

## Rapport de bourse de rédaction CICC – Hiver 2020

**Projet :** Validation à grande échelle d'un modèle de simulation pour déterminer la fréquence des haplotypes y dans la population canadienne-française

Depuis près de 30 ans, l'analyse des marqueurs génétiques est utilisée dans le milieu des sciences forensiques comme preuve à conviction dans les dossiers criminels. D'emblée, ce sont les marqueurs situés sur les chromosomes autosomaux qui sont analysés. Or, dans certains cas particuliers, l'utilisation de ces marqueurs s'avère insuffisante. Par exemple, dans les dossiers d'agressions sexuelles, il est fréquent que les pièces à conviction soient des prélèvements intimes de la victime et dans ces cas, un mélange ADN du suspect et de la victime est souvent récupéré. L'ADN de la victime se retrouvant en plus forte proportion, il devient difficile de détecter celui du suspect. C'est alors que les marqueurs situés sur le chromosome Y sont utilisés, puis que ce dernier n'étant retrouvé que chez les hommes, ces marqueurs permettent d'isoler l'ADN suspect. Lorsqu'une concordance survient entre le profil ADN d'un suspect et celui retrouvé sur une pièce à conviction, un poids statistique est attribué à celle-ci, ce qui permet de guider les experts du système de justice pour déterminer la culpabilité d'un suspect. Toutefois, la détermination de cette valeur pour les profils ADN obtenus par l'analyse du chromosome Y (haplotype Y) s'avère encore problématique.

D'abord, la majorité des méthodes actuelles reposent sur l'utilisation et la disponibilité de bases de données de référence. Plusieurs critiques ont été faites sur leur utilisation pour attribuer une valeur probante, puisqu'elles ne sont pas constituées d'échantillons aléatoires dans la population et sont souvent de taille trop restreinte pour être représentative de l'ensemble des haplotypes Y retrouvés dans la population. De plus, le chromosome Y se transmet, majoritairement sans recombinaison, de père en fils laissant comme seule variation possible les mutations. Ce mode de transmission rend le chromosome Y sensible à des effets génétiques et démographiques tel que la dérive génétique et l'effet fondateur. Ces éléments combinés font que le calcul d'un poids statistique pour ces profils ADN est plus difficile et qu'il est important, et toujours d'actualité de développer de nouvelles avenues pour contrer ces problèmes.

À ce sujet, deux chercheurs, Andersen et Balding, ont récemment proposé une nouvelle méthode permettant de contre carrer ces problématiques. Cette méthode consiste en un modèle qui

permet de simuler une population d'hommes, puis d'attribuer des profils ADN Y aux individus. Il est alors possible de calculer la distribution d'hommes portant un haplotype Y commun dans la population dite vivante, représentant les trois dernières générations de la population simulée. Bien que prometteuse, la méthode proposée par Andersen et Balding nécessitait une validation à grande échelle. La fine connaissance de la généalogie du Québec, depuis sa fondation en 1608, nous permettait de tester le modèle à l'échelle d'une population entière.

Notre premier objectif était donc de valider à grande échelle le modèle proposé. Pour ce faire, nous avons utilisé le modèle tel que proposé, c'est-à-dire en utilisant les paramètres de la population du Québec actuelle pour tenter de simuler une population s'y rapprochant. Des haplotypes Y ont par la suite été attribués aux individus formant la population simulée. Pour évaluer dans quelle mesure ces simulations se rapprochaient de ce qui est réellement observé dans notre population, nous avons également effectué des simulations en utilisant la généalogie connue. En comparant les distributions d'hommes portant un même haplotype, il nous a été possible de mettre en évidence que la population canadienne-française possède un fort effet fondateur, ce qui rend la détermination des fréquences d'haplotypes Y complexe. En effet, nous avons déterminé, en analysant les distributions d'hommes portant un même haplotype Y que 99% des haplotypes Y s'est avéré être portés par environ 4 fois plus d'hommes en utilisant les données généalogiques comparativement à la population simulée. Nos résultats ont également permis de déterminer que le modèle de simulations est un excellent outil pour estimer la fréquence des haplotypes rares dans la population, alors que pour les haplotypes plus fréquents, le modèle devrait être utilisé comme guide pour déterminer la taille d'un échantillonnage dans la population qui permettrait d'estimer adéquatement la fréquence des haplotypes Y les plus fréquents.

Le deuxième objectif du projet consistait à évaluer la variation spatiale dans la répartition des haplotypes Y dans le Québec. En effet, il est plutôt bien rapporté dans la littérature que la population du Québec possède une structure particulière en raison des mouvements migrateurs à l'origine de sa fondation. Il est connu que plusieurs petits groupes d'individus issus des fondateurs de la Nouvelle-France se sont déplacés pour coloniser, par vagues, les régions du Québec. Les individus établis dans une région se mariaient plus rarement avec des gens d'une autre région, créant une structure génétique entre celles-ci. Ces éléments ont ainsi donc pu créer une hétérogénéité spatiale dans la répartition des haplotypes Y dans la population. Nos résultats ont démontré qu'effectivement certains profils ADN

Y sont uniquement retrouvés dans une région spécifique. Les régions où la concentration des haplotypes est la plus élevée sont la Mauricie et l'Estrie où un peu plus de 20% des haplotypes retrouvés dans ces régions sont uniquement portés par des hommes vivants dans cette dite région. Cet élément est un facteur clé qui doit être considéré par les experts judiciaires lors de l'attribution d'une valeur probante suite à une concordance entre deux haplotypes Y.

En conclusion, nos résultats ont permis de mettre en évidence les divers avantages et limitations de la méthode de simulations proposées tout en évoquant les difficultés rattachées à la population canadienne-française pour l'attribution d'une valeur probante à la suite d'une concordance entre deux haplotypes Y permettant ainsi de mieux guider les experts judiciaires dans les méthodes à utiliser avec la population du Québec.