

BALSAC et la démogénétique : une interface entre les sciences biologiques et sociales

HÉLÈNE VÉZINA

1. INTRODUCTION

Le fichier de population BALSAC représente une des grandes réalisations de Gérard Bouchard, qui a piloté son développement pendant 40 ans. Aujourd’hui, BALSAC est une infrastructure de recherche majeure qui contribue au rayonnement scientifique national et international du Québec. Il constitue une référence mondiale pour les études généalogiques appliquées à la génétique des populations et à l’épidémiologie génétique. Ce texte ainsi que celui de Claude Bhérier portent sur le volet génétique de l’histoire de BALSAC, qui a été lancé au début des années 1980 et qui, en plus d’avoir suscité de nombreux travaux à caractère épidémiologique, historique ou démographique, a entraîné des choix déterminants concernant le développement du fichier.

Dans cet article, il est d’abord question de la mise en place par Gérard Bouchard et ses collaborateurs d’un programme de recherche sur les maladies héréditaires, programme dont le déroulement s’est échelonné sur un peu plus d’une décennie et qui a culminé en 1991 avec la publication d’une synthèse de ces travaux pionniers dans le désormais classique *Histoire d’un génome*. La seconde partie porte sur les études génétiques réalisées à l’aide de BALSAC depuis 25 ans. Comme il n’est pas possible de rendre justice à tous les chercheurs-utilisateurs de BALSAC ainsi qu’à tous les sujets traités au cours de cette période, je me concentre sur un domaine de recherche que l’on nomme *démogénétique*. En dernière partie, je dresse un portrait de la situation actuelle et des travaux en cours à BALSAC et je partage quelques réflexions sur les orientations souhaitées pour les prochaines années.

2. LE PROGRAMME DE RECHERCHE SUR LES MALADIES HÉRÉDITAIRES À BALSAC : DE SA MISE EN PLACE À LA PUBLICATION D'*HISTOIRE D'UN GÉNOME*

En 1979, Gérard Bouchard entreprend une étude sur la dystrophie myotonique au Saguenay–Lac-Saint-Jean (appelé Saguenay dans la suite du texte), jetant ainsi les bases de ce qui deviendra un vaste programme de recherche en génétique humaine. Bien qu'il ait pu compter sur la collaboration de deux médecins-généticiens, Claude Laberge à l'Université Laval et Charles Scriver à l'Université McGill, il fallait certainement beaucoup d'audace pour s'engager en tant qu'historien dans une telle entreprise. Par la suite, Gérard Bouchard a su réunir autour de cette thématique des chercheurs¹ provenant de diverses disciplines, amorçant ainsi une approche multidisciplinaire qui caractérise encore aujourd'hui une grande partie des travaux appuyés sur les données du fichier. Les projets réalisés au sein du programme de recherche sur les maladies héréditaires ont donné lieu à une production scientifique impressionnante². En effet, entre 1981 et 1993, on recense une trentaine d'études portant sur 16 maladies dont 11 présentes au Saguenay, une vingtaine d'articles de synthèse ainsi que deux ouvrages. Ces travaux peuvent être regroupés sous deux thèmes principaux. Une première série repose sur la reconstruction et l'analyse des généalogies d'individus atteints de maladies héréditaires et vise à comprendre les modalités d'introduction et de diffusion des gènes associés à ces maladies sur le territoire. Ces études ont porté surtout sur des maladies ayant une fréquence accrue dans les régions de l'est québécois et plus particulièrement au Saguenay. La seconde série se penche sur la démographie génétique de la population du Saguenay avec l'objectif de documenter l'histoire du peuplement et ses effets sur la struc-

-
1. En plus de Laberge et Scriver, on pense à Bernard Lemieux, médecin-généticien, à Marc De Braekeleer, médecin et épidémiologiste, à Évelyne Heyer, généticienne des populations, à Bartha Knoppers, juriste qui fut une des premières au Québec à se pencher sur les questions éthiques reliées à ce type de recherche, ainsi qu'à plusieurs médecins-cliniciens du Saguenay, notamment Jean Larochelle, Jean Mathieu et Charles Morin.
 2. On peut trouver une revue de ces travaux ainsi qu'une bibliographie exhaustive dans Vézina, 1996. On peut aussi consulter les chapitres 16 et 18 d'*Histoire d'un génome* (De Braekeleer, 1991c et 1991d).

ture généalogique et la composition génétique de la population contemporaine.

Concernant le Saguenay, les divers projets de recherche ont permis de dégager des constats importants. D'abord, on a pu vérifier que, contrairement à une croyance largement répandue, les mariages consanguins n'ont pas été plus nombreux dans cette région qu'ailleurs au Québec et qu'ils ne peuvent donc pas expliquer la fréquence observée de certaines maladies génétiques (Bouchard, 1991). Les analyses généalogiques ont aussi démontré que les gènes associés aux maladies héréditaires ont été introduits dans la région par plusieurs individus, et non pas par un ou quelques couples fondateurs, et que ceci s'explique par le mode de formation de la population, notamment l'importance de Charlevoix comme région d'origine des pionniers (Bouchard *et al.*, 1988). Enfin, on a pu constater que la diffusion de ces gènes à l'échelle régionale s'expliquait principalement par l'endogamie et par une fécondité élevée (Bouchard, 1989).

Les travaux sur les maladies héréditaires ont aussi permis de montrer certaines tendances concernant la structure génétique de la population du Québec. Ainsi, on a constaté que les modalités du peuplement ont entraîné une division entre l'est de la province, plus homogène, et l'ouest, plus diversifié, et que ceci s'expliquerait en partie par le lieu d'installation après la fin du Régime français (1760) des immigrants qui étaient en majorité d'origine autre que française, ainsi que par les courants migratoires interrégionaux (Bouchard, 1990b). Cette différenciation entre l'est et l'ouest du Québec s'observe entre autres dans le type de maladies génétiques propre à chacune des deux parties : dans l'est, on retrouve des maladies qui se voient très peu ailleurs, alors que les principales maladies identifiées dans l'ouest sont aussi présentes dans les autres populations européennes ou d'origine européenne (Bouchard, 1990b et 1993a).

En 1991 paraît *Histoire d'un génome*, sous la direction de Gérard Bouchard et Marc De Braekeleer (Bouchard et De Braekeleer, 1991a). Cet ouvrage constitue une synthèse des travaux pionniers réalisés sur le Saguenay et trace la voie pour la réalisation de recherches sur de nouvelles thématiques comme les maladies multifactorielles ou les considérations éthiques et

juridiques entourant ces travaux. Le livre représente une contribution majeure sinon unique au développement des études en génétique des populations au Québec et on peut même se demander s'il existe, à ce jour, une autre population de la taille de celle du Saguenay (près de 295 000 personnes au tournant des années 1990) qui a pu être étudiée de façon aussi approfondie sous l'angle de son histoire démographique, de sa structure généalogique et des conséquences sur le patrimoine génétique et la santé de la population.

Pour la population du Saguenay, ces études effectuées à l'aide du fichier BALSAC ont favorisé une meilleure compréhension du rôle des comportements démographiques et des caractéristiques du peuplement dans l'introduction et la diffusion des maladies héréditaires. D'ailleurs, en 1980, Gérard Bouchard participe à la création de la Corporation de recherche et d'action sur les maladies héréditaires (CORAMH) dans le but de favoriser ce qu'on appelle aujourd'hui le transfert de connaissances vers les communautés et les familles touchées par les maladies étudiées. En 1992, un an après *Histoire d'un génome*, paraît un petit livre intitulé *Pourquoi des maladies héréditaires?* (Bouchard et De Braekeleer, 1992) qui témoigne encore une fois de ce souci de faciliter le partage des connaissances et des résultats de la recherche avec le grand public. CORAMH est aujourd'hui devenu un acteur incontournable pour tout ce qui a trait à la prévention des maladies héréditaires au Saguenay.

La mise en place du programme de recherche sur les maladies héréditaires, en plus de constituer le point de départ de l'utilisation de BALSAC en génétique, a été déterminante pour la poursuite du développement du fichier. Rappelons que la première étape de sa construction, complétée en 1986, a consisté en la reconstitution de la population du Saguenay effectuée à partir des 660 000 actes d'état civil de la période 1842-1971. Par la suite, les travaux se sont poursuivis et ont d'abord porté sur la région de Charlevoix puis sur quelques régions de l'est du Québec. Le nom BALSAC est d'ailleurs un acronyme tiré des premières lettres du nom de certaines de ces régions (Bouchard, 1984). En 1989, l'équipe de Gérard Bouchard prend la décision d'étendre le fichier à l'ensemble du Québec et de restreindre la saisie aux actes

de mariage³. Ce dernier choix a été dicté par les coûts associés à un tel projet, mais aussi par l'importance accordée aux reconstructions généalogiques – qui reposent d'abord et avant tout sur les actes de mariage – pour la recherche en génétique.

3. L'ÉTUDE DÉMOGÉNÉTIQUE DES POPULATIONS RÉGIONALES DU QUÉBEC : UN APERÇU DE QUELQUES TRAVAUX RÉALISÉS DEPUIS 25 ANS

Des années 1990 à aujourd'hui, la recherche génétique appuyée sur BALSAC s'est poursuivie, évoluant notamment en fonction des avancées scientifiques et technologiques de la génétique moléculaire. L'approche généalogique, c'est-à-dire l'exploitation du fichier basée sur la reconstitution des lignées généalogiques, demeure au cœur des travaux alors qu'avec l'expansion progressive du fichier, les projets d'abord focalisés sur la région du Saguenay vont s'étendre à l'ensemble des populations régionales du Québec. Dans cette section, je trace un portrait des travaux en démogénétique en les présentant selon quatre thématiques qui n'ont pas toutes la même importance eu égard au nombre d'études et de publications, mais qui ensemble sont représentatives des sujets abordés. Chacun de ces thèmes sera illustré à l'aide d'un projet spécifique.

Précisons d'abord que la démogénétique est un champ de recherche qui constitue une interface entre les sciences biologiques et sociales préoccupées par les questions de population. En effet, tout comme la génétique des populations, elle vise à décrire la structure génétique d'une population, mais pour ce faire elle a recours aux données et aux méthodes de la démographie historique et en particulier aux reconstructions généalogiques. Plus particulièrement, la démogénétique s'applique à documenter le rôle de l'histoire démographique d'une population dans l'évolution de son patrimoine génétique. Les lignées généalogiques revêtent une importance cruciale en démogénétique, car la

3. Plus précisément, il a aussi été convenu que le travail de saisie porte sur les actes des XIX^e et XX^e siècles, les actes des XVII^e et XVIII^e siècles ayant déjà été dépouillés par le PRDH pour la construction du *Registre de la population du Québec ancien*. Les actes de mariage du PRDH ont par la suite été intégrés au fichier BALSAC par le biais d'une entente de collaboration.

généalogie d'une population contient de l'information sur les ancêtres qui ont participé à la reproduction et sur les liens d'apparentement qui unissent ces ancêtres entre eux. Sous l'angle de la génétique, ces ancêtres sont les individus qui ont transmis leurs gènes jusqu'à la population contemporaine, et les liens d'apparentement sont les chemins de transmission de ces gènes au fil des générations. Le patrimoine génétique d'une population est donc formé de l'ensemble des gènes transmis par ces ancêtres, et l'étude des liens généalogiques permet de décrire et d'expliquer certaines caractéristiques de ce patrimoine.

3.1 DU SAGUENAY AUX POPULATIONS RÉGIONALES DU QUÉBEC : UNE ÉTUDE COMPARATIVE DE LA CONSANGUINITÉ

Les travaux réalisés sur la démographie génétique des populations régionales s'inscrivent dans le prolongement des études sur le Saguenay. Ils se sont étendus d'abord aux populations de l'est de la province puis à l'ensemble des régions en adoptant dans certains cas une perspective comparative pour mettre en contexte les particularités du Saguenay et du nord-est québécois ou plus globalement pour montrer les ressemblances et les différences entre les caractéristiques démogénétiques des régions du Québec. Plusieurs études se sont penchées sur la contribution des immigrants fondateurs de diverses origines au patrimoine génétique québécois et ont permis de confirmer l'importance des fondateurs français, mais aussi de montrer le poids non négligeable d'autres groupes, comme les Acadiens et les Irlandais. Les résultats de ces travaux ont donné lieu à une production scientifique variée pouvant servir autant aux chercheurs en sciences sociales qui s'intéressent à l'histoire régionale qu'à ceux des sciences biomédicales qui travaillent sur des cohortes de patients recrutés dans la population québécoise⁴.

D'autres projets ont porté sur les caractéristiques du peuplement des régions du Québec et sur leur influence sur la structure généalogique et génétique de la population contemporaine. Ainsi,

4. Voir par exemple Bergeron *et al.*, 2008; Bhérer *et al.*, 2011; Heyer *et al.*, 1997; Tremblay et Vézina 2010; Tremblay *et al.*, 2009; Vézina *et al.*, 2005.

j'ai entrepris à partir de 1998, avec mon collègue Marc Tremblay de l'Université du Québec à Chicoutimi, le développement de corpus généalogiques couvrant tout le territoire québécois. Notre premier projet a porté sur la mesure de l'apparement et de la consanguinité dans les populations régionales. Comme nous l'avons vu plus haut, le fichier BALSAC a d'abord couvert le Saguenay, ce qui donne à cette région l'avantage d'avoir été énormément étudiée et d'être sans doute celle dont on connaît le mieux le patrimoine génétique au sein de la population québécoise. Mais cela a aussi contribué à l'apparition de quelques stéréotypes dont celui d'une fréquence élevée de certaines maladies héréditaires qui s'expliquerait par l'isolement géographique et par un niveau de consanguinité plus important qu'ailleurs. Bien que dès les années 1980 Gérard Bouchard ait démontré dans ses travaux que ceci était faux, cette idée a continué (et continue toujours) à être véhiculée (Bouchard, 2004b). Nous avons donc voulu tester à notre tour, à l'aide d'une approche comparative, l'hypothèse que la consanguinité n'était pas plus élevée au Saguenay qu'ailleurs au Québec.

Pour réaliser notre étude, nous avons divisé le territoire québécois en 26 régions en nous fondant sur des critères géographiques et démographiques et nous avons reconstruit les généalogies d'individus mariés entre 1935 et 1974 dans chacune de ces régions. À l'aide des données généalogiques, nous avons ensuite calculé le coefficient de consanguinité de chacun de ces individus, ce qui nous a permis d'estimer le coefficient moyen de consanguinité dans chaque région. En démogénétique, on distingue deux formes de consanguinité : la consanguinité proche et la consanguinité éloignée. La consanguinité proche se calcule en prenant en compte les ancêtres communs dans les trois ou quatre premières générations d'ancêtres. Il s'agit donc d'une forme de consanguinité qui est généralement connue par les personnes concernées et qui nous renseigne notamment sur les pratiques en matière de choix du conjoint dans une société. La consanguinité éloignée considère les ancêtres communs ayant vécu il y a plusieurs générations et elle permet de cerner les conséquences génétiques de l'histoire démographique d'une population. Au Québec (et à l'exception des Amérindiens bien sûr), on peut remonter dans la

plupart des cas jusqu'aux premiers ancêtres d'origine européenne arrivés sur le territoire.

Nos résultats ont démontré que la consanguinité proche n'est pas plus élevée au Saguenay que dans les autres régions. Les mariages entre proches apparentés ne peuvent donc absolument pas constituer une explication à la fréquence accrue de certaines maladies héréditaires observées dans la région. Nos analyses ont aussi révélé que la consanguinité éloignée est généralement plus élevée dans les régions de l'est de la province (incluant donc le Saguenay) que dans l'ouest, ce qui indique que ce sont plutôt les caractéristiques de l'histoire du peuplement qui constituent un facteur explicatif des fréquences observées (Vézina *et al.*, 2004).

3.2 L'apport des données moléculaires : la contribution amérindienne au patrimoine génétique des Québécois

À partir du début des années 2000, les avancées de la biologie moléculaire et de l'informatique ont rendu possible la collecte et l'analyse de matériel génétique à des coûts permettant d'envisager la constitution d'échantillons de grande taille et donc la réalisation d'études en génétique des populations basées sur des données moléculaires. À BALSAC, un programme de recherche s'appuyant sur l'exploitation conjointe des données moléculaires et généalogiques pour caractériser le patrimoine génétique de la population québécoise a été mis en place en collaboration avec Damian Labuda, de l'Université de Montréal. Des participants recrutés dans plusieurs régions du Québec ont fourni les informations pour la reconstruction de leur généalogie à BALSAC ainsi qu'un échantillon biologique permettant l'extraction de leur ADN. Les analyses génétiques ont été effectuées au laboratoire du professeur Labuda et ont d'abord porté sur les marqueurs des lignées maternelles et paternelles, puis sur l'ensemble du génome. Plusieurs projets ont été menés à bien au cours des 15 dernières années, en particulier sur la Gaspésie, qui a été la première région étudiée dans le cadre de ce programme (Moreau *et al.*, 2009 ; Moreau, Vézina, Jomphe *et al.*, 2011 ; Vézina, Tremblay *et al.*, 2014). Les travaux ont notamment permis de montrer que la diversité génétique dans les groupes étudiés était importante, remettant ainsi en question un autre mythe, celui concernant

l'homogénéité génétique des Canadiens français (Moreau *et al.*, 2007). Claude Bhérer y revient au chapitre suivant.

Au sein de ce programme de recherche, les origines des ancêtres fondateurs et le métissage ont reçu une attention particulière. Ainsi, j'ai effectué avec Damian Labuda une étude sur les origines amérindiennes des Canadiens français, un sujet qui suscite énormément d'intérêt dans la communauté scientifique et dans la population, mais qui fait aussi l'objet de certaines affirmations mal fondées. Notre objectif était de répondre aux trois questions suivantes : Quelle est la contribution amérindienne au pool génique canadien-français du Québec ? Quel peut être l'apport des données génétiques dans la mesure généalogique de cette contribution ? Et, finalement, en quoi les données généalogiques et génétiques peuvent-elles être complémentaires ? Pour ce faire, nous avons mesuré la contribution des ancêtres amérindiens identifiés dans les généalogies au pool génique de 794 participants résidant dans les régions de Montréal, du Saguenay, de la Gaspésie et de la Côte-Nord. Comme les données généalogiques sur les ancêtres amérindiens sont incomplètes et imprécises, nous avons eu recours aux données génétiques. En effet, grâce aux informations fournies par l'ADN mitochondrial et le chromosome Y, il a été possible d'identifier des ancêtres amérindiens additionnels au sein des lignées maternelles et paternelles, respectivement. L'ADN contenu dans les mitochondries est transmis lors de la reproduction uniquement par la mère. Ceci veut dire que les filles le transmettront à leur tour, mais pas les fils, faisant de l'ADN un marqueur des lignées maternelles largement utilisé en génétique des populations. Les lignées paternelles peuvent de leur côté être retracées grâce au chromosome Y qui détermine le sexe masculin. À l'aide des analyses génétiques, nous avons déterminé l'haplogroupe de chaque individu. L'haplogroupe est une signature génétique permettant de préciser l'origine ancestrale (ethnique ou géographique) des lignées maternelles et paternelles.

Nos résultats ont montré que plus de la moitié des participants avaient au moins un ancêtre amérindien dans leur généalogie (Vézina *et al.*, 2012). Ceci signifie que la majorité des participants peuvent avoir hérité de gènes provenant de fondateurs amérindiens. Cependant la contribution génétique totale de ces ancêtres aux quatre groupes régionaux demeure faible. En

effet, elle est de moins de 1% au Saguenay et à Montréal alors qu'elle dépasse à peine 1% sur la Côte-Nord et en Gaspésie. On voit ainsi l'importance de bien distinguer la présence d'ancêtres amérindiens dans l'arbre généalogique et le poids de ces ancêtres dans la composition du génome. Grâce aux données génétiques, nous avons pu détecter de nouvelles lignées amérindiennes sur lesquelles nous n'étions pas en mesure de statuer à partir des seuls renseignements généalogiques à notre disposition, faisant la démonstration du potentiel de complémentarité des deux types de données. Dans une étude subséquente, des analyses génétiques plus poussées ont permis de vérifier la corrélation avec les informations généalogiques et de confirmer que la contribution amérindienne au pool génique canadien-français se situe vraisemblablement autour de 1% (Moreau *et al.*, 2013).

3.3 Des maladies héréditaires aux traits complexes: introduction et diffusion d'une mutation associée au cancer du sein

Plusieurs chercheurs s'intéressant aux maladies héréditaires ont utilisé les données du fichier BALSAC afin de reconstruire et d'analyser les généalogies d'individus atteints. Comme nous l'avons vu plus haut, ces travaux ont d'abord porté sur des maladies ayant une fréquence élevée dans les régions de l'est du Québec et plus particulièrement dans la population du Saguenay. En plus de montrer le rôle de l'histoire du peuplement dans l'apparition et la diffusion de ces maladies, on a pu identifier dans plusieurs cas des immigrants fondateurs qui pourraient avoir introduit les gènes responsables. Les progrès de la génétique moléculaire ont permis de localiser les gènes entraînant le développement de plusieurs de ces maladies et d'identifier une ou des mutations du gène. Il est donc progressivement devenu possible de raffiner les analyses en les réalisant sur des groupes de patients porteurs de l'une ou l'autre des mutations identifiées (Labuda *et al.*, 1996; Plante *et al.*, 2008; Yotova *et al.*, 2005). Des chercheurs ont aussi travaillé sur des maladies observées à l'échelle locale ou régionale à l'extérieur du Saguenay, ce qui a contribué à une meilleure connaissance du contexte démohistorique des maladies héréditaires au Québec (voir par exemple Laberge, Jomphe *et al.*, 2005; et Vézina, Tremblay *et al.*, 2005). Enfin, des analyses généalogiques

ont été entreprises sur des cohortes de patients atteints de maladies multifactorielles dans le but de mieux comprendre le rôle des facteurs génétiques dans l'étiologie de ces maladies et de contribuer à l'élaboration de protocoles de recherche en épidémiologie génétique. Par exemple, des travaux ont été effectués sur la maladie d'Alzheimer (Vézina *et al.*, 1999), sur les affections bipolaires (Morissette *et al.*, 1999) et sur l'hypertension (Pausova *et al.*, 2002).

Dans le cadre d'un programme de recherche sur la génétique et l'épidémiologie moléculaire du cancer du sein dirigé par Jacques Simard, de l'Université Laval, j'ai réalisé avec Francine Durocher, aussi de l'Université Laval, une étude démogénétique ayant pour objectif de documenter les modalités d'introduction et de diffusion dans la population canadienne-française d'une mutation associée au cancer du sein (Vézina, Durocher *et al.*, 2005). Des reconstructions généalogiques ont été effectuées dans 18 familles porteuses d'une mutation du gène BRCA1 qui est associé au cancer du sein héréditaire. Les caractéristiques généalogiques de ces familles ont été comparées à celles de groupes témoins afin de départager celles qui pouvaient être liées à la présence de la mutation de celles s'expliquant plutôt par la structure de la population (on peut penser par exemple à la contribution génétique plus élevée de certains ancêtres). Grâce à cette démarche, nous avons identifié le couple ayant la plus forte probabilité d'avoir introduit la mutation dans la population québécoise et, en nous fondant sur les lieux de mariage de leurs descendants, nous avons pu constater une concentration régionale de la mutation.

3.4 Généalogie et biologie évolutive : la durée moyenne des générations dans les populations humaines

La démogénétique et BALSAC trouvent aussi une place dans les travaux de biologie évolutive. On passe ainsi d'études appuyées sur BALSAC et ayant pour objet la population québécoise à des recherches de nature fondamentale au sein desquelles les fichiers de population constituent des laboratoires permettant de se pencher sur certains phénomènes qui ont pu jouer un rôle dans l'histoire évolutive des populations humaines (Austerlitz et

Heyer, 1998 ; Heyer *et al.*, 2005 ; Moreau, Bhérer *et al.*, 2011 ; Peischl *et al.*, 2018).

Par exemple, la durée des générations est un paramètre fréquemment utilisé dans les études touchant à l'estimation de l'âge et des origines des mutations dans les populations humaines. Dans un projet réalisé avec Marc Tremblay, il a été possible de constater que, dans la plupart des travaux recensés, les intervalles moyens entre générations successives étaient nettement sous-estimés. Nous avons donc utilisé les données du fichier BALSAC pour proposer de nouvelles estimations (Tremblay et Vézina, 2000). À partir d'informations tirées de généalogies et d'actes de mariage couvrant une période de plus de trois siècles, la longueur moyenne des intervalles intergénérationnels a été estimée à 30 ans. De plus, nous avons montré que la longueur moyenne des intervalles féminins (28 ans) est plus courte que celle des intervalles masculins (34 ans).

4. BALSAC D'HIER À DEMAIN : QUELQUES PISTES DE DÉVELOPPEMENT ET D'EXPLOITATION

La recherche génétique à BALSAC s'est développée à partir d'un fichier sur la population saguenéenne et d'un programme d'étude des maladies héréditaires. Par la suite, les travaux d'enrichissement du fichier se sont étendus à tout le Québec, permettant ainsi de mener des activités de recherche portant sur l'ensemble des populations régionales. Où en est BALSAC aujourd'hui ? Quelles sont les perspectives pour son évolution et son exploitation en génétique humaine ? En 2010, j'ai succédé à Gérard Bouchard à la direction du Projet BALSAC. Aussi, dans cette dernière partie, en plus de faire le point sur la situation actuelle du fichier de population BALSAC, je souhaite partager quelques réflexions qui sous-tendent les orientations prises au cours des dernières années. Au cœur de ces réflexions demeure l'objectif central de réaliser pleinement la mission du Projet BALSAC, qui est de mettre en place les conditions qui favorisent l'expansion et l'exploitation du fichier, de valoriser les possibilités de recherche auprès de la communauté scientifique et d'offrir les services les plus complets possible pour soutenir l'utilisation du fichier.

4.1 BALSAC et les microdonnées historiques québécoises

De 2001 à 2011, le fichier a connu une phase de croissance majeure qui a permis d'y intégrer près de 2 millions d'actes de mariage. Il contient maintenant tous les actes de mariage de la population catholique depuis les débuts de la Nouvelle-France jusqu'à 1965. Ces actes ont été informatisés et reliés entre eux grâce à une méthodologie de jumelage des données nominatives mise au point à BALSAC. La population non catholique est moins bien recensée puisque seulement quelques régions et périodes ont fait l'objet d'une saisie. De plus, la qualité du contenu de ces actes rend difficile leur utilisation à des fins de jumelage.

BALSAC comprend aujourd'hui plus de 2,2 millions d'actes de mariage sur l'ensemble du Québec et 550 000 actes de naissances et de décès (pour le Saguenay). Cinq millions d'individus répartis dans 2,5 millions de familles y sont répertoriés, ce qui permet la reconstruction automatique des lignées généalogiques et de la structure de la parenté sur plus de 300 ans. Le fichier est sous la responsabilité de quatre universités : l'Université Laval, l'Université de Montréal, l'Université McGill et l'Université du Québec à Chicoutimi, qui en est aussi gestionnaire. Il est utilisé par des chercheurs du Québec, du Canada et de l'étranger dans des domaines reliés principalement à la génétique des populations, à l'épidémiologie génétique et aux sciences sociales. Pour l'année 2017, on recense 47 projets de recherche faisant usage des données du fichier, dont 14 en épidémiologie génétique et 14 en biologie évolutive. De plus, il est à noter que parmi les chercheurs associés à ces projets plus d'un cinquième proviennent de l'extérieur du Canada (BALSAC, 2017).

À partir de 2010, BALSAC a entrepris une réorganisation visant à adapter la structure de la base de données pour rendre le processus d'intégration des données plus flexible et ainsi permettre l'ajout de données autres que celles provenant de l'état civil. C'est dans ce contexte qu'a débuté en 2013 la construction de l'IMPQ, l'Infrastructure intégrée des microdonnées historiques de la population québécoise, en partenariat avec le Programme de recherche en démographie historique et le Centre interuniversitaire d'études québécoises. Ce projet a permis de développer une

structure informatique pour la mise en commun des microdonnées démohistoriques existantes (tirées de l'état civil et des recensements canadiens) et l'intégration de nouveaux jeux de données. De plus, grâce à un programme de jumelage conçu à BALSAC, les données d'état civil et celles des recensements ont été jumelées entre elles, ce qui offre des possibilités de recherche tout à fait inédites. En effet, avec l'IMPQ, les chercheurs auront accès à un vaste ensemble d'informations biographiques permettant l'étude des populations historiques sur la base des trajectoires individuelles au sein des familles, des ménages et des communautés dans une perspective multigénérationnelle (Vézina *et al.*, 2015 et 2016). Grâce à l'IMPQ, les données du fichier BALSAC sont désormais intégrées à une vaste infrastructure qu'il est possible de consulter par le biais d'un portail accessible à la communauté scientifique depuis l'automne 2018. Ces développements permettront non seulement d'élargir de façon considérable les avenues de recherche en sciences sociales, mais ils auront aussi des retombées pour les études démogénétiques et plus largement les recherches biomédicales faisant usage des données généalogiques.

4.2 Vers un portrait démogénétique de la population québécoise

Comme nous l'avons vu précédemment, les travaux appuyés sur le fichier BALSAC ont permis le développement des connaissances sur la structure démogénétique de la population québécoise contemporaine et sur les facteurs sociodémographiques qui ont façonné le pool génique depuis les débuts du peuplement d'origine européenne jusqu'à aujourd'hui. Dans plusieurs cas, ils ont aussi servi à nuancer certaines affirmations, voire à démentir certaines croyances en offrant un cadre d'interprétation permettant d'élargir les perspectives. On pense par exemple à la mise en contexte de la situation saguenéenne par rapport à celle des autres régions du Québec, de celle du Québec par rapport à d'autres populations de formation récente, à la distinction entre la consanguinité proche qui relève des pratiques matrimoniales et la consanguinité éloignée qui se comprend dans le contexte de l'histoire du peuplement. Force est cependant de constater que les mythes ont la vie dure – Gérard Bouchard pourrait certainement nous entretenir sur le sujet – et que leurs effets pernicious peuvent

se perpétuer longtemps tant dans le monde scientifique qu'au sein des communautés concernées. Les travaux de recherche doivent donc se poursuivre tout comme les efforts de diffusion et de promotion des résultats.

À BALSAC, nous avons entrepris depuis quelques années un projet visant à tracer un portrait démogénétique de la population québécoise. Ce portrait, qui se veut en quelque sorte une suite à *Histoire d'un génome*, repose sur une synthèse des travaux réalisés depuis 25 ans et sur des analyses qui produiront une description globale de la structure généalogique de la population québécoise en termes de différenciation spatiale et d'évolution temporelle. Pour l'instant, les résultats sont progressivement intégrés au site Internet de BALSAC sous forme de courtes capsules. À terme, nous souhaitons que ce portrait puisse constituer un outil de référence sur la formation et l'évolution des bassins génétiques régionaux du Québec et qu'il soit présenté sous une forme permettant de rejoindre la communauté scientifique, les intervenants en santé ainsi que le grand public. On pourrait penser par exemple à la production d'un atlas interactif en ligne.

4.3 BALSAC à l'ère de la génomique

Depuis plus de 40 ans, BALSAC a suscité et soutenu des études novatrices et multidisciplinaires en génétique humaine, mais aussi, comme cela a été présenté dans les chapitres de Danielle Gauvreau et de Marc St-Hilaire, en sciences sociales. En collaboration avec les chercheurs-utilisateurs du fichier (près de 50 en 2017), l'équipe de BALSAC suit de près les progrès théoriques, méthodologiques et technologiques survenus au sein de ces divers domaines afin de s'assurer que le fichier demeure une infrastructure de recherche pertinente en mesure de répondre aux besoins et aux demandes de la communauté scientifique et de contribuer à l'avancement des connaissances. Au cours des prochaines années, BALSAC planifie son développement en réponse à trois préoccupations. D'abord, les chercheurs en génomique qui travaillent avec les informations généalogiques souhaitent l'intégration au sein d'une structure formelle et pérenne des données généalogiques et génétiques, car, dans

le mode de fonctionnement actuel de BALSAC, celles-ci ne sont mises en relation que dans le cadre de projets spécifiques pour une durée déterminée. Deuxièmement, le prolongement jusqu'au XX^e siècle de la période pour laquelle nous disposons d'une reconstitution exhaustive de la population québécoise établie à partir de l'ensemble des actes de l'état civil (naissances, mariages et décès) suscite un grand intérêt chez les chercheurs tant du côté des sciences sociales que de celui des sciences biologiques, entre autres pour les études de biologie évolutive. Enfin, nous souhaitons offrir un cadre d'analyse et d'interprétation de la dimension spatiale pour les données généalogiques et génétiques.

C'est dans ce contexte que le projet de construction de i-BALSAC a été élaboré en partenariat avec le Centre interuniversitaire d'études québécoises et la plateforme CARTaGENE. Il permettra d'effectuer une cartographie haute résolution de la population canadienne-française par une approche génomique, généalogique et géographique tout en offrant un regard historique couvrant quatre siècles. L'expertise acquise au cours des dernières décennies en matière d'exploitation conjointe des données généalogiques et génétiques sera bonifiée par l'élaboration d'outils statistiques et technologiques de pointe et par l'intégration d'une perspective spatiale.

5. POPULATION ET GÉNÉTIQUE: Y A-T-IL TOUJOURS UNE NOUVELLE FRONTIÈRE POUR LES SCIENCES SOCIALES ?

Le titre de cette dernière section est inspiré d'un article publié par Gérard Bouchard en 1993 et intitulé «Population et génétique: une nouvelle frontière pour les sciences sociales» (Bouchard, 1993b). Dans ce texte, il réfléchit sur les relations entre sciences sociales et génétique humaine, et sur les possibilités de recherche sur «des problématiques conjointes de génétique et de population» qu'il voit s'en dégager (*Ibid.*, p. 398). Il propose une matrice pouvant soutenir une démarche interdisciplinaire basée sur le concept de dynamique de population et entrevoit un avenir prometteur pour les fichiers de population, qu'il présente comme des «machines à croiser les données et à décloisonner les aires d'interrogation, ouvrant sur la diachronie, le spatial et le multidimensionnel» (*Ibid.*, p. 406).

Les contours de cette frontière se sont certainement transformés depuis 25 ans, mais cette réflexion plus large sur le potentiel de collaboration entre chercheurs des sciences sociales et biomédicales, et en particulier sur l'apport possible de la démographie et des sciences sociales à la génétique, paraît toujours d'actualité. L'existence d'une interface contribuant à ouvrir cette frontière et fournissant un lieu de contacts et d'échanges pour les scientifiques préoccupés par les questions de population et de génétique semble plus que jamais pertinente alors que les problématiques (on peut penser par exemple à l'épigénétique), les questionnements et les enjeux qu'elles soulèvent ainsi que les données et les méthodes nécessaires pour y répondre sont de plus en plus complexes. À ce titre, BALSAC, au sein duquel sont réalisées des études interdisciplinaires depuis plusieurs décennies, et son créateur Gérard Bouchard font certainement figure de précurseurs. Avec le lancement prochain de l'IMPQ et le projet de création d'i-BALSAC, notre objectif est que BALSAC et ses partenaires puissent continuer à jouer le rôle de laboratoire d'étude des populations permettant d'effectuer des recherches innovantes et de susciter les collaborations entre chercheurs d'horizons disciplinaires variés.