

Chapitre 2

Données et méthodologie

2.1 Échantillon

2.1.1 Recrutement

L'échantillon familial de sujets asthmatiques duquel proviennent les 226 individus asthmatiques étudiés dans ce projet comprend maintenant 260 sujets de départ avec leur famille. Ces individus ont été recrutés s'ils répondaient à au moins deux des critères suivants : 1) ils ont eu un minimum de deux visites en clinique pour l'asthme dans un délai d'un an, 2) ils ont été hospitalisés deux fois ou plus relativement à l'asthme à l'intérieur de la même année, ou 3) ils utilisent des corticostéroïdes (pris par voies orales ou par voies respiratoires). Leurs familles sont incluses dans l'étude si au moins un parent est disponible pour l'évaluation phénotypique, si au moins un parent n'est pas affecté et si les quatre grands-parents sont d'origine canadienne-française. Lorsque c'est possible, les grands-parents et d'autres membres de la famille sont aussi recrutés pour l'étude (Poon et al, 2004).

Le recrutement de cet échantillon familial a débuté en 1998 et à présent, 260 familles ont participé aux recherches pour un total de 1348 individus, incluant ceux qui ont demandé à être retirés du projet et dont l'ADN est aujourd'hui détruit. Au

total, l'échantillon familial compte 656 asthmatiques dont 471 asthmatiques allergiques, 166 asthmatiques non-allergiques et 19 asthmatiques sans statut d'allergie. Lorsque ce projet de maîtrise a débuté, en septembre 2005, 226 familles prenaient part à l'étude et un individu asthmatique par famille a alors été sélectionné pour former l'échantillon utilisé.

Le tableau 2.1 tiré d'une étude récente de Tremblay et al (2006) fournit certaines caractéristiques cliniques pour 223 des 226 sujets asthmatiques étudiés.

Tableau 2.1 Caractéristiques cliniques de l'échantillon familial de sujets asthmatiques de la région du Saguenay–Lac-St-Jean (n=223)

	Valeur chez les proposant
Ratio homme : femme	1 : 1,2
Âge moyen (étendue)	18 (3-46)
Âge médian	16
<i>Fumeur</i>	
Jamais	186 (84%)
Ex-fumeur	12 (5%)
Fumeur	25 (11%)
<i>Conditions respiratoires</i>	
% de FEV ₁ (σ) ^a	92,2 (16,3)
PC ₂₀ en mg/ml (σ) ^b	2,66 (3,33)
IgE sérique en μ g/l (σ)	229,1 (4,6)
<i>Sous-phénotypes</i>	
Atopie ^d	182 (82%)
HRB ^e	169 (90%)
IgE > 100mg/l	141 (63%)
IgE > 280 μ g/l	95 (43%)

Abréviation : IgE, immunoglobuline E

^aFEV₁ = Volume expiratoire forcé en 1 sec.

^bPC₂₀ = Concentration de métacholine induisant une chute du FEV₁ de 20%; moyenne arithmétique

^cHistoire d'asthme actuelle ou passée et documentée

^dDéfini comme ayant au moins une réponse positive suite à un test d'allergie cutané (diamètre du cercle \geq 3mm en 10 min)

^eHRB = Hyper réactivité bronchique

Source : Tremblay et al, 2006

2.1.2 Phénotype

Suite à la sélection des sujets et des membres de leur famille, le statut de tous les participants est déterminé grâce à l'évaluation clinique et à un questionnaire standardisé qui a été modifié afin d'inclure des questions concernant la sévérité de l'asthme, l'histoire familiale d'asthme, l'âge auquel l'asthme s'est manifesté et la présence d'autres diagnostics de problèmes respiratoires. Les participants sont définis comme ayant de l'asthme 1) si une histoire d'asthme est rapportée (d'après le questionnaire) et s'il est possible d'avoir accès à un diagnostic d'asthme (passé ou présent) fait par un médecin ou 2) si l'asthme est confirmé par un test positif de provocation à la méthacholine (pour les sujets âgés de 12 ans et plus). Tous les sujets signent un formulaire de consentement éclairé et remplissent un questionnaire généalogique (Poon et al, 2004).

2.1.3 Génotype

Des échantillons de sang de tous les participants sont recueillis et l'ADN est extrait à partir des leucocytes du sang entier en utilisant la trousse Genomic-tip 100/G (Quiagen Inc.) selon les instructions du fabricant. Selon les technologies disponibles au laboratoire, le génotype peut être obtenu avec 3 techniques distinctes, soit FP-TDI, TaqMan et Illumina.

Pour chacune des techniques, les étapes à suivre sont semblables. Les réactifs utilisés et les logiciels diffèrent, mais le principe est sensiblement le même. Il faut d'abord amplifier la région d'ADN à l'étude par PCR. Ensuite, il faut purifier le produit PCR, puis l'allonger en y incorporant soit des oligonucléotides hybridés ou des marqueurs fluorescents qui permettent de visualiser les changements de bases. Finalement c'est l'utilisation de logiciels qui permet d'interpréter les résultats obtenus et de déterminer les génotypes.

2.2 Polymorphismes (SNP) à l'étude

Afin de sélectionner les déterminants génétiques à l'étude et de procéder au génotypage, des gènes candidats doivent d'abord être choisis. Ils peuvent être ciblés soit en se basant sur l'expression différentielle du gène dans les tissus d'individus atteints comparés à des individus témoins (puces à ADN), sur la position chromosomique du gène située dans une région liée à la maladie ou en se basant sur le rôle biologique connu ou possible du gène dans la pathophysiologie de l'asthme. Le gène CX3CR1 a été retenu car il présente une expression différentielle dans les biopsies bronchiques de personnes atteintes d'asthme (Laprise et al, 2004). En ce qui a trait au gène PLAU, il a été sélectionné parce qu'il fait partie de la voie biologique du gène SERPINB2 qui lui, est différemment exprimé dans les tissus d'individus atteints d'asthme (Bégin et al, 2007), tandis que les gènes IL13 et IL4R ont été choisis car ils sont déjà reconnus pour avoir un rôle dans l'asthme (Van der Pouw Kraan et al, 1999; Kelly-Welch et al, 2003).

Pour chacun des gènes à l'étude, CX3CR1, PLAU, IL13 et IL4R, un éventail de polymorphismes d'une seule base (SNP pour *Single Nucleotide Polymorphisms*) a été sélectionné à l'aide des banques de données publiques du Centre National d'Information Biotechnologique (NCBI) et du *SNP Consortium*. Une investigation plus poussée, à l'aide d'analyses statistiques, a été entreprise afin de voir s'ils présentaient une association à l'asthme.

L'utilisation d'analyses statistiques sert à démontrer, si elle existe, une association génétique entre ces gènes et l'asthme. C'est le logiciel FBAT (pour *Family-Based Associations Tests*) (Laird et al, 2000) qui effectue un test de déséquilibre de transmission pour vérifier les déséquilibres dans la transmission des allèles parentaux vers les enfants atteints (aussi appelés *proposants*) (Spielman et Ewens, 1996). Cet outil permet également l'évaluation des fréquences alléliques et la vérification des erreurs mendéliennes. Ces analyses ont été faites sous les

modèles génétiques additif, récessif et dominant pour chaque SNP à l'étude. Ces études d'association ont révélé qu'un certain nombre de SNP sont associés à l'asthme dans l'échantillon familial de sujets asthmatiques du SLSJ. Le tableau 2.2 montre les SNP retenus pour chacun des gènes à l'étude, en plus d'indiquer l'allèle fréquent, l'allèle mineur, l'allèle associé à l'asthme ainsi que les fréquences alléliques. Les SNP décrits dans cette étude sont cités selon leur numéro de référence (rs#) figurant dans la banque de données NCBI (<http://www.ncbi.nlm.nih.gov/>).

Tableau 2.2 Polymorphismes associés à l'asthme pour les quatre gènes étudiés

Gène	SNP	Allèle fréquent (fréquence allélique)	Allèle mineur (fréquence allélique)
CX3CR1	rs1050592	T (0,686)	C (0,314)
	rs3732378	G (0,838)	A (0,162)
	rs3732379	C (0,704)	T (0,296)
	rs938203	A (0,844)	G (0,156)
	rs2669849	A (0,743)	G (0,257)
PLAU	rs2227564	T (0,800)	C (0,200)
	rs2227566	C (0,590)	T (0,410)
IL13	rs1295686	C (0,742)	T (0,258)
	rs20541	G (0,749)	A (0,251)
	rs2243204	G (0,902)	A (0,098)
	rs2243300	C (0,913)	A (0,087)
IL4R	rs3024570	C (0,932)	T (0,068)
	rs3024630	A (0,933)	G (0,067)
	rs3024660	A (0,873)	G (0,127)
	rs1805016	A (0,899)	C (0,101)

Note : les allèles en caractère gras sont les allèles associés à l'asthme

Comme l'indique le tableau 2.2, pour les SNP des gènes CX3CR1 et PLAU, ce sont les allèles fréquents qui ont été associés à l'asthme et les analyses FBAT montrent que ces allèles sont plus souvent transmis aux proposants asthmatiques, suggérant un effet de susceptibilité de ces allèles pour l'asthme. En contrepartie, les allèles mineurs sont moins souvent transmis aux sujets asthmatiques, suggérant un effet protecteur de ces allèles pour les individus non-asthmatiques.

Pour les SNP des gènes IL13 et IL4R, ce sont plutôt les allèles mineurs qui sont associés et les analyses FBAT montrent que ces allèles sont plus souvent transmis aux proposants asthmatiques, suggérant un effet de susceptibilité à ces allèles pour l'asthme.

2.3 Formation des sous-groupes

Pour cette étude, quatre sous-groupes basés sur les quatre gènes en cause sont formés. Pour qu'un individu asthmatique se retrouve dans un sous-groupe, il doit être porteur de tous les SNP associés à l'asthme pour un gène étudié. Par exemple, pour le sous-groupe du gène CX3CR1, l'individu doit avoir respectivement les allèles T, G, C, A et A pour les SNP rs1050592, rs3732378, rs3732379, rs938203 et rs2669849 (voir tableau 2.2). Si les sous-groupes contiennent uniquement les individus porteurs de tous les SNP associés à l'asthme pour un même gène, c'est pour mieux faire ressortir les différences qui pourraient exister entre les porteurs et les témoins. En effet, c'est le proposant porteur de tous ces SNP qui est le plus susceptible de présenter des caractéristiques généalogiques particulières.

Tel que représenté par les tableaux 2.3 à 2.5, les sous-groupes sont de tailles différentes. Le sous-groupe de CX3CR1 est constitué de 153 proposants, tous porteurs des 5 SNP à l'étude et celui de PLAU en compte 168, porteurs des 2 SNP étudiés. Les sous-groupes pour IL13 et IL4R comptent moins de proposants, soit respectivement 36 et 25 individus chacun, qui sont porteurs des 4 SNP étudiés pour chacun des 2 gènes en cause. La taille de ces deux derniers sous-groupes est moins importante, car pour ces gènes, c'est l'allèle mineur qui est associé à l'asthme (voir tableau 2.2).

Tableau 2.3 Nombre de proposants porteurs des différents polymorphismes (SNP) de CX3CR1

Gène	5 SNP		4 SNP		3 SNP			2 SNP				1 SNP		0 SNP (non-porteur)	Manquants
	11111		11110	1111	11100	1110	111	11000	1100	110	11	10	1	00000	
CX3CR1	153		5	19	2	2	7	4	1	1	21	2	1	0	8

1 = porteur

0 = non-porteur

_ = donnée manquante

Tableau 2.4 Nombre de proposants porteurs des différents polymorphismes (SNP) de PLAU

Gène	2 SNP		1 SNP		0 SNP (non-porteur)		Manquants
	11		10	1	00	0	
PLAU	168		27	11	1	2	17

1 = porteur

0 = non-porteur

_ = donnée manquante

Tableau 2.5 Nombre de proposants porteurs des différents polymorphismes (SNP) d'IL13 et d'IL4R

Gènes	4 SNP		3 SNP		2 SNP		1 SNP		0 SNP (non-porteur)		Manquants
	1111		1110		1100		1000		0000		
IL13	36		6		58		4		105		17
IL4R	25		7		13		28		136		17

1 = porteur

0 = non-porteur

_ = donnée manquante

2.4 Reconstitutions généalogiques

Les reconstitutions généalogiques sont effectuées à une profondeur maximale, selon ce que les sources permettent. Dans la majorité des cas, il est possible de remonter jusqu'à l'arrivée des premiers ancêtres sur le continent nord-américain (12 générations et plus). Les données généalogiques sont conservées dans le fichier BALSAC-RETRO et un rigoureux travail de validation (vérification de la cohérence des liens généalogiques qui ont été établis et de la saisie de données) est fait avant d'utiliser la généalogie pour la recherche. Toutes ces étapes sont réalisées conformément au Règlement sur la protection des renseignements personnels et l'éthique de la recherche concernant le fichier de population BALSAC et sont soumis à l'approbation du Comité d'éthique de la recherche de l'Université du Québec à Chicoutimi.

Diverses sources sont utilisées pour la reconstitution des généalogies. Il s'agit principalement du fichier de population BALSAC, du registre de population du Québec ancien (RPQA) du Programme de recherche en démographie historique (PRDH) de l'Université de Montréal, des recueils et microfilms de l'Institut généalogique Drouin et de données généalogiques rendues accessible par le groupe « BMS2000 », résultat de la collaboration de 24 sociétés de généalogie du Québec et de l'Ontario (BMS2000, 2007). D'autres sources comme les répertoires de mariages par paroisse, diocèse, ville ou comté, les dictionnaires généalogiques ou le fichier Loiselle sont également utilisées.

Le fichier de population BALSAC contient des informations historiques, démographiques et généalogiques de la population du Québec des origines à aujourd'hui. Actuellement, il comporte des informations sur près de 3,5 millions d'individus, tirées principalement d'actes de mariage, mais également de baptême et de sépulture s'étalant du 17^e au 20^e siècle et couvrant l'ensemble du territoire québécois. Les actes de mariages qui s'y retrouvent sont jumelés, facilitant ainsi

la construction de l'ascendance. En 2008, le répertoire contenait 2 525 240 actes de l'état civil (1 667 309 mariages, 677 128 baptêmes et 180 803 sépultures) en format électronique et atteindra à terme plus de 4,5 millions d'actes, dont 3,7 millions de mariages, en ajoutant les événements plus récents (Bouchard, 2008).

Pour la région du Saguenay–Lac-Saint-Jean, le fichier-réseau BALSAC contient l'ensemble des actes de l'état civil saguenéen (665 000 actes pour la période 1838-1971) permettant ainsi de reconstituer l'ensemble de la population depuis les débuts de la colonisation jusqu'à 1971 (Bouchard, 2008).

Le RPQA du PRDH contient essentiellement de l'information sur des mariages célébrés avant 1800 (PRDH, 2007), tandis que les informations fournies par l'Institut de généalogie Drouin concernent des mariages de familles canadiennes-françaises ayant eu lieu entre 1760 et 1935 (Institut généalogique Drouin, 2007). Le Groupe « BMS2000 » quant à lui, permet de consulter 1 670 700 actes de baptêmes, 2 640 810 actes de mariages et 759 415 actes de sépultures (BMS2000, 2007).

2.5 Sélection des témoins

Des groupes témoins sont utilisés dans cette étude afin de déterminer si les caractéristiques généalogiques des individus asthmatiques sont liées à leur statut de porteur ou si elles représentent simplement la structure de base de la population. Les sujets témoins sont sélectionnés dans le fichier BALSAC-RETRO qui a été construit parallèlement au fichier-réseau BALSAC afin d'alimenter divers projets de recherche en démographie génétique, en génétique des populations et en épidémiologie génétique (Jomphe et al, 2002). En plus de servir à l'entrée de données généalogiques, il est également utilisé comme outil de recherche (Jomphe et Casgrain, 2000). À ce jour, le fichier RETRO contient près de 435 000 mentions d'individu et 225 000 mentions de couple. Il est estimé que 87% des

couples se trouvant dans RETRO proviennent du fichier BALSAC et que le reste concerne en grande majorité des mariages célébrés hors du Québec ou des mariages retrouvés dans diverses sources et qui ne sont pas encore saisis dans BALSAC (Bouchard, 2008). Pour chacun des individus figurant dans RETRO, il est possible de retrouver leur nom et prénom, le lieu et la date de leur mariage, les noms de leurs parents et de leurs enfants, ainsi que leur statut migratoire et leur origine s'ils sont nés hors du Québec. Il faut cependant préciser que ce ne sont pas tous les enfants des individus en question qui se retrouvent dans le fichier. Également, il est à noter qu'aucun statut migratoire ni origine n'est accordé aux gens mariés au Québec et ayant des parents également mariés au Québec, car ce sont des critères jugés suffisants pour considérer que ces individus sont d'origine québécoise.

Un sujet témoin, sélectionné dans le fichier RETRO, est associé à chaque sujet asthmatique de l'échantillon. Il est apparié pour le lieu et la date de mariage des parents, ainsi que pour le sexe. L'étude des groupes témoins permet de déterminer les caractéristiques généalogiques qui sont propres aux porteurs des SNP donnés et celles qui définissent l'ensemble de la population d'origine des sujets. L'apparementement entre les sujets et les témoins ne doit pas être supérieur à un coefficient de 0,125, c'est-à-dire oncle-tante/neveu-nièce, tout comme l'apparementement entre les cas ou entre les témoins ne doit pas dépasser cette valeur. Un coefficient maximal de 0,125 est utilisé par souci de comparaison avec d'autres études.

À chaque couple de parents de sujets atteints, un couple résidant dans la même paroisse, marié la même année et ayant un enfant du même sexe que le sujet de départ, car c'est celui-ci qui sert de témoin, doit être associé. En premier lieu, la requête informatisée tient compte de tous ces critères en plus de considérer uniquement les individus n'ayant pas un coefficient d'apparementement plus élevé que 0,125 avec d'autres sujets de l'étude. Si la recherche ne propose aucun candidat, un écart de plus ou moins 2 ans entre la date de mariage des parents du

sujet et celle des parents du témoin est accepté. Si la recherche demeure infructueuse, un écart de plus ou moins 5 ans est considéré. Après ces tentatives, si aucun témoin n'est associé, c'est l'exactitude du lieu de mariage qui est modifiée. La région exacte est acceptée en remplacement de la paroisse exacte, suivant ensuite le même protocole pour l'année de mariage.

De l'information est manquante pour quelques couples de parents de sujets asthmatiques. Il y en a 24 qui portent la mention « non mariés », 2 la mention « mariage introuvable » et 1 qui porte la mention « inconnu » comme lieu de mariage. Afin d'être en mesure de leur apparier un témoin, certaines données sont estimées. Pour les couples sans lieu de mariage, c'est celui des parents de l'épouse qui est utilisé, car traditionnellement, le couple se marie plus souvent dans la paroisse de la femme. Pour les couples sans date de mariage, la moyenne entre la date de mariage des parents de l'époux et celle des parents de l'épouse est effectuée. À cette moyenne sont ajoutés 30 ans, soit une génération (Tremblay et Vézina, 2000), ce qui donne l'année de mariage estimée qui est utilisée pour la recherche d'un témoin.

La requête informatisée a mené à l'obtention de 202 témoins sur un total de 226 recherchés. Un travail manuel a été fait pour les 24 manquants. Comme il n'existait aucun individu pouvant être associé à un sujet asthmatique dans le fichier RETRO, des généalogies de sujets choisis dans le répertoire « BMS2000 » (BMS2000, 2007) ont été complétées et ajoutées au fichier. Ces 24 généalogies supplémentaires sont issues d'individus appariés pour les mêmes critères utilisés lors de la recherche informatisée. Afin d'avoir un groupe de témoins complet, les conditions de sélection ont dû être assouplies, en augmentant l'écart permis entre le mariage des parents des individus de départ et des parents des témoins. Par la suite, les reconstitutions généalogiques du groupe témoin sont effectuées selon la même procédure que pour les cas.

Comme les témoins sont choisis à l'aide de la base de données RETRO, aucune évaluation clinique ne peut être faite et leur statut de porteur ne peut pas être vérifié. C'est la fréquence observée dans la population d'origine qui doit être considérée, ce qui pourrait diminuer la capacité à trouver d'éventuelles différences d'apparentement et de consanguinité entre les cas et les témoins.

2.6 Analyses généalogiques

Les analyses généalogiques sont effectuées à l'aide de la bibliothèque de fonctions GENLIB (GRIG, 2008) développée par le Groupe de Recherche Interdisciplinaire en démographie et épidémiologie Génétique (GRIG) à partir du logiciel statistique S-Plus de la compagnie Insightful (Insightful, 2007). Les comparaisons entre les groupes sont réalisées à l'aide de méthodes statistiques qui tiennent compte de la structure particulière des données généalogiques. Les procédures de calcul permettent de décrire les corpus généalogiques et de les comparer à l'aide de tests d'hypothèses.

2.6.1 Analyses descriptives

À prime abord, il faut caractériser les généalogies. Il s'agit entre autres de dénombrer les mentions d'ancêtres (nombre total de liens généalogiques qui ont pu être identifiés) et les ancêtres distincts. À une génération donnée, le nombre d'ancêtres retrouvés est obtenu en cumulant chacune des apparitions des ancêtres dans toutes les généalogies à cette génération, tandis que le nombre d'ancêtres distincts est obtenu en comptant chaque ancêtre une seule fois par génération (Jomphe et al, 2002). La génération maximale atteinte peut être déterminée lorsqu'il y a interruption des branches généalogiques. Le nombre moyen d'apparitions qui est l'occurrence moyenne des ancêtres (nombre de mentions d'ancêtres/nombre d'ancêtres distincts) sert également à caractériser les ascendances.

La profondeur généalogique (P) est aussi utilisée pour décrire les généalogies. Elle est une mesure du degré d'enracinement des ascendances dans un territoire donné et indique la valeur moyenne de la génération des fondateurs dans les ascendances (Cazes et Cazes, 1996). Elle se calcule sur un ensemble d'ascendances (profondeur généalogique totale) ou sur une ascendance unique (profondeur généalogique par ascendance) (Jomphe et al, 2002). Pour cette étude, la profondeur généalogique moyenne est calculée selon la méthode de Cazes et Cazes (1996) :

$$P = \sum_{x=0}^n X \frac{F_x}{T_x}$$

La variance de la distribution est de :

$$\sigma^2 = \sum_{x=0}^n X^2 \frac{F_x}{T_x} - \left(\sum_{x=0}^n x \frac{F_x}{T_x} \right)^2$$

où : n = génération maximale

F_x = nombre de fondateurs à la génération x

T_x = nombre d'individus attendus à la génération x

La complétude (C_x) permet également de décrire le corpus. Il s'agit du rapport entre le nombre d'ancêtres retrouvés et le nombre d'ancêtres attendus, à chaque génération x (adapté de Jetté, 1991).

$$C_x \text{ (en \%)} = \left(\frac{\text{nombre d'ascendants connus à la génération } x}{\text{nombre d'ascendants attendus à la génération } x} \right) * 100$$

Par la suite, les analyses d'apparentement et de consanguinité et la caractérisation des ancêtres et des fondateurs régionaux peuvent être entreprises.

2.6.2 Apparentement et consanguinité

Les coefficients d'apparentement et de consanguinité sont mesurés par génération, ce qui permet de quantifier le phénomène observé pour différentes périodes.

La parenté entre deux individus, x et y , se mesure par le coefficient de parenté (Phi ou P) qui se définit comme la probabilité qu'à un même locus, un gène choisi au hasard chez x soit identique à un gène choisi au hasard chez y . Le calcul du coefficient de parenté se fait de la façon suivante (adapté de Wright, 1922) :

$$\Phi_{x,y} = \sum_A \sum_C (1/2)^k (1+F(a))$$

où l'ensemble des ancêtres communs à x et y correspond à A , l'ensemble des chemins généalogiques reliant x à y en passant par l'ancêtre commun est représenté par C , le nombre d'individus dans le chemin généalogique considéré est k et où le coefficient de consanguinité de l'ancêtre commun est symbolisé par $F(a)$.

Pour ce projet, c'est l'apparentement moyen de chaque regroupement d'individus qui sera utilisé aux fins de comparaisons. Il s'agit en fait de la moyenne des coefficients de parenté obtenus entre toutes les paires d'individus possibles. Lorsqu'il s'agit de l'apparentement intragroupe (à l'intérieur d'un même groupe), il faut additionner tous les coefficients de parenté des paires de sujets du groupe et effectuer une division par le nombre de paires d'individus qu'il est possible de former à l'intérieur de ce groupe, soit $n(n-1)/2$. Pour un groupe de 25 personnes par exemple, la somme des coefficients de parenté est alors divisée par 300, puisque c'est le nombre de combinaisons possibles $((25*24)/2)$. Pour l'apparentement intergroupe (entre deux groupes différents) il faut considérer les liens de parenté entre les individus d'un groupe par rapport à ceux d'un autre

groupe. Les coefficients d'apparentement pour chaque paires d'individus sont calculés pour ensuite en faire une moyenne, soit en divisant la somme par le nombre de paires distinctes qui est donné par la formule $n_1 * n_2$. Pour deux groupes de 25 personnes, la somme des coefficients de parenté est divisée par 625, soit le nombre de paires possible entre ces deux groupes ($25 * 25$).

La consanguinité est le résultat d'une union entre deux individus apparentés. Pour un individu z , dont les parents sont x et y , elle se mesure par le coefficient de consanguinité F qui se définit comme la probabilité que cet individu ait hérité de deux allèles identiques par ascendance, soit deux allèles provenant d'un même ancêtre, l'un transmis par son père et l'autre par sa mère. Le coefficient de consanguinité d'un sujet est alors égal au coefficient de parenté (Φ) de ses parents. Le calcul du coefficient de parenté se fait de la façon suivante (adapté de Wright, 1922) :

$$F = \Phi_{x,y} = \sum_A \sum_C (1/2)^k (1 + F(a))$$

où l'ensemble des ancêtres communs à x et y correspond à A , l'ensemble des chemins généalogiques reliant x à y (le père et la mère) en passant par l'ancêtre commun est représenté par C , le nombre d'individus dans le chemin généalogique considéré est k et où le coefficient de consanguinité de l'ancêtre commun est symbolisé par $F(a)$.

Afin de comparer les résultats obtenus chez les individus asthmatiques avec ceux des témoins, des tests statistiques sont effectués. Ces tests ont pour but de démontrer s'il existe ou non une différence entre la population d'origine (les témoins) et les individus atteints. Pour les coefficients d'apparentement, il s'agit d'un test de permutation et pour les coefficients de consanguinité, c'est un test de Student qui est utilisé.

2.6.3 Caractérisation des ancêtres

Les individus portant l'appellation d'ancêtres sont tous ceux qui contribuent à la généalogie d'un sujet de départ. C'est l'ensemble des individus qui sont retrouvés dans les arbres généalogiques, à l'exception des sujets asthmatiques ou des témoins.

Les ancêtres sont caractérisés par l'occurrence, qui est le nombre de fois qu'un ancêtre apparaît dans un corpus d'ascendances, ainsi que par le recouvrement qui indique le nombre de sujets distincts auxquels un ancêtre contribue génétiquement, soit la somme des généalogies dans lesquelles on le retrouve (Jomphe et al, 2002).

La contribution génétique (C_g) des ancêtres est également considérée. Cette mesure se définit de deux façons : l'une étant la probabilité qu'un gène d'un ancêtre soit transmis à un sujet et l'autre, la proportion des gènes d'un sujet qui provient d'un ancêtre donné (Jomphe et al, 2002). La contribution génétique dépend du nombre d'apparitions des ancêtres dans les généalogies et du nombre de générations qui séparent le sujet de l'ancêtre. Effectivement, plus un ancêtre apparaît souvent dans les ascendances, plus sa contribution génétique est grande, de même que son impact sur le patrimoine génétique du groupe considéré. La contribution génétique totale d'un ancêtre à un groupe de sujets correspond à la somme de ses contributions à chacun des sujets d'un corpus généalogique et se mesure de la façon suivante (adapté de Heyer et al, 1997) :

$$CG = \sum_{\mathbf{S}} \sum_{\mathbf{C}} (1/2)^g$$

où l'ensemble des sujets de l'échantillon correspond à \mathbf{S} , l'ensemble des chemins généalogiques entre l'ancêtre et chaque sujet est représenté par \mathbf{C} et où g est le

nombre de générations, dans chaque chemin généalogique, qui sépare l'ancêtre du sujet.

2.6.4 Caractérisation des fondateurs régionaux

Dans le cadre de ce projet, les fondateurs régionaux sont définis comme des individus mariés au Saguenay–Lac-St-Jean, dont les parents se sont mariés en dehors de la région. En théorie, ils sont les premiers de leur lignée à avoir contribué au pool génique de la région.

Tout comme les ancêtres, les fondateurs régionaux peuvent être caractérisés par l'occurrence et le recouvrement. La première mesure dépend du nombre de fois que chacun apparaît dans les généalogies, tandis que la deuxième indique le nombre de sujets auxquels le fondateur est relié. Règle générale, plus la valeur d'occurrence d'un fondateur est élevée, plus il recouvre de généalogies.

Le lieu d'origine s'avère une donnée intéressante dans la caractérisation des fondateurs régionaux. Il est attribué en fonction du lieu de mariage des parents. Cette donnée peut permettre la division du groupe, aidant ainsi à réaliser des analyses plus fines faisant ressortir certaines particularités plus difficiles à cerner ou pouvant être plus spécifiques à certaines régions, comme permettre de vérifier la part du pool génique provenant des différentes régions.

Au même titre que la contribution génétique totale des ancêtres, celle des fondateurs régionaux s'obtient en faisant la somme de ses contributions génétiques des fondateurs à chacun des sujets d'un groupe. De plus, la contribution génétique des fondateurs régionaux peut être définie en fonction de leur lieu d'origine. Il s'agit de sommer les contributions génétiques des fondateurs ayant une même région comme lieu d'origine, ce qui permet de connaître la part