



# **Manuel de référence L'ADN généalogique**

Par Denis Beauregard  
SGCF 5753  
Février 2024



**Manuel de référence**

**L'ADN généalogique**

**Par Denis Beauregard**

Manuel de référence  
L'ADN généalogique  
Par Denis Beaugard

isbn : 978-2-922422-23-8  
Copyrights 2024 Denis Beaugard

## *Table des matières*

---

1	Introduction.....	10
2	Principes de base.....	13
2.1	Haplogroupes et haplotypes.....	13
2.2	Les principaux laboratoires.....	14
2.3	Valider sa généalogie.....	15
2.4	Débloquer une généalogie.....	16
2.5	Valider une origine amérindienne.....	16
2.6	La confidentialité.....	17
2.7	La situation en France.....	18
3	Techniques.....	19
3.1	Quatre bases et des données omises.....	19
3.2	Trois sortes de tests génétiques.....	20
3.3	Haplogroupes et haplotypes.....	21
3.4	Optimiser ses résultats.....	21
4	Le chromosome Y (test STR).....	22
4.1	Aperçu.....	22
4.2	But.....	23
4.3	Les triangulations.....	23
4.4	Le test STR de FTDNA.....	24
4.4.1	Haplogroupe et haplotype.....	24
4.4.2	Nombre de marqueurs.....	25
4.5	Exemples de nombre de résultats.....	28
4.5.1	Familles avec le plus de descendants.....	28
4.5.2	Signature de référence.....	30
4.5.3	Parents inconnus.....	31
4.5.4	Les événements non parentaux (ÉNP).....	32
4.5.5	Les sortes d'événements non parentaux.....	33
4.5.5.1	Deux brisures flottantes.....	33
4.5.5.2	Brisure délimitée.....	33
4.5.5.3	Brisure flottante.....	34
4.5.5.4	Brisure précise.....	35
4.5.5.5	Brisure double.....	35
4.5.5.6	Adoptions et parents inconnus.....	35
4.6	L'affichage de résultats STR.....	36
4.6.1	Autres données du client.....	36
4.6.2	Résultats STR des autres clients d'un projet.....	37
4.7	Des cas spéciaux.....	38
4.8	Les autres laboratoires.....	39
4.9	Les sites de comparaison.....	40

5	Le chromosome Y (test SNP).....	41
5.1	Aperçu.....	41
5.2	La séquence des SNP menant vers les Beauregard.....	42
5.3	Le test Big Y.....	43
5.4	Description des SNP Y.....	44
5.5	Les noms des SNP.....	45
5.6	Les résultat d'un Big Y.....	46
5.6.1	Le Block Tree.....	47
5.6.2	Les rapports Discover.....	49
5.6.3	Les matches Big Y.....	51
5.6.4	Les <i>results</i> .....	51
5.6.5	Les résultats STR.....	51
5.7	Les SNP à la pièce.....	51
5.8	Les autres laboratoires (génom complet).....	52
5.8.1	L'analyse du chromosome Y.....	52
5.8.2	L'analyse du génome complet.....	53
5.8.3	Les fichiers de résultats.....	53
5.8.4	L'analyse par YFULL.....	53
5.9	Les autres laboratoires (SNP de 4000 ans).....	54
6	Le chromosome MT (lignée maternelle).....	56
6.1	Aperçu.....	56
6.2	La signature MT.....	56
6.3	Les haplogroupes MT.....	57
6.4	Les haplotypes MT.....	58
6.4.1	L'hétéroplasmie.....	59
6.5	Haplogroupes de quelques pionnières.....	60
6.6	Les différents tests MT.....	60
6.7	La recherche de cousins MT.....	62
6.7.1	Confirmer la lignée menant vers un ancêtre (triangulation).....	62
6.7.2	Détecter une erreur dans la documentation.....	62
6.7.3	Débloquer une filiation.....	63
6.8	L'ADNmt de Catherine Pillat.....	64
6.9	L'ADNmt des sœurs Langlois.....	66
6.10	Les sœurs Aucoin et l'hétéroplasmie.....	67
7	L'ADN autosomal.....	69
7.1	Aperçu.....	69
7.2	Les 23 paires de chromosomes.....	69
7.3	La combinaison des chromosomes.....	69
7.4	Le test ADN autosomal.....	70
7.5	La recherche de cousins.....	73
7.6	Le type de cousins trouvés par ADN.....	74
7.7	Confirmer un cousinage avec l'ADN.....	76
7.7.1	Comparer des frères et sœurs.....	76
7.7.2	Comparer des cousins.....	77
7.7.3	Confirmer un grand-parent.....	78

7.8	Résoudre une énigme généalogique avec l'ADN.....	78
7.8.1	Recherche des parents biologiques.....	78
7.8.2	La triangulation.....	80
7.8.3	Recherche des grands-parents biologiques.....	80
7.8.4	Contacteur un cousin biologique.....	81
7.9	L'ADN longue durée des zones d'empilement.....	81
7.10	Les origines ethniques.....	83
7.11	Le génome complet pour un usage autosomal.....	84
7.11.1	Peut-on aller plus loin?.....	85
7.12	L'information médicale.....	85
8	Le laboratoire FTDNA.....	87
8.1	Introduction.....	87
8.2	Les tests disponibles.....	87
8.2.1	Tests ADN Y.....	87
8.2.2	Tests ADNmt.....	88
8.2.3	Test Family Finder.....	89
8.3	Visite guidée.....	89
8.3.1	Panneau d'entrée ( <i>Dashboard</i> ).....	89
8.3.2	La section [ <i>Y-DNA Results &amp; Tools</i> ].....	91
8.3.3	La section [ <i>Big Y</i> ].....	93
8.3.4	La section [ <i>mtDNA Results &amp; Tools</i> ].....	95
8.3.5	La section [ <i>Autosomal DNA Results &amp; Tools</i> ].....	96
8.4	Les projets.....	98
8.4.1	Résultats publics.....	101
8.4.2	Exemple de résultats d'ADN Y :.....	102
8.4.3	Exemple de résultats d'ADNmt :.....	102
8.4.4	Les numéros de trousse.....	103
8.5	Exporter et importer des données.....	104
8.5.1	Exporter des données.....	104
8.5.1.1	Données Family Finder.....	104
8.5.1.2	Données ADNmt.....	104
8.5.1.3	Données ADN Y.....	105
8.5.2	Importer des données.....	105
8.6	Versions.....	106
9	Le laboratoire 23andme.....	107
9.1	Introduction.....	107
9.2	Les tests disponibles.....	107
9.2.1	Le test généalogique.....	107
9.2.2	Le test médical.....	108
9.3	Visite guidée.....	108
9.3.1	Panneau d'entrée.....	108
9.3.2	Ancestry Report.....	108
9.3.3	DNA Relatives.....	109
9.3.4	Liste des cousins.....	111
9.3.5	Arbre en ligne (lien « <i>Family Tree</i> ».).....	113

9.3.6	Section <i>Ancestry &amp; Traits</i> .....	113
9.4	Utilisation pour la généalogie.....	114
9.4.1	Les haplogroupes Y.....	114
9.4.2	Les haplogroupes MT.....	114
9.4.3	Les autres lignées.....	115
9.4.4	Débloquer la lignée d'un adopté.....	116
9.5	Exporter et importer des données.....	116
9.5.1	Exporter des données.....	116
9.5.2	Importer des données.....	116
9.6	Versions.....	116
10	Le laboratoire AncestryDNA.....	118
10.1	Introduction.....	118
10.2	Le test disponible.....	118
10.2.1	Le test généalogique.....	118
10.3	Visite guidée.....	118
10.3.1	Panneau d'entrée.....	118
10.3.2	ThruLines.....	119
10.4	Utilisation pour la généalogie.....	120
10.4.1	Les cousins trouvés.....	121
10.4.2	Les haplogroupes Y.....	121
10.4.3	Les haplogroupes MT.....	121
10.4.4	Les autres lignées.....	121
10.4.5	Débloquer la lignée d'un adopté.....	122
10.5	Exporter et importer des données.....	122
10.5.1	Exporter des données.....	122
10.5.2	Importer des données.....	123
10.6	Versions.....	123
11	Les autres laboratoires.....	124
11.1	myHeritage.....	124
11.1.1	Exporter et importer des données.....	125
11.1.1.1	Exporter des données.....	125
11.1.1.2	Importer des données.....	125
11.2	LivingDNA.....	125
11.3	Yseq.....	125
11.4	Les génomes complets.....	125
11.5	Le site YFULL.....	126
11.5.1	Importer des données.....	126
11.6	Gedmatch.....	127
11.6.1	Importer des données.....	127
11.7	Geneanet.....	127
11.8	Comparaison.....	127
11.9	Pour en savoir plus.....	129
12	Vocabulaire.....	130
12.1	Termes francophones.....	130
12.2	Termes anglophones.....	132



# 1 Introduction

---

L'ADN (acide désoxyribonucléique) est le dictionnaire à partir duquel tous les organismes vivants et virus se construisent. Il est maintenant utilisé en généalogie à des fins variées, comme confirmer des relations entre des cousins ou débloquent des filiations que la documentation traditionnelle n'a pu conserver.

À ces fins, le client dépose un peu de salive dans une éprouvette qu'il envoie à un laboratoire spécialisé. Par la suite, une analyse est faite à partir de cet échantillon et les résultats sont comparés avec ceux des autres clients du même laboratoire.

Finalement, le client reçoit ses résultats sous plusieurs formes.

- La liste des correspondants (aussi appelés *matches* ou *relatives*) est le principal volet et souvent le seul que l'utilisateur pourra utiliser.
- Les résultats bruts permettront l'exportation vers un autre laboratoire ou vers un site de comparaison afin d'augmenter l'utilité du test, parfois afin de trouver plus d'informations, parfois pour découvrir d'autres cousins éloignés.
- Les origines ethniques sont pour certains la principale raison pour faire ces tests. Toutefois, l'ADN n'étant pas distribué uniformément de génération en génération (sinon, tout le monde serait le sosie de ses frères et sœurs), ce volet est le moins utile et le plus perturbateur. Ces résultats doivent être reçus avec légèreté et ne pas être pris comme exacts.
- L'aspect médical est en général absent. Certains laboratoires ont même éliminé les valeurs qui seraient utilisables pour détecter certaines prédispositions. D'autres laboratoires en font tout de même leur spécialité.

Il existe plusieurs sortes de tests ADN pour la généalogie et nous les verrons dans des sections séparées.

**L'ADN Y** se trouve sur le chromosome Y, que seuls les hommes possèdent. Il est transmis par le père à ses fils avec parfois une variation et en comparant ces variations, on peut parfois déterminer si deux hommes ont un ancêtre commun dans des temps historiques.

**L'ADN X** suit un parcours mixte et certains laboratoires ne l'utilisent pas pour les comparaisons ou appliquent des règles spéciales. Avec le chromosome Y, il constitue la paire des chromosomes sexuels.

## Introduction

L'**ADNmt** se trouve dans les mitochondries, des organelles que tout le monde possède, mais qui ne sont transmises que par les mères à leurs enfants. <sup>1</sup>

L'**ADNau** (ou **autosomal**) est constitué par les 22 paires de chromosomes autres que le X et le Y. Il est reçu de l'ensemble de nos ancêtres et son origine diminue de moitié à chaque génération, mais dès qu'on dépasse les parents, l'apport des ancêtres commence à varier.

L'analyse ADN se fait avec différentes méthodes. Autrefois, on se contentait de détecter et mesurer la longueur des séquences répétitives des nucléotides A, C, G et T (appelées **STR**). Par la suite, une fois le génome humain analysé, on a pu détecter des valeurs situées à des endroits précis (**SNP**). Des technologies plus récentes (**WGS**) permettent maintenant la lecture de séquences de 100 à 400 valeurs de suite, puis de les situer dans l'ensemble du génome humain (qui a été cartographié grâce à des milliards de dollars investis par le gouvernement américain). Cette cartographie n'est toutefois pas terminée car le chromosome Y contient de longues séquences répétées que la technologie de 2<sup>e</sup> génération ne permettait pas de placer, ce qui fait qu'en 2023, sa longueur est passé de 59 à 62 millions de positions, lorsque le séquençage de plus longs segments (3<sup>e</sup> génération) a permis d'améliorer la référence qui passait de Hg38 à T2T.

Ce document contient de nombreuses références au laboratoire FTDNA ou FamilyTreeDNA<sup>2</sup> qui est la principale source des données génétiques dans les travaux de l'auteur. En fait, c'est, et de loin, le meilleur laboratoire pour les tests ADN uniparentaux qui suivent les lignées paternelles ou maternelles. Les autres laboratoires seront aussi mentionnés.

Beaucoup de statistiques viennent du projet Héritage français (ou *French Heritage*)<sup>3</sup> hébergé chez FTDNA. En 2023, il a dépassé la barre des 11 000 membres. Fondé vers 2004, il est le plus gros regroupement de données génétiques sur la Nouvelle-France et probablement sur la France.

Plusieurs exemples sont tirés du catalogue<sup>4</sup> de signatures ADN ancestrales hébergé sur le site Francogene. Celui-ci comprend en page d'accueil les nouveautés, puis séparément les signatures de référence Y et MT par ordre alphabétique des immigrants (Nouvelle-France) ou lorsque l'information est disponible, l'ancêtre le plus ancien connu

---

<sup>1</sup>Les mitochondries du père sont considérées comme perdues, sauf en de très rares cas pathologiques. Parmi les milliers de cas évalués par l'auteur, il n'existe aucun cas d'ADNmt paternel et lorsque les résultats ne sont pas ceux prévus, c'est qu'il y a une erreur dans la documentation ou qu'il y a une adoption récente ou une hypothèse qu'on finit par corriger.

<sup>2</sup> <https://www.familytreedna.com>

<sup>3</sup> <https://www.familytreedna.com/groups/frenchheritage/about>

<sup>4</sup> <http://www.francogene.com/triangulation/>

## Introduction

en Europe. Comme les signatures ancestrales sont validées par la méthode de triangulation, les exemples seront identifiés TRI0001 pour la triangulation numéro 1.

Le lecteur trouvera un grand nombre d'informations sur les laboratoires qui analysent ADN<sup>5</sup> et sur les outils disponibles<sup>6</sup> sur le site de l'ISOGG (*International Society of Genetic Genealogy*), un organisme de coordination et d'information.

---

---

<sup>5</sup> [https://isogg.org/wiki/List\\_of\\_DNA\\_testing\\_companies](https://isogg.org/wiki/List_of_DNA_testing_companies)

<sup>6</sup> [https://isogg.org/wiki/Autosomal\\_DNA\\_tools](https://isogg.org/wiki/Autosomal_DNA_tools) et autres

## 2 Principes de base

---

Il existe plusieurs méthodes pour analyser l'ADN, plusieurs sortes de résultats, plusieurs positions analysées à l'intérieur des chromosomes et plusieurs bases de données avec un contenu variable (clients et résultats).

Pour le néophyte, c'est souvent à partir de son budget, mais surtout de ses besoins, que le laboratoire sera choisi. Il est toutefois important de savoir qu'en 2023, les plus grosses bases de données sont américaines et donc que l'on trouvera souvent des cousins nord-américains. De plus, les données analysées ne sont pas les mêmes selon le laboratoire. Enfin, certains laboratoires ne permettent pas l'importation des résultats obtenus auprès des autres laboratoires alors que l'exportation de leurs résultats est possible.

Dans ce chapitre, vous verrez les principes de base qui seront développés plus loin,

### 2.1 Haplogroupes et haplotypes

---

Les haplogroupes servent à regrouper des résultats de tests génétiques d'ADN Y ou MT. Les haplotypes sont des données plus individuelles.

L'haplogroupe permet de comparer rapidement des résultats pour savoir s'il y a un air de famille ou pas, un peu comme des patronymes qui se ressemblent ou pas. Ainsi, les Beauregard, Tousignant et Pagé partagent l'haplogroupe R-A431. Si le chromosome Y d'un homme portant un de ces patronymes est dans l'haplogroupe G-M201 ou I-M253, par exemple, on voit tout de suite qu'il n'y a pas de cohérence dans les résultats et donc qu'il n'appartient pas à une de ces familles du point de vue biologique. Donc, la lettre initiale doit être la même.

Avec le temps, les mutations s'enchaînent. Les haplogroupes reflètent les étapes dans ces mutations dans le temps. Ainsi, si on est R-L151 ou R-DF27, le résultat est cohérent car il s'agit d'étapes menant vers R-A431, mais ces haplogroupes désignent aussi un grand nombre d'autres familles. Un haplogroupe Y peut donc être plus ou moins précis selon son ancienneté. Un haplogroupe plus ancien couvre beaucoup plus de familles alors qu'un haplogroupe terminal ou apical serait le plus récent dans la succession des mutations vers la personne testée.

Nous verrons plus loin que l'haplogroupe et l'haplotype permettent de catégoriser une signature ADN Y ou MT afin de comparer les résultats de descendants de différentes familles.

## 2.2 Les principaux laboratoires

---

En 2023, les principaux laboratoires sont les suivants (voir la page<sup>7</sup> de l'ISOGG pour les adresses et un résumé).

- **Ancestry** qui a le plus grand nombre de clients (ADN autosomal), mais un test ADN incomplet;
- **23andme** qui a un volet médical (avec un coût supplémentaire) et qui révèle des haplogroupes Y et MT (une information partielle sur les lignées paternelle et maternelle);
- **myHeritage**, un site web généalogique dont les tests ADN sont effectués par FTDNA;
- **FTDNA** ou **FamilyTreeDNA**, le pionnier, avec la plus grosse base de données de résultats Y et MT, un test autosomal propre et, en plus, la possibilité d'importer les résultats obtenus des autres laboratoires, ainsi qu'un arbre phylogénétique qui est une sorte d'arbre généalogique de l'ADN Y et MT;
- **YSEQ** offre plusieurs tests et construit peu à peu sa propre base de comparaison;
- Quelques laboratoires offrent le séquençage du génome complet (**WGS**) ou séquençage par segments de 100 à 400 valeurs contiguës comme **Dante**, **FGC**, **Nebula**, **Yseq**, etc. N'ayant pas leur propre base de comparaison, leurs clients peuvent utiliser YFULL. Les tests de 3<sup>e</sup> génération sont encore trop coûteux pour le consommateur moyen et ne seront pas discutés ici. Ils pourraient permettre la lecture d'un million de valeurs contiguës.
- **Gedmatch**, acquis par la firme Verogen en 2019, n'est pas un laboratoire mais offre à la fois l'importation de données et plusieurs outils intéressants;
- **Geneanet**, site français acquis par Ancestry en 2021, permettait également d'importer des données ADN jusqu'à la fin de 2023;
- **YFULL** est un site web qui reçoit des données ADN brutes et fait la comparaison entre les résultats afin de construire un autre arbre phylogénétique dont les résultats ne proviennent d'un seul laboratoire. **YDNA-Warehouse** avait également un arbre en ligne mais semble maintenant abandonné. Certains sites offrent aussi leur propre analyse de résultats à un niveau moindre, parfois en se concentrant sur une région ou une grande famille d'haplogroupes.

---

<sup>7</sup> [https://isogg.org/wiki/List\\_of\\_DNA\\_testing\\_companies](https://isogg.org/wiki/List_of_DNA_testing_companies)

Certains laboratoires offrent des tests de paternité ou une exploration des origines en se basant sur une technologie désuète. S'il n'y a pas de base de comparaison, il est préférable de ne pas les utiliser à des fins de généalogie. De plus, des laboratoires ont livré des faux résultats, avec par exemple une proportion de 20% d'ADN amérindien à tous leurs clients, y compris un journaliste africain.

### 2.3 Valider sa généalogie

---

Deux approches sont disponibles : vérifier les cousinages et comparer la signature Y ou MT d'ancêtres lointains.

Les cousinages sont évalués grâce à l'ADN autosomal partagé avec d'autres personnes qui ont fait le même test autosomal. Il faut que les tests comparés se trouvent dans la même base de données de comparaison. Si vous voulez comparer l'ADN de deux personnes (qui seront donc testées), il est préférable de choisir le même laboratoire. La relation trouvée est en réalité une plage possible et non une valeur exacte parce que l'ADN n'est pas distribué uniformément.

Il est aussi possible de comparer les signatures Y et MT avec d'autres personnes ayant fait un test ADN similaire. Un catalogue permet d'avoir des signatures de référence concernant surtout la Nouvelle-France, soit l'actuel Québec, l'Acadie et la Louisiane. Pour les autres régions, la signature Y sera souvent comparable via un test sur l'ADN Y et en examinant les noms de famille des correspondants, lorsque le nom des autres clients dont l'ADN Y est similaires. Le meilleur laboratoire pour ce test est FTDNA qui donnera comme résultats le nom des autres clients dont l'ADN Y est similaire et, parfois, l'ancêtre connu le plus lointain. Le laboratoire 23andme permet de comparer l'ADN Y avec moins de précision (on trouve l'ADN Y d'ancêtres ayant vécu il y a plus de 4000 ans, donc un grand nombre de familles possibles). D'autres laboratoires contiennent aussi cette information mais il faut utiliser un outil extérieur pour le savoir. Il y a aussi des tests sur l'ADN Y sans base de comparaison, ce qui implique qu'il faudra alors exporter les résultats vers un laboratoire de comparaison.

Dans notre culture, les noms de famille des femmes changent à chaque génération. Pour avoir une bonne quantité de signatures MT de comparaison, il faut avoir une masse critique, soit un groupe géographique avec un nombre élevé de tests ainsi qu'une disponibilité de données généalogiques. C'est le cas pour la Nouvelle-France. Ailleurs, il pourrait y avoir des projets visant l'époque coloniale (comme les passagers du Mayflower) mais cela ne fait pas l'objet de la présente publication.

## 2.4 Débloquer une généalogie

---

Si l'ancêtre dont on veut découvrir les parents est à 5 générations ou moins, un test ADN autosomal peut aider. Le travail consiste en général à documenter toutes les correspondances trouvées et à regrouper les descendants.

S'il s'agit des parents ou grands-parents biologiques, il est probable que les correspondants les plus proches, sauf un gros coup de chance, soient éloignés. Il faudra donc recouper beaucoup d'informations pour avoir des résultats dans la plupart des cas.

Si c'est un ancêtre en ligne seulement masculine, il est parfois possible de trouver le nom de famille à partir des correspondances avec un test ADN Y de FTDNA, à condition d'avoir une bonne quantité de tests faits dans la population cible. Ici, la personne testée doit être un homme.

En lignée féminine, on trouve 2 cas principaux:

- Si la personne à trouver a un nom de famille et qu'il manque ses parents parce que l'acte de mariage est perdu, on pourra regarder parmi les correspondances trouvées s'il y a un couple vivant à la bonne époque et ayant le bon nom de famille.
- Sans nom de famille ou avec un nom de famille obtenue à l'adoption, la chance devient un facteur important et on arrive souvent à plusieurs parents possibles. Par contre, si la personne recherchée est localisée directement dans la lignée maternelle, un test en ADN mitochondrial peut aider.

## 2.5 Valider une origine amérindienne

---

L'ADN Y et l'ADNmt sont des tests à longue portée. Ils suivent les lignées uniquement paternelle ou maternelle et peuvent donner un résultat identifiable comme amérindien, ce qui permet de confirmer un ancêtre (parmi tous les ancêtres identifiables), jusqu'à plus de 50 générations, bien avant l'arrivée des Européens en Amérique. Ceci est basé sur les haplogroupes (lignée uniquement paternelle ou maternelle) ainsi que sur les correspondances Y ou MT (à cause de l'ADN provenant des invasions asiatiques en Europe, une personne pourrait avoir de l'ADN asiatique sans qu'il s'agisse d'ADN amérindien et les correspondances enlèvent tout doute de ce côté, pourvu que le test soit assez précis).

S'il est impossible d'atteindre la personne visée sans mélanger les pères et les mères, l'ADN n'étant pas distribué uniformément et la partie amérindienne pouvant devenir trop faible pour être détectée, on ne peut généralement confirmer une origine amérindienne qu'avec un nombre limité de générations entre la personne testée et l'autochtone. Un résultat typique serait, en moyenne, l'ancêtre détecté à environ 7 générations (avec 1%

d'ADN) ou avec une trace (avec 0,1% d'ADN) à 10 générations. La trace pourrait être asiatique (nord ou centre) par exemple, ou même d'Europe de l'Est (Hongrie ou Finlande). Il s'agit ici d'une question de populations de référence utilisées par les laboratoires.

Ceci s'applique aussi à des origines africaines subsahariennes (ce qu'on appelle l'Afrique noire), mais pas au bassin méditerranéen où le brassage de population était trop important.

### 2.6 La confidentialité

---

Un test ADN à des fins généalogiques est par définition une comparaison des résultats de plusieurs clients. Par exemple, on peut confirmer la signature ADN Y parce qu'on a des résultats cohérents de deux hommes et leur lignée documentée, ou exceptionnellement, si ces deux hommes descendent d'une même famille par des branches identifiées par des noms différents. Il y a donc, au départ, des données confidentielles qui doivent être partagées si l'on veut que le test ADN soit utile.

Avec le test autosomal, ce concept s'étend à la parenté proche. Les adoptés peuvent souvent identifier des ancêtres biologiques, parfois les parents, parce que des cousins de ces parents biologiques ont fait un test ADN. La même remarque s'applique à la recherche de criminels ou l'identification de victimes par la généalogie génétique. Il est impossible d'isoler l'ADN d'une personne de celui de sa parenté proche, ce qui oblige les personnes testées à accepter d'un côté que leur ADN soit visible par d'autres, du moins par les segments partagés, et de l'autre, que cela s'étende à leur parenté.

Outre l'aspect génétique, la documentation saisie par le client peut parfois servir à l'identifier. Une façon de diminuer la dérive possible est sans doute d'inscrire soi-même les données accessibles aux correspondances pour qu'il n'y ait pas d'enquête généalogique pour compléter une lignée. Une autre manière est d'utiliser les restrictions disponibles chez les grands laboratoires, mais en même temps, cela enlève une partie de l'utilité du résultat s'il est fait dans un but généalogique.

En général, il faut accepter de participer au système de cousinage (*matching*) des laboratoires. Avec FTDNA, il faut signer un formulaire de consentement et lire la longue liste de ce que les autres peuvent faire. Avec 23andme, il faut également activer cette fonction pour voir les cousins, sinon les résultats sont limités aux parties médicales et ethniques.

## 2.7 La situation en France

---

La loi française interdit la prise d'empreintes génétiques ou l'examen des caractéristiques génétiques à des fins autres que médicales ou scientifique. Il s'en suit que les laboratoires les plus intéressants ne livrent pas à une adresse en France.

- **Ancestry** n'a jamais livré en France.
- **23andme** a expédié des tests jusqu'à l'automne 2022.
- **myHeritage** a permis leur envoi entre 2018 et le 1<sup>er</sup> janvier 2023, tout en faisant de la publicité durant cette période.
- **FTDNA** a mis fin à ses livraisons le 1<sup>er</sup> février 2023 et demande un numéro de téléphone.
- Les autres laboratoires ne semblent pas affectés. Certains ont rapporté que les douanes ont intercepté des colis concernant l'ADN généalogique à des fins de loisir, mais ceci semble anecdotique et il est possible d'expédier une trousse ADN en utilisant une adresse intermédiaire. Il est préférable que le contenu du colis ne soit pas mis en valeur. Certains suggèrent un achat par un site Amazon en dehors de la France (qui n'offre pas de test ADN) mais européen (afin de livrer sans passer par des douanes).
- En mai 2023, l'article 16-10 de la loi française aurait été modifié mais ceci n'a pu être vérifié sur le site de Legifrance. Toutefois, depuis la fin de 2022, les trousse d'analyse ADN sont interceptées aux douanes françaises et retournées à l'expéditeur. Un Français doit plutôt examiner régulièrement ce qu'offrent les sites Amazon en Europe (sauf en Suisse) pour commander un test ADN qui ne franchirait pas les douanes. Le cas éventuel, il faut aussi éviter de donner un numéro de téléphone en France et éventuellement utiliser les versions non-françaises des sites web des laboratoires.

## 3 Techniques

---

Il existe des méthodes variées pour analyser l'ADN. Ce chapitre est plus technique et se destine aux amateurs de technologies même si le fonctionnement n'est pas détaillé.

Il existe 3 méthodes pour analyser l'ADN et elles sont toutes basées sur des tests chimiques pour déterminer les séquences des bases de l'ADN analysé.

### 3.1 Quatre bases et des données omises

---

L'ADN humain se compose de 47 chromosomes (dont 46 chromosomes nucléaires) qui sont autant de molécules. Cet ensemble comprend 22 paires de chromosomes autosomes (ADN<sub>au</sub>), 2 chromosomes sexuels (ADN X et ADN Y) et le chromosome des mitochondries<sup>8</sup> (ADN<sub>mt</sub> ou génome mitochondrial humain). Les 23 chromosomes paternels viennent du père et les 23 maternels, de la mère. L'homme possède des chromosomes X et Y alors que la femme a deux X. Les mitochondries paternelles sont considérées comme perdues et il ne reste donc, pour l'usage généalogique, que les mitochondries maternelles.

Ces chromosomes sont des séquences de longueurs variables comprenant 4 bases (ou paires de bases) et des atomes pour les retenir ensemble. La lecture des données brutes d'un test ADN inclura donc les lettres A, C, G et T (adénine, cytosine, guanine et thymine), ainsi qu'un tiret si la valeur n'a pas été lue. Dans ce dernier cas, on utilise souvent l'expression anglaise *no-call*.

On ne lit pas les chromosomes au complet, mais des parties qu'on assemble ou dont on identifie la position à l'intérieur du génome de référence. Il s'en suit deux effets qu'il est utile de connaître.

- Un test ADN ne lit pas toutes les valeurs recherchées. Ainsi, pour un même laboratoire, la longueur du fichier de données brutes peut varier. Une absence d'environ 2% des résultats est suffisante pour que des jumeaux identiques aient des résultats différents dans ce qu'on appelle les origines ethniques (bien qu'il ne s'agisse ni d'origines, ni d'ethnies).
- Les données brutes ne sont pas associées au chromosome paternel ou maternel. Dans le fichier des données brutes autosomales, pour chaque position, on trouve souvent 2 valeurs dont l'ordre est souvent alphabétique, éventuellement l'ordre

---

<sup>8</sup> Le génome des mitochondries forme un chromosome, bien que certains réservent ce terme aux 46 chromosomes nucléaires. Voir par exemple <https://www.genome.gov/genetics-glossary/Mitochondrial-DNA>

dans lequel les valeurs ont été lues. Il est toutefois possible, en comparant avec des populations semblables, d'identifier les parties paternelles et maternelles.

### 3.2 Trois sortes de tests génétiques

---

Les trois principaux types de tests génétiques ou de mesures génétiques sont appelés STR, SNP et WGS.

Le test **STR** (pour *Short Tandem Repeat*) consiste à rechercher des séquences répétées et à compter le nombre de répétitions. C'est le vieux test utilisé par la police pour identifier des personnes sur une scène de crime (avec le codis). Ce type de test est également utilisé pour comparer deux personnes et évaluer leur degré de parenté, mais le test SNP décrit plus bas est beaucoup plus précis.

Ce test STR est également utilisé pour caractériser le chromosome Y en mesurant un certain nombre de chaînes appelées marqueurs. Plus le nombre de marqueurs est similaire, et plus l'ADN est similaire, ce qui ne signifie pas que la parenté soit plus proche.

Le test **SNP** (*Single Nucleotide Polymorphism*) se base sur la recherche de certaines séquences et retient la valeur trouvée après cette séquence. Ces séquences sont suffisamment longues pour qu'il n'y ait pas de confusion. Une puce d'analyse ADN est programmée pour détecter un certain nombre de séquences trouvées sur les 22 paires de chromosomes, l'ADN X, l'ADN Y et l'ADNmt. Le choix des valeurs lues dépend du laboratoire et certains ont changé plusieurs fois leur liste de positions lues.

Le test **WGS** (*Whole Genome Sequencing*) consiste à lire environ 100 valeurs de suite ou plus, puis à tenter de trouver leur position dans l'ensemble de l'ADN (plus de 3 milliards de positions). Le fichier obtenu, appelé BAM, est gigantesque car il contient plusieurs fois chaque lecture. Un test "30 X" veut dire qu'on tente de lire 30 fois chaque valeur, ou en fait, que le fichier de résultats contient 90 milliards de valeurs ou 30 fois la séquence de référence. Certaines compagnies peuvent lire 400 valeurs, mais les résultats comportent plus d'erreurs. Des technologies en développement permettraient d'atteindre le million de valeurs.

Un test de 1<sup>re</sup> génération (séquençage de **Sanger**) a demandé 10 ans pour décomposer le génome humain complet pour obtenir le génome de référence.

Un test de 2<sup>e</sup> génération (**NGS** ou **Next Generation Sequencing**) décode le génome complet en une journée.<sup>9</sup>

La 3<sup>e</sup> génération est en train de faire son apparition.

---

<sup>9</sup> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3841808/>

Le test de type **WGS** permet de séquencer le génome humain au complet (dans un cadre médical, par exemple) ou seulement le chromosome Y afin d'en faire une étude plus approfondie.

Ces tests ne sont pas parfaits. La région située au centre des chromosomes, appelée centromère, est plus difficile à lire. De plus, certaines sections à incidence médicale sont volontairement omises par certains laboratoires.

### 3.3 Haplogroupes et haplotypes

---

Les haplogroupes servent à regrouper des résultats de tests génétiques d'ADN Y ou MT. Les haplotypes sont des données plus élémentaires. Nous verrons plus loin les protocoles utilisés selon que le Y ou le MT doit être regroupé.

Les haplotypes sont les détails, les mesures individuelles. Ainsi, l'ADN Y peut se résumer à une étiquette courte comme R1b ou R-M269 ou même une combinaison comme R1b-M269. L'haplotype comprendrait alors les valeurs STR lues pour prédire cet haplogroupe.

La recherche de correspondances parmi les autres clients d'un laboratoire comme FTDNA faisant des tests sur les chromosomes Y utilise les haplotypes pour retrouver des résultats similaires mais pour l'ADNmt, le processus de recherche est un peu différent comme nous le verrons plus loin.

### 3.4 Optimiser ses résultats

---

La plupart des tests se font avec de la salive, sauf ceux de FTDNA et myHeritage qui utilisent le grattage intérieur de la joue. Dans tous les cas, il faut avoir conscience de la présence de l'ADN dans nos aliments et il faut donc réduire leur présence. Pour cela, il vaut mieux ne rien manger durant l'heure précédant la prise de salive, même si certains labos suggèrent 30 minutes. De plus, on peut augmenter le nombre de cellules prises dans les joues en se frottant les joues avant de prendre la salive.

---

## 4 Le chromosome Y (test STR)

---

Le chromosome Y n'existe que chez les hommes. Il est transmis de père en fils avec peu de différences.

On trouve en moyenne une modification ou mutation aux 83<sup>10</sup> ou 144<sup>11</sup> ans sur une longue période. Comme l'ADN peut être modifié durant le cours normal de l'existence et non seulement lors de la procréation, le nombre de générations n'est pas un facteur important pour calculer la probabilité d'une mutation. En d'autres mots, sur plusieurs siècles, on pourra se faire une idée du nombre d'années jusqu'à un ancêtre commun mais cela ne dépend pas de la longueur des générations. Par ailleurs, il peut se passer plusieurs siècles sans mutation comme il peut y avoir plusieurs mutations durant une génération.

Dans notre culture, le nom de famille a été longtemps transmis par le père. En théorie, si deux hommes ont le même nom de famille et des résultats ADN Y similaires, ils devraient descendre d'un même homme ayant vécu après l'année 1200, époque où les noms de famille ont commencé à être hérités de père en fils. Pour les descendants de la Nouvelle-France, on remonte en général à un seul pionnier, ou rarement à des cousins éloignés portant le même nom de famille à leur arrivée mais venant de la même région.

### 4.1 Aperçu

---

Dans le cadre d'une analyse du chromosome Y, le test trouvera des hommes descendants d'un ancêtre commun par la lignée masculine des deux personnes testées. Ce sont donc des cousins masculins. Les termes alternatifs comprennent les correspondants et les *matches*.

Le test STR de l'ADN Y vise à trouver une signature Y, soit un résumé des résultats, et une cohérence de ces signatures, entre deux descendants d'un ancêtre éloigné, ceci afin d'avoir une signature Y de référence pour cet ancêtre. Cette cohérence est évaluée à partir de la ressemblance des résultats ou plutôt, de leur faible différence. L'analyseur mesure différentes informations, additionne les différences pour chaque marqueur, et indique les cousins ayant de l'ADN Y le plus près.

En ayant la signature de référence, il suffira alors de la comparer avec celle d'un descendant pour confirmer ou non qu'il y a un ancêtre commun ayant vécu après l'année 1200. Dans le cadre d'une colonie, cet ancêtre commun sera l'immigrant, sauf exception. Pour comparer, il faut un langage commun, soit un test similaire (STR dans ce cas-ci) ou

---

<sup>10</sup> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8228294/> (section 2.2.1, d'après l'ADN Y des tests commerciaux)

<sup>11</sup> <https://www.yfull.com/faq/definitions/>

avoir l'expertise permettant de faire la comparaison entre deux sortes de tests si les données disponibles viennent de tests différents.

### 4.2 But

---

Le but du test STR de FTDNA est de trouver des clients avec de l'ADN Y semblable. Cela ne veut pas dire qu'ils sont des cousins proches ou éloignés, ni qu'ils descendent du même ancêtre immigrant (pour la Nouvelle-France) ou ancien (pour l'Europe ou les Amérindiens). Cela donnera une signature de comparaison. Dans 5% des cas (pour la Nouvelle-France), il peut y avoir une brisure génétique (adoption, infidélité d'une mère dans la lignée paternelle, une mère cachant la grossesse de sa fille). En 2023, la probabilité d'avoir un match avec un test YSTR-37 ou similaire est autour de 50% au Québec. En France, c'est très faible à cause du manque de tests génétiques effectués dans ce pays.

### 4.3 Les triangulations

---

La notion de triangulation couvre deux méthodes très différentes. L'ADN Y et l'ADNmt reposent sur un chromosome unique transmis de génération en génération, et les mutations trouvées sont identifiées par rapport à un chromosome de référence. La triangulation autosomale sera expliquée plus loin et utilise des concepts très différents.

Une triangulation Y implique que pour 2 fils d'un ancêtre, il existe:

- Une lignée documentée à chaque génération. On accepte parfois une parenté virtuelle si les lignées documentées viennent de deux immigrants porteurs du même nom de famille et venant de la même région. Par exemple, on peut accepter Marin et Gaspard Boucher du Perche, les quatre Gagnon également du Perche, les deux René Landry acadiens, deux familles Gauthier et Doré, etc.
- Des tests ADN Y cohérents. Si ce sont des tests STR de FTDNA, idéalement des tests d'au moins 37 marqueurs avec matches qui se voient l'un l'autre en examinant les résultats.
- Idéalement, ce sont des tests du même type. Parfois, on peut accepter un test STR de FTDNA et un test SNP de 23andme si ce dernier test est d'une sorte plus rare. On peut accepter un test à 12 ou 25 marqueurs si le nombre de résultats (ou de matches) est faible. On peut également accepter une différence de plus que celle utilisée automatiquement par FTDNA si les marqueurs différents sont rapides ou des groupes de marqueurs (comme 19-19 contre 19-19-19).

## Le chromosome Y (test STR)

Un catalogue<sup>12</sup> réunit ces triangulations classées par ordre alphabétiques avec les pionniers venus en Nouvelle-France et dont la signature est connue et confirmée. En août 2023, à partir de 300 couples mariés au Québec en 1939-1940, nous trouvons 150 hommes sans signature, 106 avec une signature triangulée, 10 lignées simples non triangulées et 34 signatures Y basées sur l'ancêtre du client (sans savoir qui est adopté).

### 4.4 Le test STR de FTDNA

---

Le laboratoire FTDNA est le principal laboratoire utilisé pour analyser le chromosome Y et pour trouver des cousins avec un chromosome Y similaire. Son test de base, appelé Y-37, consiste à lire 37 séquences identifiées dites marqueurs STR. Le client pourra alors savoir qui, parmi les autres clients de ce laboratoire, a un chromosome Y similaire ou ayant le moins de différences (la distance génétique).

Il faut aussi comprendre que des résultats identiques ne permettent pas de connaître le niveau de parenté. Deux clients peuvent avoir une distance génétique de zéro (ou zéro différence avec 37 ou 111 marqueurs) et descendre de deux fils d'un ancêtre commun ayant vécu il y a 300 ans. Ils pourraient aussi avoir une différence de deux sur une génération (père et fils).

L'analyse des résultats donnera pour chaque groupe de marqueurs le nombre de différences avec les autres clients du même laboratoire. Même si cela se faisait lors des premiers tests, FTDNA ne permet plus l'importation de résultats provenant d'autres laboratoires.

#### 4.4.1 Haplogroupe et haplotype

---

Le résultat peut être résumé par une valeur simple, l'haplogroupe, présenté sous la forme d'une lettre et d'une mutation. Voici les principaux haplogroupes prédits à partir de tests STR en comparant avec des tests sur les SNP. Dans l'affichage de résultats chez FTDNA, les haplogroupes prédits sont, en ce qui concerne des origines françaises:

E-M35 (E1b), G-M201, I-M170 (I), I-M253 (I1), I-P37 ou I-M223 (I2), J-M267 (J1), J-M172 (J2), L-M20 (rare), N-M231, N-M232, R-M173 (R), R-M198 (R1a), R-M512 (R1a), R-M269 (R1b), T-M70.
--

Certains haplogroupes qu'on peut retrouver chez des descendants de Français sont toutefois considérés comme asiatiques. Ils peuvent être amérindiens si leurs matches le sont également. Quelques familles françaises sont classées comme C-M216 mais les résultats similaires sont tous européens. L'haplogroupe R (précédant ses sous-divisions en R1a, R1b et R2) était aussi en Asie et on a trouvé des Amérindiens avec cet haplogroupe R au centre du continent.

<sup>12</sup> <http://www.francogene.com/triangulation/y.php>

## Le chromosome Y (test STR)

C-M216, O-M175, Q-M242.

Cette façon de regrouper les principales familles est cependant récente et les lettres initiales auraient pu être plus nombreuses. Dans certains cas, les sous-groupes sont assez importants pour qu'ils soient indiqués dans certaines publications. Ainsi, R-M269 est aussi appelé R1b-M269. Certaines étiquettes ont disparu, comme E-L117 et E-M183 remplacés par E-M35.

L'haplotype Y est formé par la liste des valeurs des différents marqueurs en fonction du test. Par exemple, les douze premiers marqueurs de l'haplotype le plus commun dans le projet French Heritage:

13	24	14	11	11-14	12	12	12	13	13	29
----	----	----	----	-------	----	----	----	----	----	----

### 4.4.2 Nombre de marqueurs

FTDNA offre cinq niveaux de précision dans ses tests Y. Plus un test comprend de marqueurs et plus l'on acceptera de différences lors de l'affichage des résultats. En réduisant le nombre de marqueurs, on tolérera moins de différences mais on trouvera aussi beaucoup plus de résultats parce que d'un côté, la précision est moindre, et de l'autre, il y a plus de clients. Voyons les limites liées au nombre de marqueurs, ainsi que l'année d'apparition des différents tests Y par STR, le nombre de marqueurs similaires acceptés dans un match, et les deux plus grands nombres d'haplotypes selon le test.

Année	2000	2001	2003	2006	2011	2013	2018	2019
Test	Y-12	Y-25	Y-37	Y-67	Y-111	Big Y	Big Y-500	Big Y-700
Distance	0 ou 1	2	4	7	10			
Marqueurs similaires	11	23	33	60	101			
Max haplotypes	203	18	8	7	7			
2 <sup>e</sup> rang	97	10	7	6	3			

Le premier test proposé mesurait 11 séquences puis on découvrit qu'une de ces séquences pouvait apparaître à 2 endroits et ces tests sont devenus des Y-12. Les autres tests sont apparus un après l'autre. Pour simplifier la présentation, nous utilisons [Y-12] ou [test 12 marqueurs] et ces expressions sont équivalentes.

Comparons le nombre de marqueurs et les caractéristiques. Plus le nombre de marqueurs est élevé et plus:

- le prix est élevé.

## Le chromosome Y (test STR)

- le nombre de correspondants trouvés est faible. Ceci s'explique par le nombre plus faible de clients achetant un test plus précis. De plus, les résultats sont plus proches ou en d'autres mots, l'ADN Y est plus semblable parce qu'il faut plus de valeurs semblables (11 ou 12 sur 12 pour Y-12 mais 101 sur 111 pour Y-111).
- le nombre de différences tolérées ("Distance" dans le tableau précédent) est plus élevé mais surtout, le nombre total de marqueurs similaires requis est beaucoup plus élevé.
- la probabilité qu'un *match* valide ou probant soit un cousin dans un horizon généalogique (depuis l'année 1600) ou dans un horizon des noms de famille (depuis l'année 1200) devient aussi plus élevé.
- le nombre de résultats différents ("Max haplotypes" dans le tableau ci-dessus, pour le projet French Heritage, ainsi que "2° rang" pour l'haplotype au 2° rang) diminue;

FTDNA a un système interne de recherche de ces cousins. Une fois que le client a retourné sa salive au laboratoire, il y a plusieurs étapes comme l'envoi d'un accusé-réception, l'ajout à un lot pour traitement (*batch*), l'analyse ADN proprement dite, la recherche parmi les autres clients de ceux ayant de l'ADN Y similaire. Par la suite, le client reçoit un autre message chaque fois qu'un nouveau cousin est découvert.

À partir de son écran d'accès (*dashboard*), le client a plusieurs options dans la section des tests ADN Y.

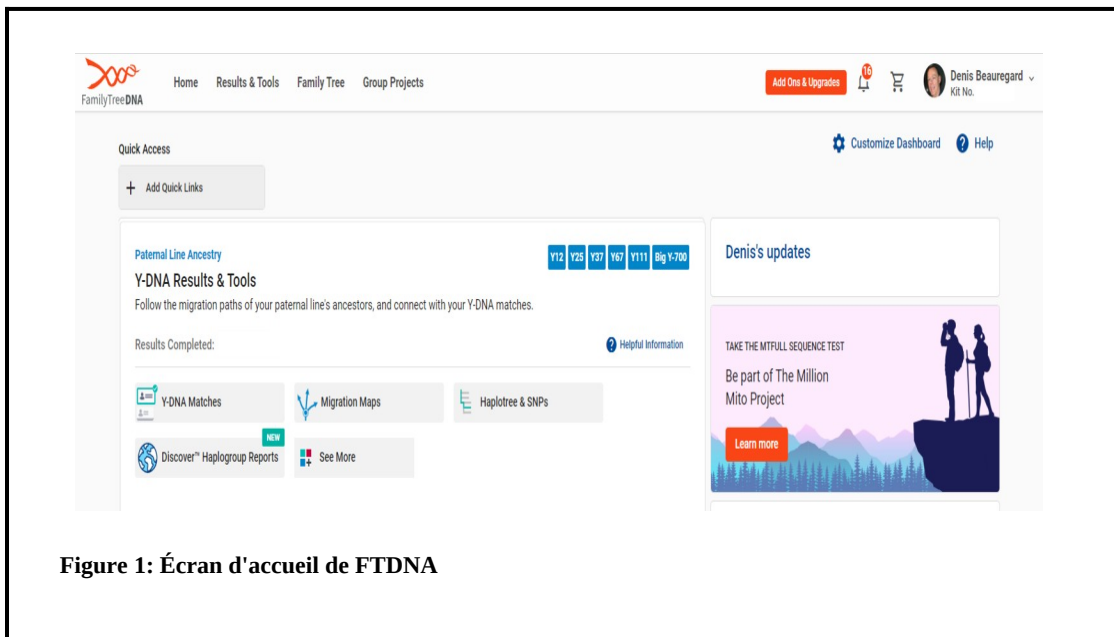


Figure 1: Écran d'accueil de FTDNA

## Le chromosome Y (test STR)

Au centre se trouve la liste des tests effectués par le client.

La section [**Y-DNA Matches**] montre une liste des cousins trouvés.

**Figure 2: Résultats ADN Y de FTDNA**

- ① La comparaison avec les autres clients peut se limiter à ceux ayant fait tel test. Comme le test de 111 marqueurs comprend les autres, en choisissant 67 marqueurs, on verra aussi les cousins ayant fait le test Y-111 mais pas ceux qui n'ont fait que Y-37 ou moins. On peut éliminer les matches avec moins de marqueurs s'ils sont trop nombreux par la configuration sous le nom du client en haut à droite, [*Account Settings*] puis [*PRIVACY & SHARING*] puis [*Y-DNA Match Levels*]. Par exemple, s'il y a 1000 matches avec Y-12, on peut choisir de ne pas les voir. Les autres clients ne verront pas eux non plus vos résultats.

Si un client fait le test Y-111, il peut sélectionner n'importe quel autre test et voir ses concordances basées seulement sur cette sélection. S'il a fait le Y-37, il pourra voir ceux qui ont fait un Y-111 ou Y-67 et dont la partie Y-37 est cohérente.

- ② La liste des cousins apparaît par la suite avec un certain nombre d'informations. S'il y a une photo, elle sera affichée avec son nom (ou des initiales s'il veut demeurer anonyme). Suit la liste des tests ADN qu'il a fait. Au bout de la ligne, 3 icônes ajoutent des fonctions, à savoir:
- Un arbre symbolique en bleu indique que le client a un arbre sur FTDNA (mais cet arbre est parfois presque vide). Si l'icône est faite de carrés vides, il n'y a pas d'arbre en ligne. Si elle est biffée, l'arbre est privé.
  - La 2<sup>e</sup> icône permet d'ajouter une note visible pour vous mais pas par l'autre client.
  - La dernière icône est un lien vers un outil de prédiction de la parenté avec le cousin en question. Ce tableau est statique et donne les mêmes valeurs pour tout le monde. Une ligne comme "1600 CE (950 - 1950 CE)" est interprétée comme "un ancêtre commun aurait vécu autour de l'an 1600 de notre ère, entre les années 950 et 1950". Les experts dans le domaine déconseillent ce genre d'outil qui n'est pas vraiment précis.

## Le chromosome Y (test STR)

La ligne suivante montre le test Y du client, la distance génétique pour la catégorie choisie, le nombre de STR communs s'il a fait le test Big Y (que nous verrons au prochain chapitre), le pays d'origine selon le client (certains peuvent écrire Canada ou États-Unis alors que l'ancêtre vient de France), et finalement le plus ancien ancêtre paternel connu.

- ③ La liste affichée peut être modifiée grâce à l'activation de certains filtres comme la distance génétique (ou le nombre de différences), l'appartenance à un projet commun (les projets de FTDNA font l'objet d'un autre chapitre), la présence de certains tests, etc.
- ④ La liste peut être triée de différentes façons. On peut utiliser le lien "Sort", ou encore afficher en mode *Table View* et cliquer sur le nom d'une colonne. Cette 2<sup>e</sup> méthode est la seule permettant un tri en ordre des noms de famille.

### 4.5 Exemples de nombre de résultats

---

#### 4.5.1 Familles avec le plus de descendants

---

Chaque client a des valeurs différentes en ce qui concerne le nombre de résultats dans chaque catégorie. Voyons quelques exemples.

Dans mon test personnel, les résultats avec Y-111 et Y-67 sont seulement des descendants de mon ancêtre immigrant, André Jarret, sieur de Beauregard [TRI0041]. Il y a 2 résultats de plus dans les Y-67 parce que ces cousins n'ont que le test Y-67. Avec le test Y-37, d'autres familles apparaissent. Selon le test Big Y (vu au prochain chapitre), l'ancêtre commun aurait vécu il y a plus de 1900 ans. Avec le test Y-25, apparaissent des cousins plus éloignés, mais aussi certains dont l'ancêtre commun aurait vécu il y a seulement 1200 ans. Finalement, le nombre de cousins par Y-12 est plus faible que les Y-25, ce qui n'est pas habituel.

Les Gagnon [TRI0139, TRI0151] sont une des plus grosses familles au Québec et un de ses membres est très actif dans le domaine de l'ADN généalogique. Ici, le nombre de résultats augmente lentement et de façon continue. Le nombre de Y-25 et Y-12 est à peine plus élevé que les Y-37. Selon les résultats du test Big Y, qui seront expliqués au chapitre suivant, tous les descendants Gagnon ayant fait ce test partagent entre eux 44 mutations qui ne se retrouvent chez aucune autre famille, pour le moment, ce qui mènerait à un ancêtre familial il y a 3500 ans.

Les Tremblay [TRI0164] ont le plus grand nombre de descendants du nom au Québec. Leurs nombres de résultats sont très différents. Le nombre de correspondances totales (donc, avec d'autres noms que les Tremblay) est assez bas de Y-111 à Y-37, mais approche les 1500 avec Y-25 et les 6000 avec Y-12. Leur Big Y montre 31 mutations uniques et l'ancêtre familial aurait vécu il y a 3200 ans.

## Le chromosome Y (test STR)

Le cas des Côté [TRI0029], une autre famille nombreuse, est dans une situation intermédiaire avec 175 résultats avec Y-12 (aussi, de plusieurs familles différentes) mais avec un résultat inattendu, soit la présence d'un grand nombre de résultats arabes, avec qu'il y a presque uniquement des Côté dans le Y-25. Dans cette famille, le Big Y indique un ancêtre familial né vers l'an 250.

Les Leblanc [TRI0149] forment la famille acadienne la plus nombreuse. Ici, les Y-37 sont le triple des Y-67, les Y-25 le triple des Y-37 et les Y-12, le triple des Y-25. Il y a donc une hausse logarithmique du nombre de résultats. Les Leblanc auraient leur chromosome Y particulier depuis 4400 ans.

Les Doucet forment deux familles. L'immigrant venu de France [TRI0181], frère de Germain Doucet, a un seul fils, Pierre. Le nombre de matches pour des gros tests est modéré, mais dépasse les 1200 correspondances avec Y-25 et redescend à près de 400 avec les Y-12. L'autre famille vient d'un autre Germain Doucet [TRI0046] dont l'ADN Y révèle son origine amérindienne. Alors que les tests Y-111 des descendants de Pierre sont rares, on en trouve 36 pour les descendants de Germain. Le nombre de Y-12 est toutefois modeste. Dans ce dernier cas, c'est sans doute parce qu'il y a peu d'Autochtones qui font des tests ADN.

Il faut remarquer que pour les grosses familles, souvent, il n'y a pas de résultats de tests ADN Y avec des familles rapprochées en France, faute de testeurs disponibles (à cause de la loi française). Cela isole ces familles. Mais pour la plupart des familles, l'ancêtre partagé avec d'autres immigrants a vécu après l'an zéro. Avoir l'exclusivité d'un ancêtre ayant vécu il y a 2000 ou 3000 ans signifie simplement que pour le moment, il n'y a pas de cousin plus proche qui ait fait un tel test.

À l'autre extrémité des résultats, certains tests Y-111 n'ont aucun correspondant sauf avec Y-12 ou Y-25. Il s'agit souvent de familles ayant un petit nombre de descendants en ligne masculine.

Si quelqu'un veut une signature de référence pour sa famille et qu'elle n'est pas déjà répertoriée, même avec un seul test, il est probable qu'il faudra trouver un autre homme descendant idéalement d'un autre fils de l'ancêtre plutôt que se fier au hasard pour trouver des cousins parmi les résultats.

La conclusion est sans doute qu'il est difficile de prédire quels seront les résultats avant d'avoir fait un test ADN et que si l'on est à la recherche d'une signature "type" ou de "référence" pour une lignée d'hommes, mieux vaut procéder par une triangulation.

### 4.5.2 Signature de référence

---

Le Québec est privilégié avec un petit nombre de pionniers et une généalogie assez bien documentée pour les années anciennes (avant 1800). Plusieurs font un test ADN Y dans le but de connaître la signature ADN Y de référence pour leur ancêtre. Le but est alors de réussir une triangulation, soit d'avoir 2 lignées documentées vers deux fils d'un ancêtre commun, idéalement un pionnier de Nouvelle-France. L'Acadie est dans une situation similaire, sauf que la moitié des registres n'ont pas survécu à la Déportation de 1755-1758. La proportion de descendants d'Acadiens testés est toutefois plus élevée. Dans le reste de l'Amérique du Nord, on trouve d'un côté les descendants de ces familles, et de l'autre d'autres immigrants venus d'Europe, comme les Huguenots dans les anciennes colonies anglaises, ou des Français qui ont peuplé directement le centre du continent en arrivant par la Louisiane. Selon les chemins migratoires, le nombre de personnes testées et la qualité de la documentation disponible, il est possible de dégager une signature triangulée pour un grand nombre d'arrivants. Pour le Québec et l'ancienne Acadie, on peut s'attendre à 3000 signatures en couvrant le régime français.

Il existe plusieurs laboratoires ADN et le produit livré varie beaucoup. L'intérêt du test Y-37 de FTDNA est qu'il y a déjà une bonne quantité de clients ayant fait un test ADN et qu'il y a donc de bonnes chances pour trouver par hasard un autre client de la même famille biologique et ayant fait un test cohérent. Dans ce contexte, on regardera les matches avec au moins 37 marqueurs. Avec 12 ou 25 et un nombre le moins important de cousins, il est néanmoins possible d'avoir un faux positif, soit un autre descendant apparaissant dans les matches, avec une lignée documentée vers un ancêtre commun, mais avec une brisure génétique apparaissant quand on augmente le nombre de marqueurs.

Ainsi, les Fortin [TRI0110] et Fortier [TRI0286] peuvent avoir le même résultat avec le test Y-12, mais atteignent 33 différences avec le test Y-111. Dans le projet French Heritage, il y a 202 tests avec le même résultat Y-12 sur plus de 5300 résultats Y-12 ou mieux. Ces 202 résultats (plus en réalité, car on tolère une différence au sein d'un même projet) sont les mêmes pour ces familles. On peut penser que ces lignées familiales descendent d'un même ancêtre commun qui aurait vécu il y a fort longtemps.

Archambault, Aubin, Bergeron, Bourgeois, Brodeur, Brun, Chabot, Dionne, Garneau, Langlais, Lauzon, Lavoie, Levasseur, Maheu, Miville, Moreau, Perrot, Proulx, Racine, Sauvé, etc.
---

L'inverse est aussi possible. Les Lévesque dit Rompré et Dusablon [TRI0766] ont seulement 5 correspondances avec un Y-12 et aucun avec plus de marqueurs. Dans ce cas, il a suffi d'un test 23andme, moins précis, pour confirmer la signature, parce que le

## Le chromosome Y (test STR)

nombre de résultats est très faible tout comme la probabilité d'une erreur. Si le nombre de matches trouvés est très faible, un test Y-12 peut parfois être suffisant pour confirmer une triangulation.

En général, donc, il faudra au moins 37 marqueurs pour confirmer la signature de l'ancêtre et avec 111, on diminue la possibilité d'un hasard.

### 4.5.3 Parents inconnus

---

Plusieurs généalogistes font un test ADN afin de débloquent une lignée ancienne. En général, au Québec et en Acadie, les grosses familles ont déjà fait l'objet d'une étude. En juin 2023, dans la base de référence de l'auteur, sur 76 430 fiches, 44% ont une signature ADN Y et 66,8% ont une signature ADNmt (parce qu'il y a moins de pionnières venues durant le régime français et que les femmes ne peuvent faire que le test d'ADNmt). Comme ces fiches contiennent également les familles des ancêtres en France et que tout le monde n'est pas testé, la probabilité de trouver le nom de famille d'un ancêtre né de père inconnu frôle le 50%. Il faudra utiliser un homme pour faire ce test, que ce soit le père, le frère, le cousin paternel, etc. et celui-ci devra descendre de l'ancêtre de parents inconnus en ligne paternelle seulement.

En général, un test Y-37 donnera un petit nombre de cousins de la même famille biologique et d'autres qui y sont par hasard. Ce phénomène est la **dérive génétique**. Les valeurs STR peuvent augmenter ou diminuer au hasard. Deux lignées avec des valeurs éloignées il y a 5000 ans par exemple, pourraient avoir des valeurs STR qui ont changé dans des directions opposées pour se rapprocher. Plus le nombre de marqueurs est élevé, et plus il faut qu'un grand nombre de marqueurs change pour que deux familles se rapprochent, alors qu'avec seulement 12, ce nombre est petit.

Il faut alors un peu d'expertise pour partager les cas.

- Un nombre très faible de correspondants trouvés suggère une immigration récente (les Français faisant peu de tests ADN) ou une petite famille. Si aucun nom n'est à consonance francophone, il peut s'agir d'une petite famille qui n'a pas encore fourni de signature de référence, ou d'une famille non-francophone. Il est aussi possible que le nom de famille ait été anglicisé. Dans tous les cas, il faudrait faire une recherche généalogique et même contacter l'autre personne avec tout le doigté nécessaire et ne pas parler tout de suite de parents inconnus.
- Un nombre élevé de matches montrera probablement plusieurs noms de famille possibles. On peut alors augmenter la précision du test, comme passer de 37 à 111 marqueurs (il n'est pas nécessaire d'acheter un nouveau test et des mises à niveau sont disponibles), ou même essayer le Big Y que nous verrons plus loin.

## Le chromosome Y (test STR)

- Parfois, plusieurs noms de famille sont possibles. D'un côté, des noms différents peuvent tout de même venir du même ancêtre, soit à cause des "noms dits" communs avant 1800, soit à cause de traduction ou adoption d'un nom à saveur locale lors de migrations, soit parce qu'un enfant est adopté ou a pris le nom de sa mère ou du 2<sup>e</sup> conjoint de celle-ci, soit pour d'autres raisons que vous pouvez imaginer. D'un autre côté, il peut y avoir réellement 2 souches différentes avec le même ADN Y parce que 2 ou 3 générations (ou plus) avant l'immigration, une des autres possibilités est arrivée. De plus, les noms de famille en France ne sont hérités que depuis l'an 1200 environ. Avec un Y-37, l'ancêtre commun peut avoir vécu il y a 2000 ans, longtemps avant que les noms de famille soient hérités.

Le test sur l'ADN Y permet donc, si vous êtes en Amérique du Nord, de trouver un nom de famille potentiel, ou plusieurs. Dans les régions où il y a moins de clients comme la France et les pays où le revenu moyen est plutôt faible, la probabilité d'avoir le bon nom de famille parmi ses correspondants est toutefois assez bas.

À partir du nom de famille, il est possible de s'en servir avec un test ADN autosomal que nous verrons plus loin, ceci afin d'identifier parmi les cousins potentiels ceux qui seraient apparentés en lignée paternelle.

### 4.5.4 Les événements non parentaux (ÉNP)

---

Il y a parfois une brisure génétique et l'ancêtre selon la documentation n'est pas celui indiqué par la génétique. Il s'agit d'un événement non parental ou non paternel. En anglais, l'expression est *Non-paternity event*<sup>13</sup> ou NPE qu'on décrit aussi comme *Not the Parents Expected*.

En Nouvelle-France, les registres sont assez complets. D'un côté, ces actes de baptême et autres situent le taux d'enfants nés hors mariage entre 1% et 2%. Selon Lyne Paquette et Réal Bates<sup>14</sup>, ce serait 0,8% avant 1700 et 1,5% en 1700-1729. D'autres auteurs obtiennent des résultats similaires en comptant ces naissances dans les dictionnaires de Cyprien Tanguay et de René Jetté. D'un autre côté, les résultats des tests ADN Y disponibles et de l'analyse des lignées généalogiques de soutien nous apprennent que dans une naissance sur environ 200<sup>15</sup>, le père biologique n'est pas le père légal. Ailleurs dans le monde, la moyenne serait d'environ une ou deux naissances sur cent, peut-être parce que le nombre de naissances hors-mariage est inconnu. Toutefois, la portion varie beaucoup selon la région et le milieu culturel<sup>16</sup>.

---

<sup>13</sup> [https://isogg.org/wiki/Non-paternity\\_event](https://isogg.org/wiki/Non-paternity_event) mais certains utilisent *Not paternal event*.

<sup>14</sup> Les naissances illégitimes sur les rives du Saint-Laurent avant 1730, p. 242, <https://www.erudit.org/fr/revues/haf/1986-v40-n2-haf2343/304446ar/>

<sup>15</sup> Selon les résultats des tests ADN Y examinés par l'auteur et ses proches collaborateurs

<sup>16</sup> [https://isogg.org/wiki/Non-paternity\\_event](https://isogg.org/wiki/Non-paternity_event)

## Le chromosome Y (test STR)

Comme il s'agit de lignées paternelles, on peut se demander s'il pourrait s'agir d'adoptions qui ne sont pas annoncées dans les actes. L'analyse des lignées maternelles à partir de l'ADNmt montre plutôt que si le résultat n'est pas celui attendu, c'est qu'il y a une erreur habituellement détectable dans la lignée concernée. Bien qu'il soit impossible d'identifier la cause de chaque ÉNP, il est vraisemblable qu'il s'agisse d'une infidélité conjugale ou d'une fille-mère. Le second cas est prévisible si le baptême est le dernier des enfants, si une sœur est assez vieille pour être la mère ou si celle-ci se marie tardivement.

Un taux de 1/200 par génération nous donne une réussite de 95%. En d'autres mots, dans 95% des cas où le client connaît ses parents biologiques, son test ADN Y donnera soit un résultat cohérent avec un autre client du même nom ou l'équivalent (comme un surnom ou une traduction), soit avec un autre client qui dit descendre du même ancêtre, soit aucun ancêtre commun identifiable parce que la famille n'a pas encore fait l'objet d'une étude ou d'un autre test ADN Y.

### 4.5.5 Les sortes d'événements non parentaux

---

Plusieurs situations sont possibles lors de l'analyse de ces ÉNP.

#### 4.5.5.1 Deux brisures flottantes

---

Parfois, nous voyons 2 brisures flottantes, soit deux signatures différentes menant vers un même pionnier, en attendant qu'un 3<sup>e</sup> candidat fasse un test et communique sa généalogie afin de départager les lignées.

C'est le cas d'Urbain Baudreau dit Graveline et de Pierre Dagenais dit Leprince.

#### 4.5.5.2 Brisure délimitée

---

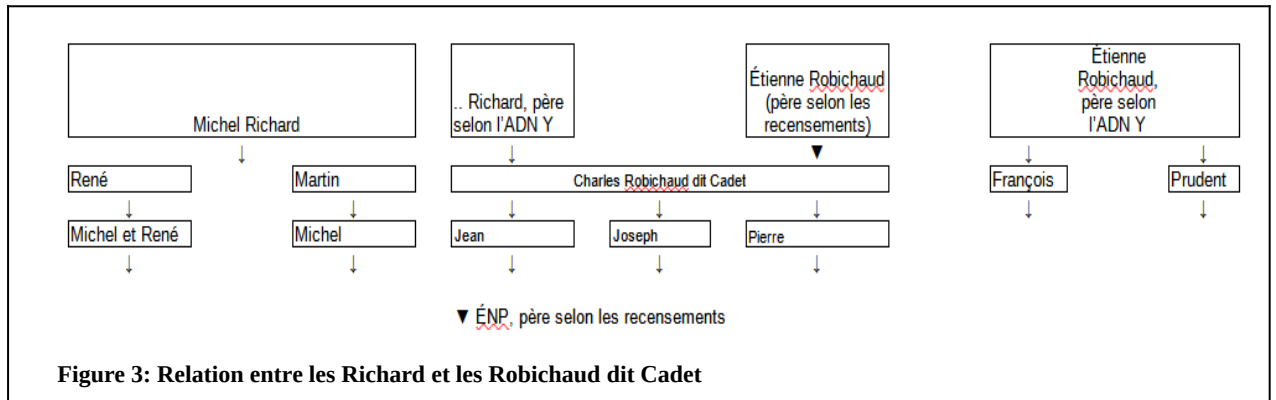
Il peut y avoir 2 lignées avec la signature divergente, ce qui donne une brisure délimitée. Dans ce cas, le nombre de générations où la brisure a pu se produire est réduit.

Lors de la première publication sur les Robichaud [TRI0132] du Nouveau-Brunswick qui sont en fait des Richard [TRI0202], deux fils de Joseph Robichaud dit Cadet, époux de Claire Leblanc, avaient une signature différente de celle des autres Robichaud [TRI0133]. La brisure pouvait concerner ce Joseph ou son père Charles. Par la suite, l'étude d'autres lignées ont démontré que c'est Charles qui avait l'ADN Y des Richard.

Ce cas est plus subtil car il se trouve que ce Charles et sa sœur Madeleine sont surnommés Cadet alors qu'ils sont les aînés de la famille Robichaud selon le recensement de 1686. Cela pourrait signifier que Michel Richard serait arrivé en Acadie avec un frère qui était peut-être l'époux de Françoise Boudreau. Puis, ce frère serait décédé et Françoise aurait alors épousé Étienne Robichaud. L'absence de documents anciens et le fait

## Le chromosome Y (test STR)

qu'Étienne ait refusé de répondre au recensement de 1671 ne permet pas de vérifier si cette théorie tient la route. Une infidélité est aussi possible.

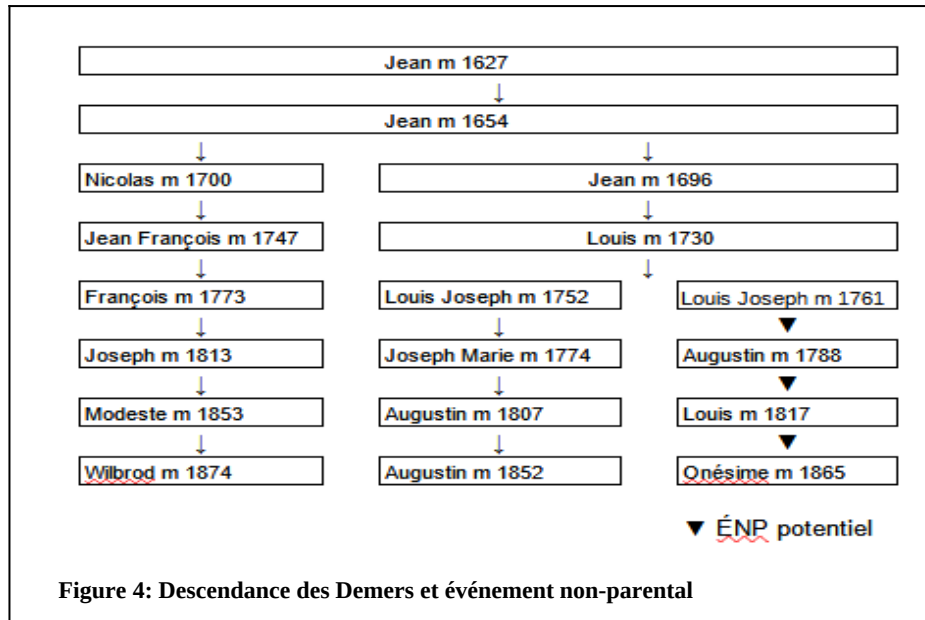


### 4.5.5.3 Brisure flottante

Le cas le plus fréquent est celui d'une famille triangulée, donc validée par deux lignées documentées et des résultats cohérents, et d'une troisième lignée documentée et rejoignant une des deux autres lignées documentées, mais avec des résultats divergeants. C'est la brisure flottante. Dans un tel cas, il faut tenter de couper la poire en deux et trouver un descendant d'un des ancêtres situé entre la personne testée et l'ancêtre dans la partie triangulée.

Un exemple est celui d'un Demers dont la documentation sur papier mène au pionnier Jean Dumais [TRI0183]. Dans son cas, la lignée triangulée a été déplacée vers Louis Joseph Demers et Françoise Paquet. La filiation inconnue se trouve entre leur fils Augustin et les quatre générations qui suivent.

## Le chromosome Y (test STR)



Un cas plus compliqué apparaît dans une famille Rousseau. Joseph Rousseau qui épouse Marguerite Drolet en 1827 est le fils d'Augustin, dont la lignée triangulée [TRI0626] montre qu'il est de l'haplogroupe I1-M253. Joseph a 3 fils dont 2 ont fait l'objet d'un test. Le descendant de son fils Joseph a l'haplogroupe I2-M223 alors que celui de son autre fils François est R1b-M269 (il est un Gagnon biologique).

### 4.5.5.4 Brisure précise

Il peut arriver qu'il y ait zéro générations entre la brisure et la lignée triangulée, ce qui donne une brisure précise.

C'est le cas de Charles Fournier, mari d'Élisabeth Bouchard [TRI0080]. L'ADN Y (G-PF4202) de trois de ses fils est connu et proviendrait d'un des frères Jean ou Gilles Goteau [TRI0375]. Il est différent de celui (R1b-M269) de son père Guillaume, époux de Françoise Hébert [TRI0357], connu grâce à 3 de ses fils.

### 4.5.5.5 Brisure double

Le dernier cas est celui de la brisure double, alors qu'il y a 2 triangulations voisines menant vers un même ancêtre, mais avec des signatures non cohérentes.

L'ADN Y de l'immigrant Mathurin Martineau dit Saintonge est encore inconnu. Celui de son fils Joseph (m Marie Anne Boucher) a comme haplogroupe I2-M223 [TRI0237] alors que pour son fils Simon (m Geneviève Arcand) et son petit-fils Simon (m Madeleine Pichet), on trouve R1b-M269 [TRI0238].

### 4.5.5.6 Adoptions et parents inconnus

## Le chromosome Y (test STR)

En parallèle, on trouve un certain nombre d'enfants dont l'adoption est connue et dont le père biologique est recherché.

Jacques Pierre, qui passa un contrat de mariage avec Angélique Choret en 1746, est né de parents inconnus avec l'haplogroupe I2-M253>Y129360 [TRI0653]. Comme ses enfants ont parfois le patronyme de Jacques, une première hypothèse était que son père est un Jacques. Pierre Jacques dit Pierre-Jacques est étiqueté R1b-Z381 [TRI0209]. Puis, un Châtigny est apparu parmi ses correspondants et en testant d'autres descendants de cette famille, il est apparu que le père ou grand-père de Jacques Pierre devait être Pierre Châtigny, époux d'Angélique Martin [TRI0649].

De même, Jean Baptiste Roy, époux de Marie Cantin en 1815 est un Drapeau dont il partage l'ADN Y R1b-A8451 [TRI0185].

### 4.6 L'affichage de résultats STR

#### 4.6.1 Autres données du client

En plus de la liste des correspondants, le client peut trouver d'autres informations dans le bloc de l'ADN Y. Les liens indiqués n'apparaissent que s'ils sont pertinents.

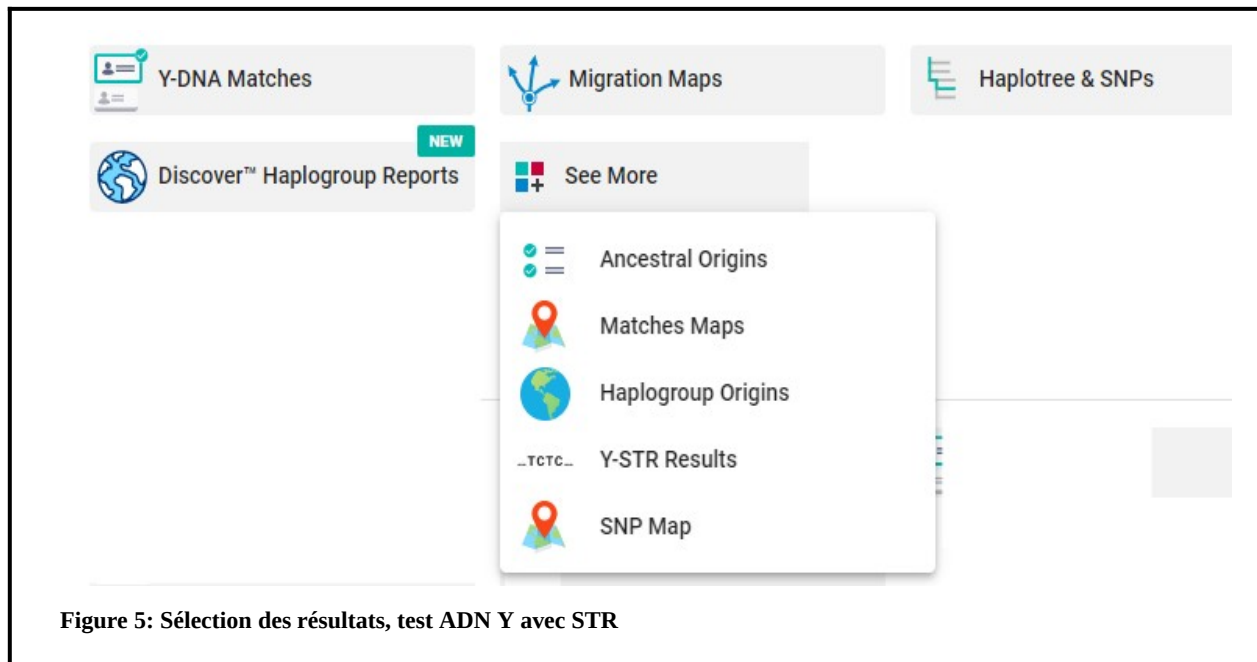


Figure 5: Sélection des résultats, test ADN Y avec STR

Le lien [*Migration Maps*] montre le chemin ancien suivi par la lignée paternelle depuis l'Adam africain. Au bas de la page, il y a un texte qui résume les étapes principales. On peut aussi choisir [*All Migration Maps*] pour voir les autres chemins à

## Le chromosome Y (test STR)

partir des lettres au bas de la page. Ces lettres représentent les principales familles dans une vue d'ensemble de l'humanité.

Le lien [*Haplotree & SNP*] montre la position du SNP le plus récent (si un test SNP a été effectué) ou le SNP prédit par le test STR (en le comparant avec ses correspondants).

Le lien [*Discover*] sera discuté au chapitre sur les Big Y.

Finalement, le lien [*More*] affichera d'autres informations comme l'origine indiquée par les matches (qui n'est pas nécessairement votre origine), le lieu d'origine en fonction du SNP des matches, les données brutes pour chaque marqueur, soit une séquence de chiffres qu'il faut comparer avec ceux des autres clients (voir plus bas) et finalement une carte permettant de pointer chaque SNP mais qui n'est plus utile avec la popularité des tests Big Y.

### 4.6.2 Résultats STR des autres clients d'un projet

Les STR permettent de comparer vos résultats avec ceux d'autres clients et éventuellement de trouver des cousins, parfois très éloignés. Le test STR en soi donne une liste de clients avec de l'ADN semblable. En rejoignant un projet qui affiche publiquement ses résultats, vous pouvez voir comment l'ADN se compare avec d'autres. Les sous-groupes sont nommés et construits de façon indépendante dans chaque projet et permettent de voir quels STR sont différents dans un même sous-groupe. Cela permet de mieux étudier des cas limites (un peu plus loin que la détection par défaut des cousins si vous regardez vos matches, par exemple). Voyons un exemple de regroupement.

JARRET de BEAUREGARD, Andre m 1676 Marguerite ANTHIAUME (GFAN 1933) (TRI0041)												
MIN					14	24	14	11	11-15	12	11	24
MAX					14	24	14	11	11-15	12	11	25
MODE					14	24	14	11	11-15	12	11	25
M134432	André Jarret m.1676 Marguerite Anthiaume	France	R-FT167459		14	24	14	11	11-15	12	11	24
865156	Andre Jarret Beauregard, b. 1641 and d. 1691	France	R-M269		14	24	14	11	11-15	12	11	25
279620	Andre Jarret, m 1676 Marguerite Anthiaume	France	R-FT167459		14	24	14	11	11-15	12	11	25
96525		France	R-M269		14	24	14	11	11-15	12	11	25
N2094	andre jarret dit beauregard,b.1723 and d.1771	Canada	R-M269		14	24	14	11	11-15	12	11	25
120340	Edouard Jalbert/Gilbert, Manchester, VT, USA	Canada	R-Y7363		14	24	14	11	11-15	12	11	25

Figure 6: Affichage public de résultats d'ADN Y avec STR

## Le chromosome Y (test STR)

Dans ce tableau, le nom du sous-groupe apparaît avec un fond jaune. Dans ce cas-ci, des liens cliquables permettent de voir la fiche familiale (GFAN 1933) sur le site de l'auteur, ainsi que les données sur la triangulation [TRI0041].

Les deux premières lignes donnent les valeurs limites pour chaque marqueur. La 3<sup>e</sup> colonne depuis la droite contient 2 valeurs, 24 et 25, et les minima et maxima de cette colonne apparaissent en haut, puis la valeur la plus fréquente (ou mode). Ensuite, les 6 personnes ayant fait le test sont affichées, avec le numéro de trousse (pour l'anonymisation des résultats), le plus ancien ancêtre connu, le lieu d'origine tel que saisi par le client, l'haplogroupe et finalement, les différents STR. La ligne bleue est un séparateur afin de montrer qu'une des valeurs est différente.

Les haplogroupes en rouge sont prédits en fonction du résultat des autres clients par régression multiple, alors que ceux en vert ont été détectés par un test approprié. La 5<sup>e</sup> colonne de marqueurs contient 2 chiffres parce que ce marqueur peut être détecté à deux endroits.

### 4.7 Des cas spéciaux

---

Les résultats sont parfois incomplets. Par exemple, si les résultats sont trop éloignés de ceux des autres clients, l'haplogroupe peut être impossible à prédire. Dans un tel cas, le client peut demander à FTDNA de faire un test supplémentaire pour trouver un haplogroupe en examinant les SNP présents dans son ADN Y. Mais il est plutôt préférable d'augmenter le nombre de marqueurs. En effet, les tests les plus anciens, de type Y-12 donc peu précis, viennent souvent de l'importation d'un test Geno 1<sup>17</sup>. Les échantillons de ce test n'ont pas été conservés et donc le test pour trouver un SNP ne pourra être effectué. En augmentant le nombre de marqueurs, le client pourra fournir un nouvel échantillon de salive.

L'absence de matches (ou leur rareté) signifie que les cousins possibles n'ont pas fait de tests chez FTDNA ou en ont fait très peu. Cette absence est commune quand elle concerne des cousins éventuels vivant dans des pays où il n'y a pas d'intérêt pour la généalogie ou encore, le coût du test est trop élevé pour la plupart des habitants. En France, une interprétation de la loi a fait que FTDNA n'y envoie plus de tests depuis février 2023.

Si un STR vaut zéro (la valeur est zéro et non un espace vide), c'est qu'il n'a pas été détecté. Un STR est identifié par une amorce et il est possible qu'une simple mutation

---

<sup>17</sup> Les tests Geno 1 et Geno 2 viennent d'un projet de la *National Geographic Society* et ne sont plus disponibles. Geno 1 proposait soit le test Y-12, soit le test MT HVR1 que nous verrons plus loin. Geno 2 montre des haplogroupes Y et MT évalués différemment. Le Y vient d'une série de SNP, ce que nous verrons au prochain chapitre. Dans les deux cas, les échantillons n'ont pas été conservés.

## Le chromosome Y (test STR)

rende cette amorce<sup>18</sup> non-délectable. S'il y a beaucoup de zéros, c'est que le chromosome Y est dégradé, par exemple suite à un traitement médical.

Autrefois, FTDNA permettait l'importation de tests Y en provenance d'autres compagnies. Certains de ces tests ne couvraient pas exactement les mêmes marqueurs STR que FTDNA. Ces tests ont un nombre différent de marqueurs, comme 33 et 46. Le visiteur verra donc que certaines valeurs sont omises dans les résultats.

### 4.8 Les autres laboratoires

---

FTDNA offre des tests ADN pour le grand public depuis 2000. D'autres laboratoires ont existé durant quelques années avant de disparaître alors que d'autres fournissent des résultats parfois équivalents, parfois différents. On trouve une liste de ces laboratoires dans le WIKI de l'ISOGG, un organisme de normalisation: [https://isogg.org/wiki/Y-DNA\\_STR\\_testing\\_comparison\\_chart](https://isogg.org/wiki/Y-DNA_STR_testing_comparison_chart)

Parmi les tests disparus, le nombre de marqueurs STR pouvait varier, ce qui explique les espaces vides dans certains résultats. Ces tests ont pu se retrouver sur des plateformes communes.

La plupart des autres laboratoires offrent, pour le chromosome Y, des séquençages du génome complet (WGS) dont la composante Y peut être analysée par un autre laboratoire. C'est le cas de Dante, FGC, Nebula et Yseq. Le séquençage complet permet de calculer certains STR mais comme il s'agit le plus souvent de séquences de 100 valeurs contiguës, ces séquences ne permettent de trouver qu'environ la moitié des STR de FTDNA. La comparaison est possible mais pas complète.

Un client de FGC nous ayant transmis ses résultats, comparons le nombre de STR calculés par bloc.

Marqueurs	FTDNA	FGC
1-12	12	7
13-25	13	3
26-37	12	6
38-67	30	18
68-111	44	25

Le laboratoire allemand YSEQ<sup>19</sup> offre plusieurs sortes de tests Y. D'un côté, les panneaux de plusieurs STR semblent bien correspondre à ceux de FTDNA mais être dans un ordre différent (détail important si on veut comparer). De l'autre, on peut aussi tester des STR individuels. Les étiquettes sont les mêmes que pour FTDNA et les valeurs

<sup>18</sup> Une amorce est une séquence ADN utilisée pour détecter le début d'une séquence STR.

<sup>19</sup> Fondé par Thomas Krahn, l'ancien généticien en chef de FTDNA.

## Le chromosome Y (test STR)

correspondent, selon les résultats reçus d'un client. YSEQ offre aussi un génome complet ainsi que des projets par nom de famille et par haplogroupe, et la possibilité de commander un SNP précis. Ils ont environ 1/10 des tests STR de FTDNA selon l'ISOGG.

Les laboratoires avec le plus grand nombre de résultats autosomaux (Ancestry, 23andme et myHeritage) n'ont aucun test Y STR.

### 4.9 Les sites de comparaison

---

Comme FTDNA ne permet pas l'importation de résultats STR, il faut comparer visuellement les panneaux publics de résultats si le test a été fait par un autre laboratoire.

La page [https://isogg.org/wiki/DNA\\_databases](https://isogg.org/wiki/DNA_databases) mentionne plusieurs de ces sites.

Le principal site en 2023 serait mitoydna.org qui reçoit plusieurs sortes de résultats. L'inscription semble obligatoire. Leurs règles interdisent l'importation d'un projet FTDNA. Elles donnent tout de même une liste assez détaillée de ce qui est accepté.

L'ancien site ysearch a disparu en 2018. Il permettait également de partager des résultats Y STR. Disparu également en 2018, *World Families Network* hébergeait un grand nombre de projets souvent rattachés à FTDNA. L'auteur a pu en récupérer une partie, ainsi que les données généalogiques.

## 5 Le chromosome Y (test SNP)

---

Le chromosome Y n'existe que chez les hommes. Il est transmis de père en fils avec peu de différences.

Le test ADN Y utilisant des SNP permet de catégoriser ce chromosome Y comme un nom de famille biologique. Ce nom a toutefois plusieurs alias car il s'agit de différentes étapes ou mutations depuis 250 000 ans.

### 5.1 Aperçu

---

Le test STR Y permet d'extraire certaines caractéristiques du chromosome Y et de trouver d'autres clients ayant de l'ADN Y similaire. Les STR sont des valeurs qui peuvent augmenter ou diminuer et il est donc possible qu'il y a une correspondance avec un cousin très éloigné. Ainsi, avec 37 marqueurs, les Beauregard et les Pagé peuvent se voir parmi les résultats des tests ADN Y alors que l'ancêtre commun a vécu il y a environ 2000 ans.

Les SNP ou mutations trouvées avec le test SNP sont considérées comme assez stables. En fait, il y a toujours une analyse des résultats des tests Big Y (le test le plus populaire pour explorer l'ensemble des SNP Y) afin de retenir les nouveaux SNP qui sont stables, cette stabilité étant confirmée par la présence d'au moins deux clients partageant le même SNP parmi la clientèle de FTDNA. En d'autres mots, le test Big Y trouvera des SNP Y déjà connus et répertoriés, ainsi que d'autres encore inconnus ou trouvés dans d'autres familles, et ces derniers ne seront retenus que s'ils se retrouvent chez plusieurs descendants d'une même branche.

En bref, en classant les SNP par quantité, on peut obtenir une évolution dans le temps (les SNP moins nombreux montrent les branches plus récentes) et voir comment les familles sont apparentées entre elles à une époque précédant le plus souvent l'apparition des noms de famille comme on les connaît aujourd'hui.

Le SNP P343 est connu chez un grand nombre d'hommes de l'ouest de l'Europe. La plupart d'entre eux ont aussi le SNP M269, mais pas tous. Ensuite, beaucoup ont le SNP P311. Ensuite, cela commence à se répartir et on trouve par exemple U106, P312 et CTS4528. Ceux qui ont la mutation U106 n'ont pas P312 ni CTS4528. Mais il peut arriver qu'une mutation apparaisse dans deux branches totalement différentes. Ainsi, nous avons des Beïque dit Lafleur, des Biroleau dit Lafleur, des Poitevin dit Lafleur, etc. qui sont des familles différentes utilisant le même surnom. Les SNP peuvent aussi faire la même chose. Les experts ont ajouté une lettre décrivant les principales familles, avec un

## Le chromosome Y (test SNP)

tiret pour les distinguer des haplogroupes d'ADNmt. Nous avons donc des haplogroupes comme R-M269, R-P311, R-U106, R-P312 et R-CTS4528. Cette science est assez récente et il y a eu des regroupements plus précis et les publications utilisent parfois des termes comme R1b-M269 ou R1b-P311.

Ce système permet une étude familiale de l'humanité, avec les branches masculines, qui s'étend sur des centaines de milliers d'années.

La mutation R-M343 (ou R1b-M343) serait apparue il y a 19 000 ans. Elle est précédée de la R-M173 ou R1-M173 de 3000 ans plus ancienne. R1 s'est aussi divisée pour donner R1a-M420 qui concerne davantage l'Europe de l'Est et qui serait apparue 2000 ans après R1b-M343.

En parallèle, on trouve un grand nombre de familles principales que nous avons vues au chapitre précédent. En ce qui concerne des origines françaises:

C-M216, E1b-M35, G-M201, I-M170, I1-M253, I2-P37 ou I2-M223, J1-M267, J2-M172, L-M20 (rare), N-M231, N-M232, O-M175, Q-M242, R-M173, R1a-M198, R1a-M512, R1b-M269, T-M70.

### 5.2 La séquence des SNP menant vers les Beauregard

Le lecteur sera sans doute intéressé par un cas réel.

Je descends d'une lignée qui part de la mutation M343 vieille d'environ 17 000 ans selon FTDNA ou presque 23 000 ans selon Yfull. Wikipedia donne une source qui propose entre 12 500 et 25 700. La variété de cette prédiction montre que l'apparition de ces mutations est une sorte de calcul et n'est pas faite à partir de sépultures anciennes qu'on aurait datées au carbone 14 ! Par contre, des sépultures anciennes permettent parfois de confirmer l'ancienneté d'une mutation.

Dans l'ADN Y, une mutation correspond à une valeur différente qu'a pris une certaine position dans le chromosome Y. Ce n'est pas nécessaire de savoir quelle est exactement cette position et il suffira de connaître le code de la mutation. Ainsi, M343 est une mutation. Par contre, cette mutation est survenue chez un homme dans la lignée masculine et chez un seul homme. Les générations suivantes ont conservé cette mutation.

L'étape suivante de l'évolution de ma lignée paternelle passe par M269, L51, P310, L151 ou P311, et P312. Nous sommes alors à environ 5 000 ans dans le passé avec l'haplogroupe R-P312.

Parmi les familles partageant ce R-P312, nous trouvons le tiers des trousseaux du projet French Heritage. En progressant dans le temps, nous atteignons les R-DF27 apparus il y a

## Le chromosome Y (test SNP)

environ 4 500 ans et 13% de ces trousseaux. À 4 250 ans arrive la mutation R-A431 avec seulement 0,5% des trousseaux.

En avançant dans le temps, de nouvelles mutations apparaissent et, à chaque fois, le nombre de trousseaux baisse. Il arrive un point où le nombre de familles concernées devient assez faible. Ainsi, notre R-A431 touche les Jarret dit Beauregard, les Pagé et les Tousignant, ainsi qu'un groupe de familles arméniennes. Il se trouve que cette mutation R-A431 est aussi testée par le laboratoire 23andme.

L'étape suivante date d'il y a 2 850 ans avec la mutation R-Y7363. Il est intéressant de constater qu'en 1 400 ans, nous retrouvons les mêmes familles. En d'autres mots, durant cette période, nous trouvons une seule lignée parmi les familles venues en Nouvelle-France. Ensuite, les branches commenceront à s'isoler, avec d'abord les Tousignant, puis les Pagé, et finalement le groupe arménien vers l'an 800. Finalement, il ne reste que la famille des Jarret dit Beauregard qu'on peut identifier avec l'haplogroupe R-FT166230<sup>20</sup>.

Voici la séquence plus complète:

```
R-M207>M173>M343>L754>L761>L389>P297>M269>L23>L51>P310>L151>  
P312>Z46516>ZZ11>DF27>ZZ12_1>FTT1>FGC78762>ZZ19_1>Z31644>  
FT318890>A431>A432>BY3288>FT12799>Y7365>BY36464>Y28672>Y7363>  
Y39310>Y41710>Y42559>FT249801>FT166230
```

### 5.3 Le test Big Y

---

Pour pouvoir catégoriser les descendants en ligne masculine, on peut faire un test de type STR Y. Il s'agit de mesurer entre 12 et 111 marqueurs et de trouver parmi les autres clients ceux qui ont des valeurs similaires. Un marqueur peut augmenter ou diminuer et cette modification est aléatoire. Il s'en suit qu'il est difficile de placer sur une échelle temporelle les résultats obtenus. Deux cousins peuvent avoir 111 marqueurs identiques même s'ils sont à 10 générations de l'ancêtre commun, mais deux autres peuvent avoir 5 marqueurs différents dans les mêmes conditions. Un marqueur pourrait augmenter à une génération et diminuer à la suivante.

Le test Big Y détecte des mutations SNP qui résistent au temps, qui apparaissent à une certaine génération et sont détectées par la suite chez tous les descendants. Cela donne une longue liste de mutations et il suffit ensuite de voir quelles mutations sont les plus fréquentes pour les classer comme plus anciennes vu qu'il y a plus de descendants qui les possèdent.

En d'autres mots, en comparant la liste des mutations trouvées chez un grand nombre d'individus, on pourra évaluer quand ces mutations se sont produites. Cela permet de

<sup>20</sup> <https://discover.familytreedna.com/y-dna/R-FT166230/path>

## Le chromosome Y (test SNP)

situer l'ancêtre commun à travers les âges. Les mutations sont tout de même aléatoires, alors l'échelle temporelle ne peut pas être figée sauf si elle s'appuie sur des lignées documentées ou des sources archéologiques. De plus, si un groupe de mutations est détecté chez plusieurs descendants, on ne sait pas dans quel ordre ces mutations se sont produites. Également, certaines mutations SNP ont été cataloguées par 2 laboratoires à la même époque, avec comme effet qu'elles ont deux noms. P311 et L151 donnent un tel exemple. Le test de FTDNA indiquera que la mutation L151 est détectée, alors que celui de 23andme l'appellera P311.

Pour trouver ces mutations SNP, FTDNA fait une analyse poussée du chromosome Y avec une méthode appelée « séquençage du génome complet » ou WGS<sup>21</sup>. Elle a pour effet de lire des segments de 100 valeurs de suite qu'on essaie ensuite de situer dans un génome de référence, puis on retient les valeurs différentes par rapport à cette référence.

Le SNP le plus récent (ou SNP apical) deviendra celui définissant le client. Ainsi, la mutation R-BY67760 trouvée chez des descendants de deux fils du pionnier André Jarret de Beauregard est la mutation familiale. La R-FT167459 trouvée chez deux descendants d'un même fils, mais pas chez le descendant de l'autre fils, est survenue plus tard et pourrait identifier une partie de cette branche de la famille, à partir de la génération où elle est apparue. Une association de famille ayant un budget illimité pourrait donc cartographier ces mutations mais le coût serait élevé car le test Big Y se détaille à environ 500\$ en 2023. Par contre, une fois le Big Y complété, un SNP apical peut être testé à la pièce pour 40\$ US chez FTDNA et 18\$ chez Yseq (qui offre un service « *Wish a SNP* »). Dans le cas de la lignée des Beauregard, FTDNA offre le SNP R-Y7363 seulement mais pas les plus récents.

### 5.4 Description des SNP Y

---

Les SNP Y (ou mutations sur le chromosome Y) sont des mutations qui obéissent à plusieurs règles.

Un SNP doit être stable et se retrouver chez plusieurs descendants. Il doit être détectable. Comme la partie centrale est difficile à analyser, les SNP se trouveront plutôt vers l'extérieur du chromosome. Une partie du chromosome X pourrait se recombiner avec le chromosome Y. Cette partie est donc éliminée.

Le résultat est tout de même impressionnant, soit environ 1 million de positions possibles. En décembre 2021, l'arbre de référence de FTDNA (Haplotype) montrait 50 000 branches, 460 000 variantes, et 214 000 clients testés.

---

<sup>21</sup> WGS pour *Whole Genome Sequencing*, avec plusieurs technologies comme la *NGS* ou *Next Generation Sequencing*

## Le chromosome Y (test SNP)

Il arrive souvent qu'on ne puisse pas séparer des séquences de SNP parce qu'on ne sait pas dans quel ordre les SNP sont apparus. Le critère principal est le nombre de descendants ayant ou non tel SNP. Si le nombre est plus petit, le SNP est plus récent. Ce sont les 50 000 branches.

Chez FTDNA, les SNP en soi sont au nombre de 460 000 et le nombre de SNP découverts continue d'augmenter régulièrement. On les appelle aussi des variantes. Une variante est privée si elle est trouvée chez une personne mais pas chez les autres de la même branche. Elle peut avoir une position et parfois un nom. Si elle a déjà un nom, c'est sans doute parce qu'elle était stable dans une autre branche. Sinon, il faut attendre d'avoir un autre client de la même branche pour que les SNP avec une position aient un nom. Sur d'autres sites de classement, on pourrait utiliser seulement la position.

### 5.5 Les noms des SNP

---

Les SNP Y sont formés d'un préfixe et d'un numéro. Le suffixe identifie le laboratoire ou le chercheur<sup>22</sup> ayant découvert le SNP, alors que le numéro est souvent séquentiel suite à la découverte du SNP.

Les SNP les plus anciens ont bien sûr un chiffre plus petit. Par exemple, M269 a été découvert par Peter Underhill, Ph.D. de l'Université Stanford, ainsi que tous ceux qui débutent par la lettre M. Voici les principaux codes décrivant les principales familles biologiques de nos ancêtres en ordre de SNP:

L-M20, E1b-M35, T-M70, I-M170, J2-M172, R-M173, O-M175, R1a-M198, G-M201, C-M216, I2-M223, N-M231, N-M232, Q-M242, I1-M253, J1-M267, R1b-M269, R1a-M512.

Les lettres ne suivent pas l'ordre de définition des SNP. Il faut se contenter d'y voir des étiquettes et ne pas y attacher une signification autre.

Lorsque FTDNA a lancé son programme Big Y qui est une analyse plus précise du chromosome Y, l'entreprise a ajouté ses propres codes que l'on verra souvent comme résultats d'un test Big Y. Ainsi, parmi les SNP de l'auteur, on peut voir

- A431 défini par Thomas Krahn, maintenant avec la compagnie YSEQ;
- BY3288 datant des tests Big Y 500;
- FT318890 et FT166230 apparus avec les tests Big Y 700;
- Y7363 défini par le laboratoire YFULL qui construit son propre arbre.

---

<sup>22</sup> <https://isogg.org/tree/>

## Le chromosome Y (test SNP)

La liste des SNP s'allongeant, la série FT comprend aussi FTA, FTB, FTC, FTD et FTE.

### 5.6 Les résultats d'un Big Y

Le client peut voir les résultats suivants en ce qui concerne un test Big Y:

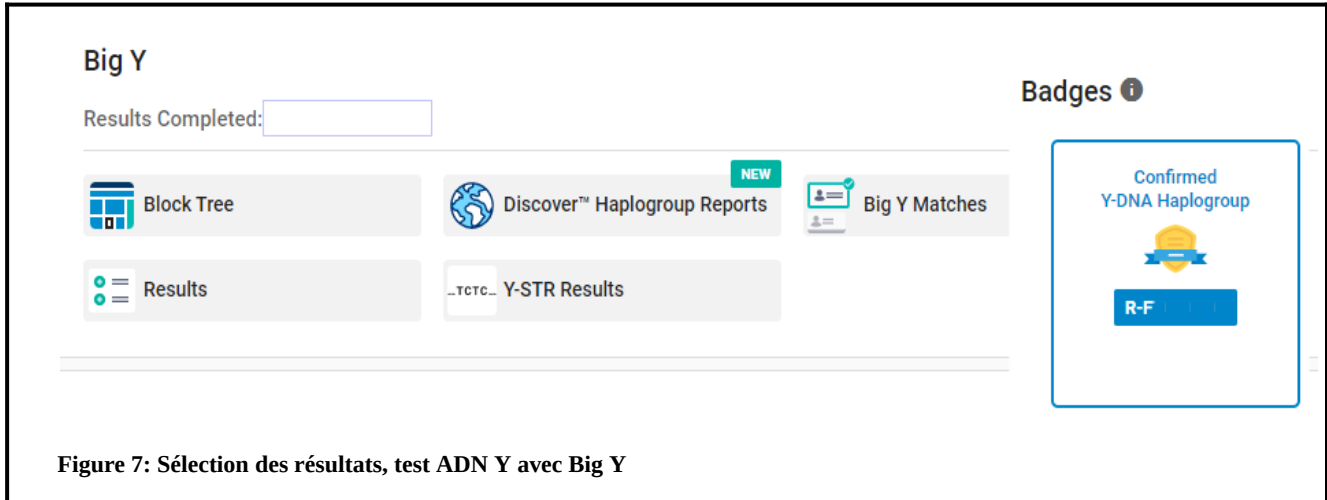


Figure 7: Sélection des résultats, test ADN Y avec Big Y

Le **Block Tree** est une structure sous forme de blocs regroupant les SNP équivalents.

Les rapports **Discover** donnent une série d'informations sur le test Big Y.

Les **matches Big Y** sont ceux ayant les résultats les plus proches.

Les **results** sont les valeurs lues par le test.

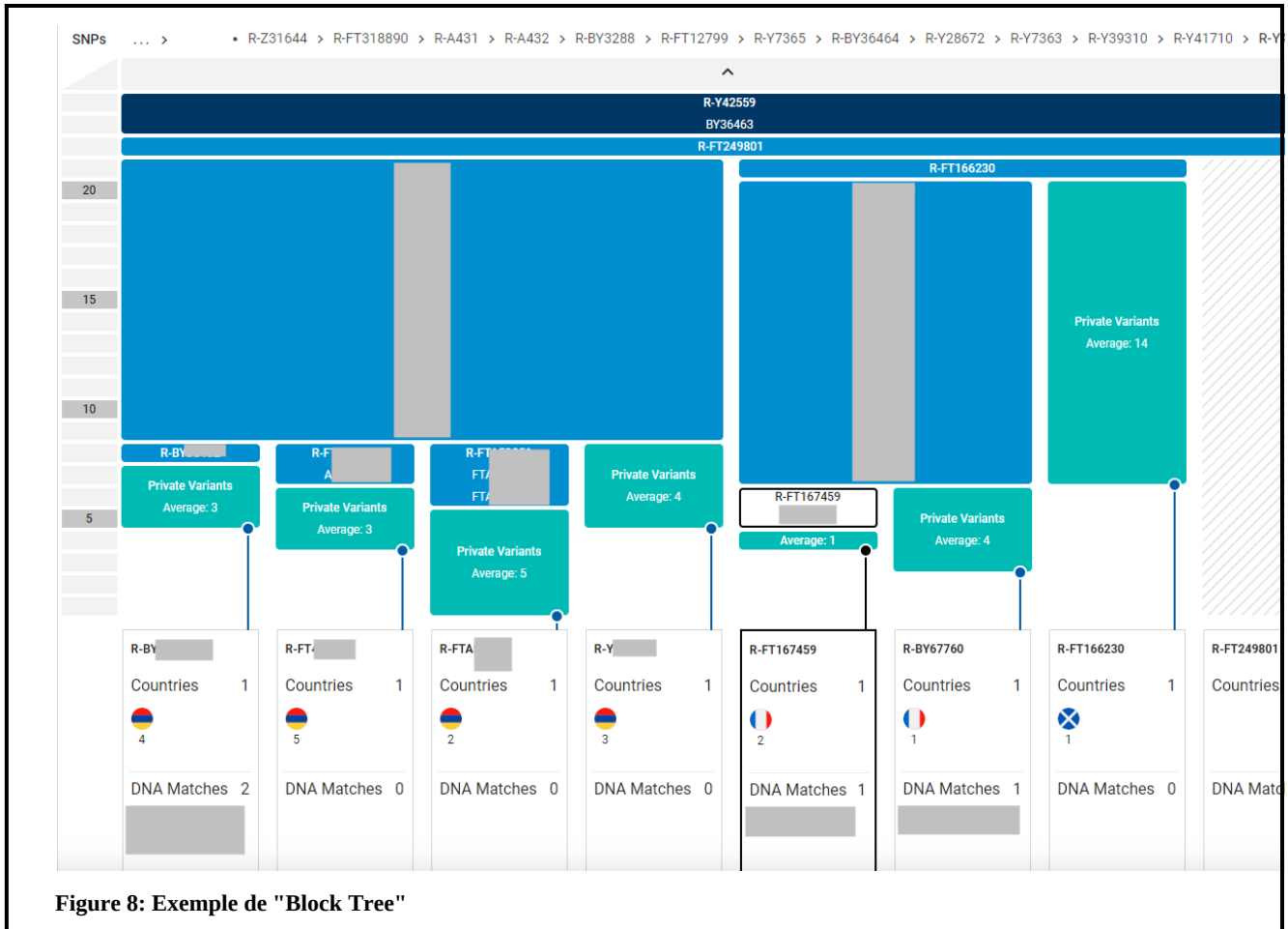
Les **résultats STR** sont parmi les 700 séquences STR programmées celles qui ont été découvertes durant ce test.

FTDNA a produit au cours des ans trois séries de tests : Big Y, Big Y 500 et Big Y 700. Les 500 et 700 correspondent au nombre de STR recherchés. Le 1<sup>er</sup> test ne faisait pas une telle recherche et quand le Big Y 700 est devenu disponible, le même algorithme a été appliqué et les 700 STR ont été recherchés parmi les tests Big Y 500 (et probablement les Big Y aussi). Plus le test est récent, plus la technologie permet de lire en profondeur le chromosome et de trouver de nouveaux SNP. En comparant un Big Y 500 avec un Big Y 700, il y a donc des parties non-comparables. De plus, la lecture se fait par bloc de 100 valeurs environ de suite plusieurs fois. Certains blocs de 100 ne sont pas lus.

Voyons cela plus en détails.

### 5.6.1 Le Block Tree

L'ensemble des SNP trouvés est placé dans une structure de blocs appelée *Block Tree*. Chaque bloc représente une série de SNP dont on ignore l'ordre d'apparition. Dans la séquence des résultats, on donnera le 1<sup>er</sup> SNP du bloc mais il est possible qu'un autre laboratoire change l'ordre des SNP ce qui donnerait un nom différent au bloc.



Pour la confidentialité, les SNP récents sont cachés par un rectangle gris.

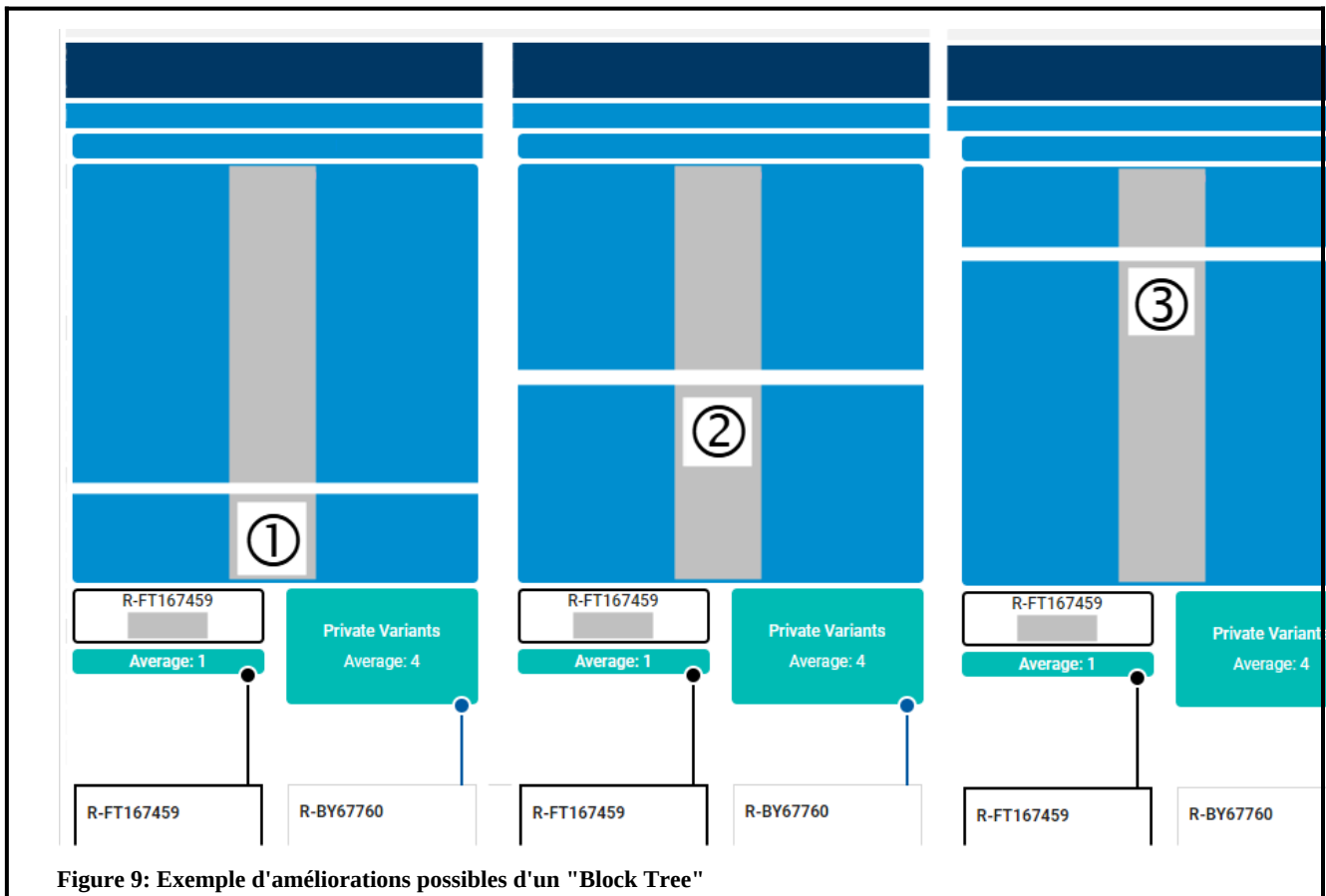
Le résultat du client est placé dans un rectangle blanc au bas de la page avec son SNP particulier. Dans ce cas-ci, un autre client assez proche a aussi fait ce test. Le premier rectangle en vert contient les variantes privées. Ce sont des variantes qui ne sont pas partagées avec les autres clients proches. Le petit bloc en blanc au bas désigne le client concerné et un autre client assez proche, les deux ayant le SNP FT167459. Un peu plus haut, le petit bloc vert montre la partie commune, soit une moyenne d'un seul SNP ou variant non-nommé. Suit un autre bloc blanc avec les SNP communs aux deux clients du

## Le chromosome Y (test SNP)

même bloc. Finalement, tout en haut du bloc bleu et caché par la zone grise, le SNP BY67760 marque le début d'une longue série et sert aussi de SNP apical pour l'autre client immédiatement à droite.

Ici, nous avons trois descendants d'un même pionnier de Nouvelle-France. Chaque client a donc ses résultats personnels jusqu'aux SNP partagés avec ses cousins les plus proches.

Au dessus se trouvent des blocs bleus qui contiennent les SNP communs à tous ceux qui sont au dessous de ces blocs bleus. Ainsi, le gros bloc bleu à droite concerne les 14 tests marqués avec un drapeau arménien. Celui au centre correspond à 3 descendants d'André Jarret de Beauregard. Le bloc vert qui suit mène à un Écossais ayant 14 variantes privées, puis une variante partagée avec ses cousins Beauregard. Finalement, la dernière colonne est une place réservée pour un futur client qui descendrait de l'ancêtre commun porteur du SNP R-FT249801.



Un nouveau client pourrait diviser un de ces blocs en deux. Ainsi, ① si un Jarret de la même région en France faisait un test Big Y, selon le nombre d'années vers l'ancêtre

## Le chromosome Y (test SNP)

commun, le 2<sup>e</sup> gros bloc bleu serait divisé en deux et le nouveau bloc bleu du bas devrait contenir un petit nombre de SNP parmi ceux déjà contenus dans le gros bloc.

Certains clients tentent de couper ce bloc en contactant d'autres clients. Ces blocs ne sont disponibles que si les clients ont fait un test Big Y. Mais on peut aussi trouver un match STR avec des cousins sans Big Y et dont le test pourrait couper ce bloc en deux. Comme les noms de famille apparaissent en France vers l'an 1200, un cousin Jarret biologique venu d'une autre région de France ② devrait avoir moins de SNP communs, donc le bloc au-dessus serait plus petit. Et si le cousinage se fait autour de l'an 1000 ③, on s'attend à ce que le nouveau bloc du dessus soit plus petit.

### 5.6.2 Les rapports Discover

C'est un ajout en 2023. Un ensemble d'outils permettant de mieux utiliser le résultat d'un test Big Y.

La colonne gauche propose ce menu que nous allons détailler.

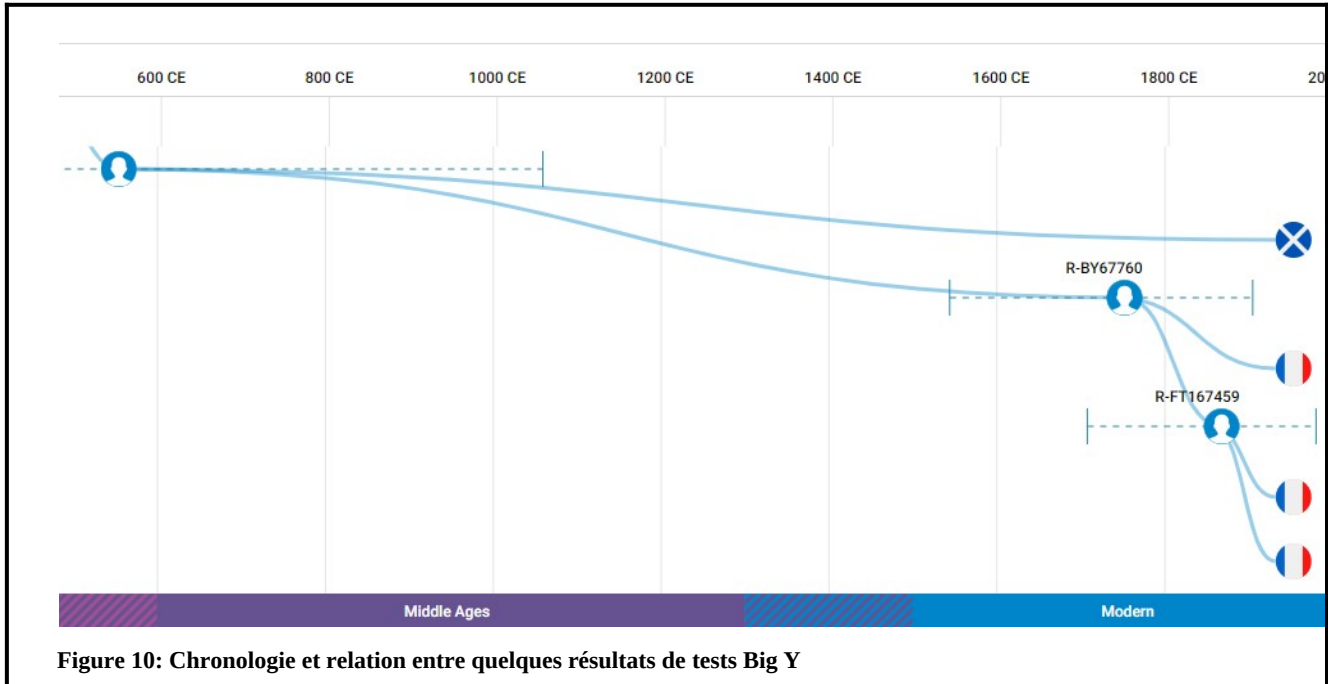
Haplogroup Story	<b>Haplogroup Story</b> - Il s'agit d'informations variées sur
Country Frequency	l'haplogroupe en soi, comme la branche plus ancienne, le
Notable	nombre de tests, et leur origine. Cette origine est celle indiquée
Connections	par le client et non son adresse postale.
Migration Map	<b>Country Frequency</b> - Une carte du monde met en lumière
Ancient	les pays d'origine des personnes testées.
Connections	<b>Notable Connections</b> - Il s'agit de célébrités qui ont une
Time Tree	partie de leur ADN Y en commun. Une indication comme
Ancestral Path	<i>Shared Ancestor - 1750 BCE</i> signifie que l'ancêtre commun a
Suggested Projects	vécu en 1750 avant notre ère. La parenté peut donc être
Scientific Details	lointaine. La liste de ces matches est dans le haut de l'écran.
About	Dans le cas de l'auteur, les connexions notables comprennent
	George Bush père et fils ainsi que Edwin Hubble (celui du
	télescope), (R-DF27), et un descendant d'Albert Perry qui a
	l'ADN Y le plus divergent et est de l'haplogroupe A00.

**Migration Map** - Cette autre carte du monde montre le chemin suivi par la lignée masculine depuis 230 000 ans, depuis le Y d'Adam en Afrique, jusqu'à certains haplogroupes populaires. Ainsi, l'auteur voit le chemin jusqu'à R-DF27 qui date d'environ 4500 ans,

## Le chromosome Y (test SNP)

**Ancient Connections** - Le lien se fait avec de l'ADN Y prélevé dans des sites archéologiques.

**Time Tree**<sup>23</sup> - C'est un graphique d'ensemble montrant les liens entre les différents résultats avec des nœuds représentant les ancêtres communs. Si l'on remonte plus loin dans le temps, on peut voir l'ADN Y provenant de sites archéologiques.



**Ancestral Path** - Cette fonction donne la série des mutations principales entre le testé et le premier Adam, ainsi que le moment des divergences menant vers l'homme de Néandertal et l'homme Denisovien.

**Suggested Projects** - Il s'agit des projets où l'on trouve des tests compatibles ainsi que le nombre de membres dans le projet et combien ont l'haplogroupe concerné. L'auteur y voit bien sûr les projets dans lesquels il a placé ses propres résultats. En remontant plus loin dans la séquence des SNP, la variété des origines augmente.

**Scientific Details** - Les détails scientifiques se résument à une courbe de probabilité au sujet de l'année où a vécu l'ancêtre commun à l'origine du SNP concerné.

**About** - Ce lien donne la source des informations utilisées pour la section **Discover**.

<sup>23</sup> Arbre de l'auteur: <https://discover.familytreedna.com/y-dna/R-FT249801/tree>

### 5.6.3 Les matches Big Y

---

La partie *Matching* est la plus intéressante pour un généalogiste. Si d'autres clients ont un nombre faible de variantes différentes, ils seront indiqués avec la liste de ces variantes. Il se peut qu'un client n'ait aucun match parce que l'ancêtre commun le plus proche parmi les autres clients est trop éloigné. Dans le cas de l'auteur, la liste comprend 4 matches, le plus ancien ayant un ancêtre commun il y a 1800 ans environ.

Cette partie est répétée dans un onglet de la section suivante.

### 5.6.4 Les results

---

Il s'agit des résultats bruts, soit la liste de toutes les valeurs examinées (et lues si la colonne *Derived* contient un +). On peut choisir les variantes nommées (qui sont trouvées chez d'autres clients), privées (qui ne se trouvent que dans nos résultats et non dans ceux des branches voisines), et correspondances (*Matching*).

### 5.6.5 Les résultats STR

---

Il s'agit des STR lus dans le cadre du test. Au début se trouvent les 111 valeurs de base du test STR qui accompagne normalement le Big Y. Puis, à partir de la position 112, les STR particuliers du Big Y.

Le lecteur notera que les chiffres sont généralement en bas de 10 parce qu'il s'agit d'une séquence trouvée dans une série de 100 valeurs. Si c'est un tiret, alors la séquence n'a pas été découverte durant le test.

À moins d'être un expert qui veut développer ses outils de prédiction, ces chiffres n'ont pas vraiment d'utilité pour le client.

## 5.7 Les SNP à la pièce

---

Les tests Big Y ne sont pas le seul chemin menant vers un haplogroupe.

Chez FTDNA, le client reçoit un haplogroupe prédit dès qu'il fait un test STR dont les matches ont des résultats cohérents. Dans le résultat des tests d'un projet, ces haplogroupes apparaissent en rouge.

Il est aussi possible d'acheter un SNP à l'unité, en cliquant sur son *badge* haplogroupe dans le profil du client. Après un certain temps, on voit apparaître l'arbre des SNP disponibles pour un achat à l'unité. Si un client a testé un SNP, celui-ci sera en vert, les SNP plus récents en bleu avec un panier et un bouton "Add" dans la marge gauche. Si le SNP n'est pas en vente (et ne se trouve que dans les Big Y), il sera en bleu sans panier et sans lien pour l'achat. Les liens présumés positifs sont sur fond orange et les liens présumés négatif, un fond gris.

## Le chromosome Y (test SNP)

FTDNA offre également des blocs de SNP. En haut du profil, le lien **Add Ons & Upgrades** est utilisé pour cela. La section **Additional Testing Options** au milieu de la page, ce qui affiche la liste de tous les groupes de SNPs disponibles. Le client ne devrait choisir cette option qu'avec la recommandation d'un administrateur de projet expert en la matière. L'expérience de l'auteur milite plutôt en faveur d'un test Big Y qui donnera un SNP beaucoup plus récent que les 2 ou 3 séries qu'on pourrait acheter.

Il est aussi possible d'acheter des SNP chez YSEQ, La présentation est différente puisqu'il faut choisir directement le SNP désiré. Ce laboratoire fait un test ADN et n'utilise pas les données d'un autre laboratoire. Il faudra donc acheter une trousse de prélèvement d'ADN. Le prix est moins élevé et il y a même un lien pour demander des SNP qui ne sont pas au catalogue (**Wish a SNP**). Le résultat ne pourra pas être exporté vers FTDNA et la comparaison se ferait donc à la main.

### 5.8 Les autres laboratoires (génomome complet)

---

FTDNA est le principal fournisseur de tests du niveau du Big Y. Il est toutefois possible de commander par d'autres sources. D'un côté, plusieurs laboratoires offrent des tests de séquençage complet (WGS) discutés plus haut. Le Big Y est dans cette catégorie. D'un autre côté, l'analyse du résultat avec son positionnement dans l'arbre généalogique ou phylogénétique se fait par un autre site, YFULL.

Deux tests sont offerts<sup>24</sup> : le chromosome Y seul ou le génome complet. Le milieu de la génétique généalogique suggère en général que ces tests sont plus précis que le Big Y, qu'ils trouvent plus de SNP, mais qu'il faut passer par une plate-forme intermédiaire avec beaucoup moins de tests pour la comparaison des résultats. À la longue, toutefois, les SNP découverts par ces autres laboratoires sont ajoutés aux SNP examinés par le test Big Y.

#### 5.8.1 L'analyse du chromosome Y

---

L'analyse du chromosome Y se fait en découpant celui-ci en morceau et en l'agrandissant chimiquement avec 100 positions à la fois, certains laboratoires offrant aussi 400 positions à la fois mais avec moins de fiabilité. Le nombre de répétition du test est indiquant comme l'agrandissement d'une lentille. Ainsi, si le laboratoire vise 30 lectures de chaque segment, le test sera annoncé comme 30X.

FGC offre un test appelé Y Elite et semble la seule à ne tester que le chromosome Y.

On peut trouver sur le site YDNA-Warehouse une comparaison entre ces différents tests du chromosome Y, sur la page <https://ydna-warehouse.org/benchmarks> mais cette

---

<sup>24</sup> [https://isogg.org/wiki/Y-DNA\\_SNP\\_testing\\_chart](https://isogg.org/wiki/Y-DNA_SNP_testing_chart)

page semble dater de plusieurs années car on ne voit que des tests Big Y 500 et pas de Big Y 700.

### 5.8.2 L'analyse du génome complet

---

Le but du séquençement complet du génome humain (WGS<sup>25</sup>) a souvent un but médical. Le format utilisé est toutefois similaire à celui des tests ne visant que le chromosome Y et les données sont donc utilisables dans ce but.

Parmi les principaux laboratoires offrant ce service, notons: Dante, FGC, Nebula. Faute d'avoir expérimenté ces laboratoires, ils ne seront pas décrits. Notons toutefois que ces tests donnent comme résultat un énorme fichier (autour de 50 Go ou davantage) et qu'il faut en général donner accès à ce fichier au laboratoire YFULL qui en fera l'analyse et la comparaison avec les tests de ses autres clients.

### 5.8.3 Les fichiers de résultats

---

L'analyse du génome complet donne en général deux sortes de fichiers : BAM<sup>26</sup> et VCF<sup>27</sup>. Le fichier BAM contient l'échantillonnage complet (une variante est appelée CRAM) alors que le VCF montre la différence entre les résultats et la référence. Le VCF est beaucoup plus petit. Il existe aussi un format SAM<sup>28</sup> qui est comme le fichier BAM mais sans comprimer les données, donc beaucoup plus gros.

Par ailleurs, le génome humain n'est pas une référence fixe. On trouve en 2023 deux références, Hg19 (alias GRCh37) et Hg38 (ou GRCh38). À mesure qu'on décode le génome de populations isolées ou provenant de sites archéologiques, de nouveaux segments peuvent apparaître. Les tests de type WGS doivent éventuellement être convertis vers l'autre norme en fonction des outils utilisés. YSEQ offre un tel service.

### 5.8.4 L'analyse par YFULL

---

YFULL n'est pas un laboratoire qui analyse la salive reçue mais plutôt le contenu de données transmises par les clients. En d'autres mots, le client fait le test avec d'autres laboratoires (comme le Big Y ou le FMS de FTDNA, ou les génomes complets de Dante, Nebula, FGC et peut-être YSEQ ou LivingDNA) et envoie les données ou un lien vers YFULL pour analyse. Il y a un prix selon ce qu'on envoie. Comme le site est en Russie, il faudra regarder comment faire le paiement.

---

<sup>25</sup> *Whole Genome Sequencing*

<sup>26</sup> BAM pour *Binary Alignment Map*, soit la carte des résultats alignés sur le génome de référence en version compressée

<sup>27</sup> VCF pour *Variant Call Format*, soit les variantes par rapport à la référence

<sup>28</sup> SAM pour *Sequence Alignment Map*, soit la carte des résultats alignés sur le génome de référence en version texte

## Le chromosome Y (test SNP)

YFULL peut analyser le fichier VCF ainsi que le fichier BAM. Les résultats les plus intéressants viennent de ce dernier et entraîne parfois la détection d'un nouveau SNP terminal. Alors que l'arbre Haplotype de FTDNA est basé surtout sur leurs propres tests Big Y et de l'ADN archéologique, celui de YFULL (ou YTREE) est construit à partir des résultats de plusieurs entreprises dont une partie des Big Y (au choix du client). L'arbre de FTDNA est plus complet et mise à jour à partir des résultats alors que celui de YFULL est basé sur moins de résultats. Comme il faut payer pour transférer les données d'un Big Y vers YFULL (FTDNA a diminué le prix de ses Big Y mais demande maintenant de payer pour avoir les données brutes), FTDNA demeure, et avec une bonne avance, le principal site avec des données Y SNP plus précises.

Une alternative est l'entrepôt de l'ADN Y (YDNA-Warehouse.org) qui n'a pas été mis à jour depuis plusieurs années et qui semble avoir été complètement refait. L'auteur n'y retrouve pas son propre haplogroupe R-Y7363 qui s'y trouvait il y a quelques années.

### 5.9 Les autres laboratoires (SNP de 4000 ans)

---

L'analyse du génome complet permet, en le comparant avec celui d'autres hommes, de le placer dans une structure couvrant l'ensemble de l'humanité ou arbre phylogénétique.

Il est toutefois possible d'examiner certains SNP dans le cadre d'un test autosomal. La puce d'analyse est programmée avec une liste de plus de 600 000 amorces et celles-ci peut comprendre des SNP Y intéressants. Le nombre de SNP Y est toutefois faible par rapport à l'ensemble des possibilités et le client obtiendra souvent un SNP qui date de 4000 à 5000 ans. C'est plus précis que par exemple le R-M269 datant d'environ 6400 ans mais il n'y a pas la discrimination que donnerait un test YSTR 37 pour distinguer les familles.

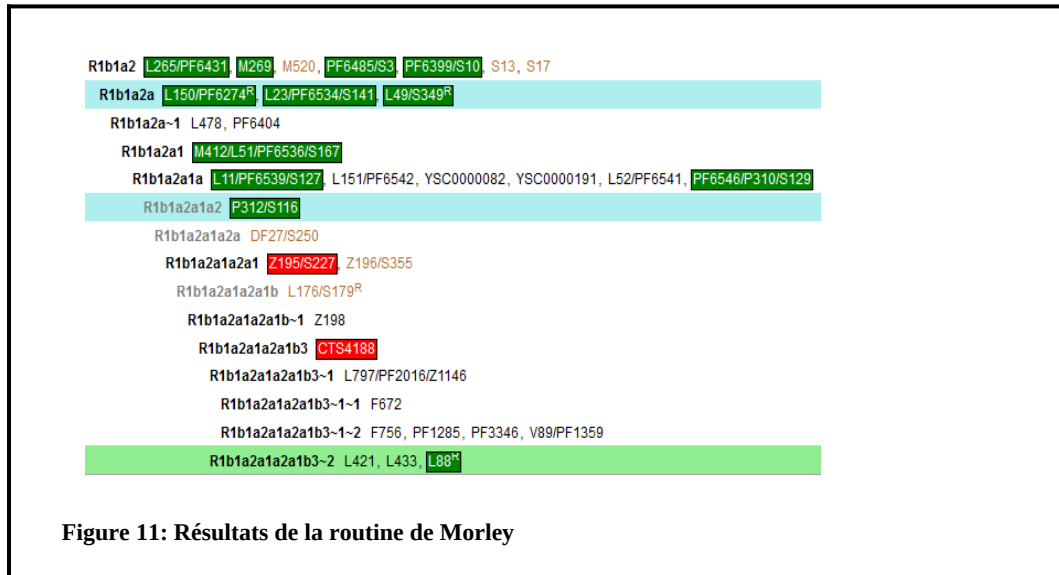
Dans ce groupe, le laboratoire le plus populaire est sans doute **23andme** qui donne directement un haplogroupe Y aux hommes ayant fait un test chez eux avec les versions les plus récentes. Un test avec la version 4 donnera un haplogroupe plus ancien et couvrant donc plus de familles différentes que la version 5. Cette information est aussi disponible si on examine la liste des *relatives* qu'on peut télécharger. Il s'agit par défaut des 1500 cousins les plus proches. Il n'est pas possible de rechercher les clients qui ont le même haplogroupe. Voici ce que l'auteur apprend à ce sujet, avec la version 5:

- Mon haplogroupe est R-A431.
- Un diagramme avec la migration de ma lignée, avec une échelle temporelle dont:
  - 27 000 ans pour R-M343
  - 10 000 ans pour R-M269

## Le chromosome Y (test SNP)

- Moins de 10 000 ans pour R-A431
- Ma lignée paternelle rejoint celle de Niall aux neuf otages (*Niall of the Nine Hostages*) bien que je comprenne mal l'opportunité de cette information car on n'a que R-M269 en commun, mais il semble que la position prédominante de sa dynastie lui a donné des millions de descendants en Irlande.

Le laboratoire utilisé par Ancestry donne aussi un haplogroupe Y mais il faut utiliser une routine externe pour le connaître, comme la routine de Morley et d'autres similaires.



Les SNP en vert sont ceux détectés par Morley. P312 est le plus récent. Z195 et CTS4188 ont été recherchés par AncestryDNA qui ne les a pas trouvés. L88\* signifie que la génération suivante est prédite (par le fond vert pâle) comme parallèle à R-L88, à une autre branche voisine. L'étoile est souvent utilisée sur les sites de référence de la même façon, pour indiquer une branche alternative.

LivingDNA, qui avait fait beaucoup de publicité il y a quelques années, semble être dans une situation semblable à celle de 23andme. Sa clientèle est surtout au Royaume-Uni et nous concerne moins.

## 6 Le chromosome MT (lignée maternelle)

---

Le chromosome MT est une façon abrégée de décrire l'ADNmt ou ADN des mitochondries. Pour alléger le présent document, le lecteur verra même MT tout court.

### 6.1 Aperçu

---

Les mitochondries sont des organelles vivant en symbiose avec l'être humain. Elles existent chez l'homme et la femme, dans les cellules. Elles sont transmises par les parents lors de la conception, mais comme il y en a très peu qui viennent du père alors que l'enfant est alimenté par la mère durant neuf mois, l'ADNmt est présumé provenir de la mère et donc suivre la lignée maternelle. Dans les rares cas où l'ADNmt venait du père, il y avait des problèmes de santé.

Si l'ADNmt trouvé chez une personne, homme ou femme, ne correspond pas à la signature trouvée en remontant sa lignée maternelle, c'est qu'il y a un problème documentaire dans cette lignée ou une adoption. Le plus souvent, l'erreur est décelée en étudiant attentivement la documentation. Parfois, elle découle des parents hypothétiques d'une fille dont les parents ne sont pas mentionnés lors du mariage et qui sont présumés d'après les témoins dans les actes de baptême ou mariage la concernant.

Alors que l'ADN Y repose sur un chromosome assez long (autour de 59<sup>29</sup> millions de positions dont 1 million sont utilisables pour la généalogie), l'ADNmt est tout petit, avec à peine 16 569 valeurs, plus ou moins 3 car il arrive que des valeurs soient insérées.

### 6.2 La signature MT

---

Comme pour le chromosome Y, les résultats de l'analyse d'ADNmt peuvent être catégorisés avec des haplogroupes et des haplotypes.

L'haplogroupe est surtout basé sur des études statistiques de l'ensemble des mutations (ou un alignement des mutations) plutôt que sur l'apparition d'une mutation précise comme avec le chromosome Y. Il peut y avoir des insertions et retraits de bases, ainsi que des remplacements. Le résultat est un code qui représente la variation théorique par rapport à l'ADNmt de l'Ève africaine ou européenne, avec une lettre suivie d'une alternance de chiffres et de lettres indiquant la hiérarchie des mutations consécutives. Ainsi, un haplogroupe appelé H3g1 serait issu de l'haplogroupe H, puis la 3<sup>e</sup> puis de la 7<sup>e</sup> puis de la 1<sup>re</sup> variation. Le nombre de variations à chaque étape peut être élevé. Ainsi, on trouve H109a et H1bu.

---

<sup>29</sup> <https://medlineplus.gov/genetics/chromosome/y/> indique plus de 59 millions, alors que <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC79676/> indique environ 60 MB.

## Le chromosome MT (lignée maternelle)

L'haplotype est la comparaison entre le résultat d'un test ADNmt et le génome MT de référence. Il existe deux génomes de référence, rCRS et RSRS. rCRS est la comparaison avec l'Ève européenne<sup>30</sup> et a constitué la norme assez longtemps parce qu'il s'agit du génome le plus fréquent trouvé durant les premières années de ces tests ADN. Par la suite, le nouveau protocole de référence RSRS, basé sur l'Ève africaine, est devenu la nouvelle référence chez FTDNA, bien que rCRS soit encore la référence dans les publications scientifiques et la plupart des laboratoires.

### 6.3 Les haplogroupes MT

---

Les noms des haplogroupes comprennent une lettre suivie d'une variation à partir de cette lettre. Dans certains cas, il y a deux lettres pour indiquer le plus souvent que 2 sous-groupes majeurs sont apparus.

Des noms typiques seraient H86 ou K2b1a. Le premier est la 86<sup>e</sup> variation de l'haplogroupe H. Le 2<sup>e</sup> cas est la succession de catégories définies une après l'autre, en suivant cette séquence:

K → K2 → K2b → K2b1 → K2b1a

Parfois, une mutation apparaît mais semble moins fréquente. Un haplogroupe peut alors se baser sur l'ajout ou la disparition d'un SNP. Ainsi, on trouve respectivement T2b3-C151T et HV0-T195C!.

T2b3-C151T indique l'haplogroupe T2b3 auquel s'est ajouté la mutation C151T. En d'autres mots, chez les clients ayant l'haplogroupe T2b3, la valeur lue à la position 151 est C, mais pour ceux avec T2b3-C151T, cette valeur est plutôt T.

De même, HV0-T195C! signifie que le SNP T195C (mutation de T à C à la position 195) normalement présent dans HV0 est ici absent, c'est à dire que la position est revenu à sa valeur d'origine: T.

Un autre cas déconcertant est celui de l'apostrophe. L'haplogroupe B4'5 provient d'un test ADNmt partiel. Il signifie que l'haplogroupe prédit (prédit parce que le test n'est pas complet, il faut le préciser) serait entre B4 et B5. Les tests plus précis ont donné B2, alors la prédiction n'était pas la bonne !

L'ensemble des lettres donne un aperçu hiérarchisé des haplogroupes et une prédiction de leur ordre d'apparition. La définition d'un haplogroupe à partir d'un autre plus ancien est souvent une question de statistiques, du nombre de descendants ayant

---

<sup>30</sup> Soit l'ADN du placenta d'une femme de Cambridge et légèrement révisé par la suite.

## Le chromosome MT (lignée maternelle)

l'haplogroupe en question. Comme les haplogroupes sont définis à partir d'une mutation précise, il pourrait même y avoir une mutation inverse pour une branche !

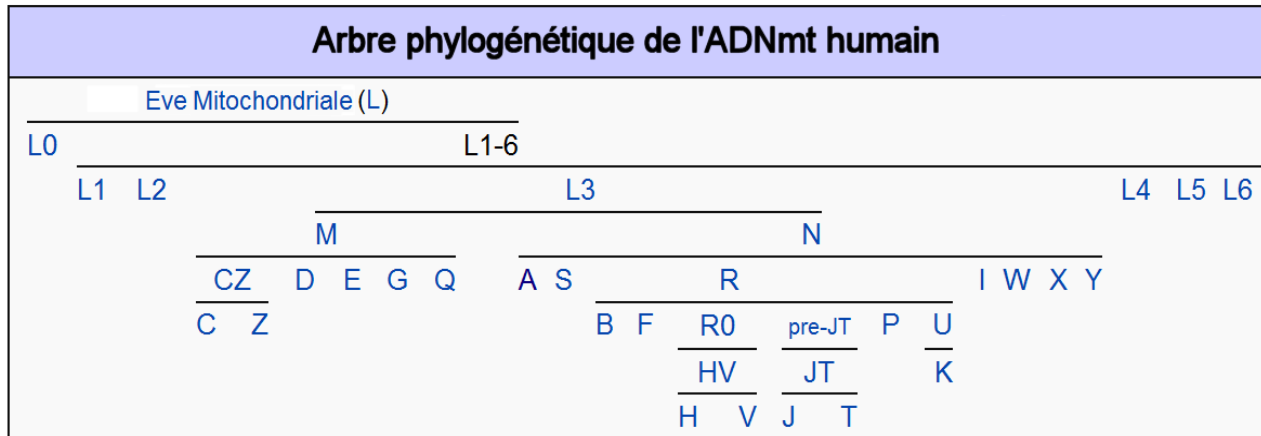


Figure 12: Arbre phylogénétique de l'ADNmt humain

On peut voir dans l'arbre phylogénétique que les haplogroupes H et V sont des descendants de HV qui est lui-même dérivé de R0 issu de R. Un test ADNmt partiel, n'explorant pas tous les SNP, pourrait donc attribuer à une personne l'haplogroupe R au lieu de H ou V, par exemple.

### 6.4 Les haplotypes MT

L'haplogroupe est basé sur une analyse des mutations en comparant avec une référence, RSRS<sup>31</sup> ou rCRS<sup>32</sup>. L'étape intermédiaire consiste à comparer les résultats du test avec une de ces références et d'indiquer les différences. Cela donne l'haplotype.

L'ADN de la mitochondrie est divisé en 3 zones : HVR1<sup>33</sup>, HVR2 et CR<sup>34</sup>.

Les premiers tests de FTDNA, appelés MT, couvraient la section HVR1, soit la partie supérieure du chromosome ou 3,6% de l'ensemble. Cette faible partie suffit en général à identifier la lettre définissant l'haplogroupe. Si on regarde des résultats chez FTDNA, on verra par exemple une lettre comme A ou H. Cela a donné la controverse sur Catherine Pillat. En effet, c'est une immigrante venue de La Rochelle. Dans les premiers résultats, elle avait l'haplogroupe A calculé d'après les haplotypes de la partie HVR1. Par la suite,

<sup>31</sup> RSRS pour *Reconstructed Sapiens Reference Sequence* ou Séquence de référence humaine reconstruite

<sup>32</sup> rCRS pour *Cambridge Reference Sequence* ou Séquence de référence de Cambridge révisée. Il semble qu'on aurait oublié une valeur dans la première version.

<sup>33</sup> HVR pour *Highly Volatile Region* ou Région hautement volatile.

<sup>34</sup> CR pour *Coding Region* ou Région de codage des protéines.

## Le chromosome MT (lignée maternelle)

des tests plus précis sont apparus et avec le génome complet, elle est devenue A10, qui est assez loin des A2 amérindiens.

FTDNA recommande de ne rendre public que les haplotypes des parties HVR1 et HVR2, soit 7% du total. La partie CR peut indiquer des problèmes médicaux et il est préférable de ne pas diffuser ce genre d'information trop sensible.

Chez **FTDNA**, les haplotypes suivant le protocole RSRS, proposé par défaut, suivent le modèle C150T qui signifie qu'en position 150, la valeur de la référence est C mais que pour ce client, on a lu T. En choisissant plutôt CRS, on verra pour le même résultat 150T, c'est-à-dire que la valeur de la référence n'apparaît pas.

Depuis un certain temps, la valeur lue est parfois en minuscule, comme A16183c. La nouvelle norme (les anciens résultats ne sont pas modifiés pour le moment) dit qu'il y a des majuscules pour une transition (A et G ou C et T) et une minuscule à la fin pour une transversion (les autres cas). Nous verrons plus loin les autres valeurs comme M, R, W, H, etc.

L'autre convention touche les insertions et retraits d'éléments. Il est en effet possible que le chromosome MT devienne plus long ou plus court. L'insertion d'un élément prend la forme 309.1C pour signifier qu'à la position 309, une nouvelle base C est apparue par rapport à la référence. «309.1C, 309.2C» signifie que 2 bases sont insérées. Le nombre d'insertions peut dépasser 10.

Chez **23andme**, dans le fichier des données brutes, le client pourra lire, par exemple:

```
i723150 MT 16129 G
```

Cela signifie que pour le test «i723150» concernant l'ADNmt, à la position 16129, on a lu G.

Un test donnant le **génomome complet** pourrait afficher chaque valeur lue, à la suite, comme:

```
GATCACAGGTCTATCACCTATTAACCACT
```

### 6.4.1 L'hétéroplasmie

Un seul chromosome Y est transmis par le père à son fils. Par contre, la mère donne plusieurs mitochondries à ses enfants et il arrive que ces mitochondries soient un mélange de 2 ou 3 variétés, c'est-à-dire qu'on voit plusieurs mutations différentes au même endroit.

Si le résultat du test prend la forme d'une série de valeurs lues, il y aura plusieurs fichiers. Si ce résultat est une suite d'haplotypes, alors on verra d'autres valeurs que A, C,

## Le chromosome MT (lignée maternelle)

G ou T comme nouvelles valeurs. FTDNA a publié une telle liste<sup>35</sup> sur son site web avec les valeurs trouvées.

### 6.5 Haplogroupes de quelques pionnières

---

L'ADNmt est utilisé pour obtenir des signatures de référence comprenant l'haplogroupe et l'haplotype HVR. Un catalogue s'est constitué en rassemblant ces signatures. En août 2023, à partir de 300 couples mariés au Québec en 1939-1940, nous trouvons 110 femmes sans signature, 173 avec une signature triangulée, 17 lignées simples non triangulées.

Voyez ce que nos pionnières les plus prolifiques ont obtenu comme signature ADN:

Pionnière(s) de Nouvelle-France	Descendance mariée <sup>36</sup>	% des Québécois qui en descendent	Haplogroupe	Triangulation <sup>37</sup>
Sœurs Langlois	3125	84%	H44a1	TRI0034
Sœurs Tourault	1985	75%	H1c15	TRI0092
Françoise Garnier	1552	76%	J2b1a1	TRI0009
Sainte Dupont	1475	84%	K1a4a1	TRI0272
Mathurine Robin	1429	80%	T2b3-C151T	TRI0045
Charlotte Maugis	1117	79%	H1ar1	TRI0036
Sœurs Lejeune (acadiennes)	926	31%	U6a7a1a	TRI0021
Marie Riton	850	64%	H5a1, H5a1-T16093C <sup>38</sup>	TRI0121
Sœurs Gaudet (acadiennes)	818	53%	J1b2	TRI0017
Louise Brodeu	807	28%	K2b1a	TRI0141

Les haplogroupes mentionnés sont tous décrits comme européens.

### 6.6 Les différents tests MT

---

Le test idéal analyse l'ADNmt au complet. Cela donne une liste d'environ 16569 valeurs parmi les lettres A, C, G et T.

<sup>35</sup> <https://help.familytreedna.com/hc/en-us/articles/4742569461263-Understanding-mtDNA-Heteroplasmy>

<sup>36</sup> Nombre de mariages de descendantes dans la base de données GFAN de l'auteur

<sup>37</sup> Selon le catalogue <http://www.francogene.com/triangulation/mt.php>

<sup>38</sup> Parfois, une mutation fait en sorte qu'il y a 2 haplogroupes officiels

## Le chromosome MT (lignée maternelle)

**FTDNA** est le principal fournisseur de tests ADNmt et le seul qui permette de rechercher d'autres clients avec des résultats similaires. Le laboratoire a d'abord offert un test simplement appelé MT et couvrant la partie HVR1, soit le haut du chromosome MT. Par la suite, le test MT Plus donnait les parties HVR1 et HVR2. Le test complet est devenu disponible sous le nom de *Full MT Sequencing* (FMS). Ensuite, les tests moins complets ont été retirés un après l'autre, mais pas les résultats.

La recherche de matches peut donc se faire avec le génome complet (FMS) et jusqu'à 3 variations, pourvu que l'haplogroupe soit le même, avec seulement les parties HVR1 et HVR2, ou avec seulement HVR1. Comme certaines variations sont rares et d'autres très nombreuses, il arrive qu'on voit parfois un petit nombre de correspondants et dans un tel cas, il y a de bonnes chances pour que l'on descende d'une pionnière commune. L'autre cas est possible, soit que le nombre de correspondants trouvés dépasse la capacité d'affichage et que cette recherche n'aboutisse pas.

Le laboratoire **23andme** offre également un résultat ADNmt depuis sa version 4, en plus de cette information sur les correspondants trouvés. Toutefois, il s'agit d'une prédiction basée sur environ 2000 SNP examinés lors de l'examen de l'ADN et non sur le chromosome complet. Le client pourrait donc recevoir un résultat comme H1e plutôt que H1e2, H3b et non H3b1a, H3 au lieu de H3s, etc. De plus, il affiche les haplogroupes d'autres clients mais ne permet pas de comparer l'haplotype ni de rechercher ceux qui ont un haplogroupe semblable. L'haplogroupe MT n'est pas importé vers FTDNA mais d'autres sites d'importation d'ADN comme Gedmatch ou Geneanet peuvent le conserver.

L'ancien test **Geno2** offrait également un haplogroupe basé sur environ 1500 SNP. Les mêmes remarques s'appliquent. Si le client a transféré son résultat vers FTDNA lorsque c'était encore possible, cet haplogroupe a suivi, tout comme pour le Geno1 (ou la première version de ce test de National Geographic) qui se limitait à la partie HVR1. Dans les deux cas, l'échantillon d'origine est perdu et il n'est pas possible de commander un nouveau test sans fournir un nouvel échantillon de salive.

Finalement, il est possible d'utiliser un test de nouvelle génération (NGS) discutés plus haut. Les données brutes sont souvent livrées avec le format FASTA qui est la suite des 16000 valeurs du chromosome MT.

L'analyse du résultat avec son positionnement dans l'arbre généalogique ou phylogénétique se fait par un autre site, YFULL. L'arbre MTREE de YFULL propose parfois un résultat différent par rapport à FTDNA et l'utilisation du même fichier. Ainsi, YFULL suggère H109a pour la pionnière Jeanne Barbier au lieu de H chez FTDNA.

## 6.7 La recherche de cousins MT

---

Dans le contexte de la Nouvelle-France, la recherche de correspondances MT sert souvent à confirmer la lignée menant vers un ancêtre, à détecter une erreur dans la documentation ou à débloquer une filiation pour un mariage perdu ou sans le nom des parents. Dans un contexte plus général, le test ADNmt montrera les migrations à l'échelle de l'humanité.

### 6.7.1 Confirmer la lignée menant vers un ancêtre (triangulation)

---

Pour confirmer une lignée, il faut une autre lignée et deux tests ADNmt cohérents. Cela s'appelle une triangulation.

Ce dernier terme s'explique par la façon de rechercher un avion avec des radars. Un premier radar indique la direction de l'avion depuis un point A. Le second radar fait de même pour le point B. L'avion se trouve au croisement de ces deux lignes.

Avec l'ADNmt, on trouve 2 lignées (ou plus) qui se rejoignent à un ancêtre commun et si l'ADNmt est cohérent, nous avons alors une triangulation et une signature MT de l'ancêtre commune la plus rapprochée (ACPR).

Des signatures sont considérées comme **cohérentes** si elles apparaissent dans les *matches* l'un de l'autre avec un test ADNmt complet (FMS chez FTDNA, même haplogroupe et pas plus de 3 différences), ou encore avec des tests HVR1 et HVR2 si les résultats sont peu nombreux. S'il s'agit de résultats venant de 23andme, cela dépend de la popularité des haplogroupes, seule information génétique disponible.

### 6.7.2 Détecter une erreur dans la documentation

---

Jusqu'à présent, les lignées MT ont toujours mené à

- une même pionnière s'il y en a plusieurs (triangulation);
- une pionnière unique, donc à un résultat non vérifiable;
- une impasse;
- une pionnière avec une signature non-compatible.

Plusieurs situations donnent une signature autre que celle prédite d'après la lignée. Il y a bien sûr une adoption récente (les clients ne disent pas toujours qu'ils veulent trouver des parents biologiques ou d'autres aïeux). Un tel cas est plutôt du ressort de l'ADN autosomal si l'impasse est à quelques générations seulement.

L'erreur documentaire est relativement fréquente quand on sort des mariages catholiques traditionnels avec le nom complet des deux parents. Par exemple, dans les mariages protestants ou américains avant 1900, il est fréquent de n'avoir que le père ou

## Le chromosome MT (lignée maternelle)

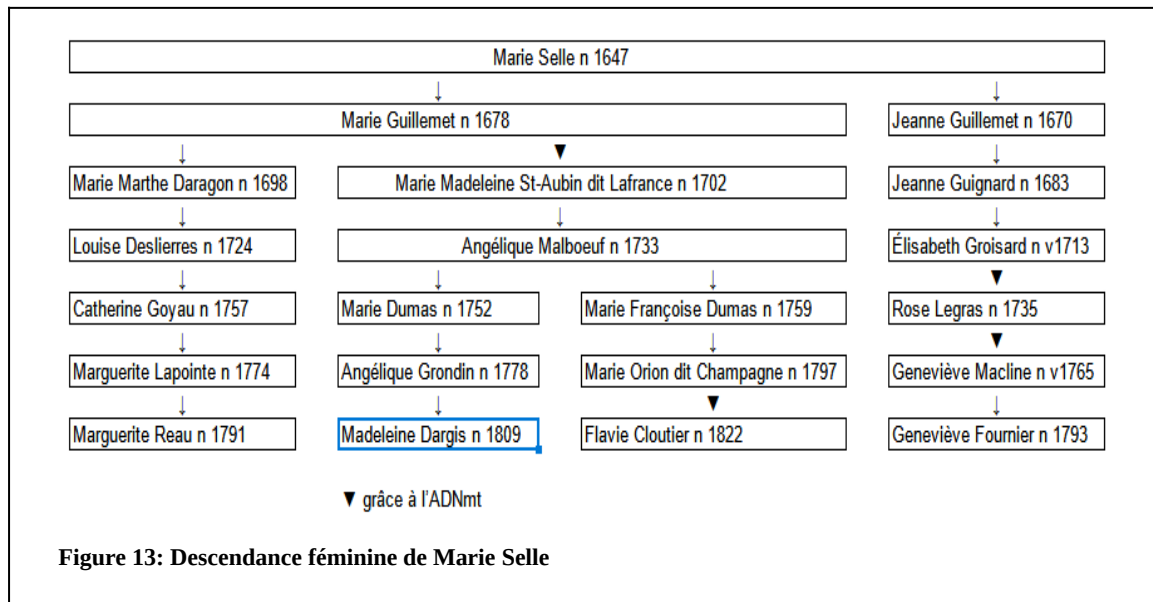
aucun parent dans l'acte original. Ceci est parfois contourné par les recensements où la mère a le plus souvent le nom de famille de son mari. Il s'en suit une génération qui prend la mauvaise direction et qui aboutit à une autre ancêtre. Dans un tel cas, il faut revoir chaque génération et trouver les points faibles et les alternatives possibles.

Il peut aussi s'agir d'une hypothèse pour une épouse qui ne mentionne pas de parents lors du mariage en se basant sur les témoins au mariage ou lors de baptêmes.

### 6.7.3 Débloquer une filiation

Si une lignée maternelle est bloquée à quelques générations seulement, la solution la plus pratique est de faire un test ADN autosomal, comme pour n'importe quelle adoption, en sachant que plus on remonte loin et plus il y a de possibilités, plus de cousins possibles. Nous verrons comment faire au chapitre approprié.

Si l'impasse est loin, ou en fait près du début de la colonisation, et que l'haplogroupe est suffisamment rare pour qu'il y ait une seule pionnière du nom, il est parfois possible de trouver la descendance féminine complète de cette pionnière jusqu'à l'époque de la naissance de l'orpheline et de voir les mères ou grands-mères potentielles. Voici un exemple.



Une première lignée sans surprise, mais avec un test moins précis HVR2 nous donne une première lignée de référence avec la séquence Marguerite Lapointe mariée à Detroit en 1790 avec Jean Baptiste Reau, puis Catherine Goyau, Louise Deslierres, Marie Marthe Daragon, Marie Guillemet (épouse de François Darragon dit Lafrance) et Marie Selle (TRI0326).

Une lignée avec U2d2 suit la séquence Marie Dumas qui épouse Nicolas Grondin en 1771, Angélique Malboeuf, et Marie Madeleine Saint-Aubin dit Lafrance. Selon le DGFQ<sup>39</sup>, p. 762, celle-ci est d'origine inconnue alors que ses parents sont connus, François et Madeleine Guillemé. Avec la même signature ADNmt, il est évident que le couple François Saint-Aubin dit Lafrance et Marie Madeleine Guillemé est le même que François Daragon dit Lafrance et Marie Guillemet.

Geneviève Macline épouse Augustin Fournier à L'Isle-aux-Coudres en 1787. Lors du mariage et d'autres actes, elle n'a pas de nom de famille. Dans plusieurs actes, elle se nomme Macline. Dans un seul acte, du 1793-03-31, son nom est Legras, et dans un deuxième acte du 1801-03-15, elle est nommée Legras, qui est rayé et remplacé en marge par McLean. Sa signature ADNmt est connue par un seul test qui lui donne U2d2. Dans les données disponibles, il y a seulement 2 autres résultats cohérents avec U2d2 et la documentation mène à Marie Selle, épouse de Nicolas Guillemet. Legras n'est pas un nom très répandu. Un seul couple Legras possède une signature U2d2, soit Jean Gabriel Legras et Élisabeth Groisard. Ce couple a 2 filles seulement, L'aînée n'a vécu que de 1733 à 1735. La cadette Marie Rose, née en 1735, a une destinée inconnue. Le DGQA<sup>40</sup> en fait la même que Barbe Lepras ou Legras décédée en 1736 à Beauport, et dont les parents sont omis. Comme le patronyme Legras est rare, ainsi que l'ADNmt U2d2, il est plus vraisemblable que Rose ne soit pas la même que Barbe et qu'elle ait survécu à l'âge adulte. Lors de la Conquête, elle a pu rencontrer un soldat McLean et lui donner comme enfant Geneviève qui portera le nom de Macline et deux fois, celui de Legras.

Finalement, une 4<sup>e</sup> lignée cohérente mène à Flavie Cloutier, fille d'Antoine et Marie Champagne. Le client fournit une lignée où cette Flavie est la fille du couple Antoine Cloutier et Marie Champagne marié en 1824 en Beauce. L'ADNmt de ce couple donne comme signature T2c, ce qui est trop différent. La solution est un autre couple, Antoine Cloutier et Marie Oriole dit Champagne, mariés à Nicolet en 1811 et dont l'ascendance mène bien à Marie Selle.

Avec 4 lignées cohérentes, les parents de 3 femmes ont été confirmés: Madeleine Saint-Aubin dit Lafrance mariée en 1721, Geneviève Macline en 1787 et Flavie Cloutier en 1843.

### 6.8 L'ADNmt de Catherine Pillat

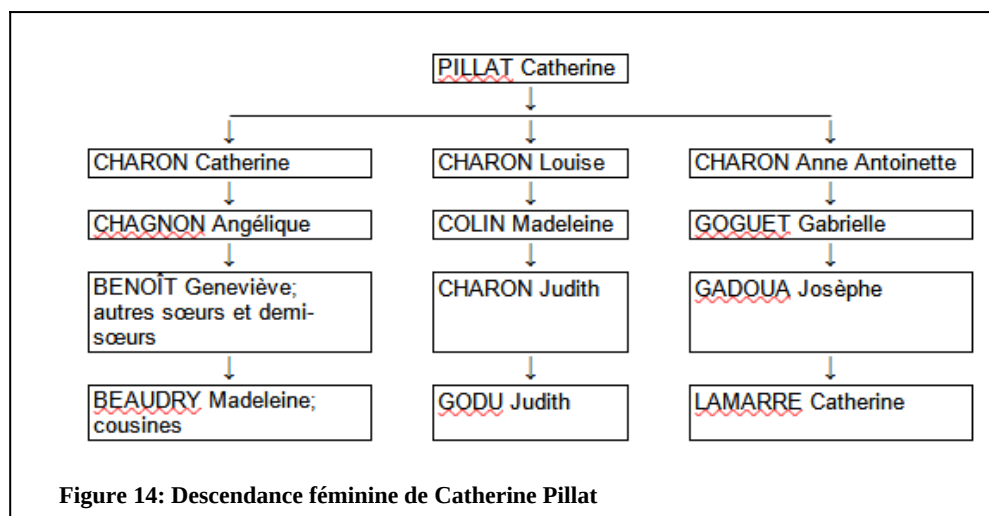
---

Voici les derniers échelons de la triangulation de Catherine Pillat, épouse de Pierre Charon (TRI0006):

<sup>39</sup> Dictionnaire généalogique des familles du Québec

<sup>40</sup> Dictionnaire généalogique du Québec ancien

## Le chromosome MT (lignée maternelle)



À l'époque où les tests ADN ont commencé à être populaires, l'ADNmt de Catherine Pillat a attiré l'attention. C'était des tests moins précis, basés sur 3,6% du chromosome MT (section HVR1) et le résultat était la simple lettre A. L'arbre phylogénétique nous apprenait, même à cette époque, que la lettre A désignait l'Asie. Comme on s'attend à ce que de l'ADN asiatique soit amérindien en Amérique du Nord, Catherine est devenue une ancêtre mystérieuse avec des origines qu'elle ne devrait pas avoir, surtout que son mariage indique clairement son arrivée de La Rochelle, un gros port d'embarquement pour la traversée de l'Atlantique à cette époque.

Le résultat étant déconcertant, plusieurs descendants en lignée matrilineaire ont refait ce test par curiosité et tous ont confirmé la lettre A initiale. Selon l'arbre de référence<sup>41</sup>, l'haplogroupe A est défini par la présence des mutations comme C16290T et G16319A trouvées dans le segment HVR1, commun aux descendantes d'Amérindiennes et de Catherine.

Ensuite, les tests sont devenus plus précis et une nouvelle catégorie a été définie autour de Catherine, la classe A10. La plupart des Amérindiennes étaient déjà identifiées comme A2, un haplogroupe défini quelques années plus tôt. On a aussi trouvé des A10 en Asie mais aucun parmi les populations autochtones d'Amérique. Alors que la première conclusion était que Catherine était amérindienne ou du moins métisse à cause d'une petite partie de son génome, le développement des tests a permis de corriger cette hypothèse pour en faire une Française avec des mitochondries venues d'Asie par un chemin inconnu. Ces A10 sont malgré tout assez rares et on ne les trouverait qu'en

<sup>41</sup> <http://phylotree.org/tree/A.htm>

## Le chromosome MT (lignée maternelle)

Sibérie<sup>42</sup> en ce moment. Il s'agirait donc d'une lignée venue d'Asie centrale à l'époque des invasions barbares (mongoles ou autres) plutôt qu'une lignée passant par l'empire romain.

Dans un tel dossier, avec une douzaine de personnes indiquant directement Catherine Pillat comme ancêtre ou dont la lignée est connue, en plus d'une vingtaine de résultats cohérents, la triangulation établie avec 3 filles de la pionnière a démontré que la signature obtenue était bien celle de Catherine. Les autres signatures A10 trouvées menaient toutes vers Catherine quand une lignée était établie.

Une autre information pertinente est la nature des correspondances trouvées lors du test chez FTDNA. Vu la nature même du moteur de recherche, à partir d'un client ayant A10, on ne peut trouver que d'autres clients ayant aussi A10. En juin 2023, la liste montre 24 cousins de cette façon. Tous ceux qui donnent une information généalogique pointent vers Catherine ou vers des ancêtres du Québec. En diminuant la précision du test, donc en regardant les parties HVR1 et HVR2, on augmente le nombre de cousins avec 2 informations contradictoires dont l'une ressemble à une erreur documentaire, et l'autre à l'acceptation de l'hypothèse, maintenant reconnue comme erronée, que Catherine Pillat était plutôt une Huronne Catherine Plat qui a vécu à la même époque. En contrepartie, quelqu'un avec A2 trouvera la même ancêtre ou d'autres Amérindiennes si le cousin a fait une recherche généalogique.

### 6.9 L'ADNmt des sœurs Langlois

---

Les sœurs Marguerite et Françoise Langlois sont sans doute les pionnières avec la plus importante descendance en ligne féminine. Si on peut évaluer la descendance d'un homme avec la fréquence de son nom de famille, il faut une autre méthode pour faire des statistiques similaires avec les femmes.

Ainsi, dans la base de données GFAN de l'auteur, fiche 3966, on trouve 3125 mariages de descendantes pour les deux sœurs. 84% des Québécois descendent de l'une ou l'autre. Marguerite Langlois épousa Abraham Martin, celui qui donna son nom aux Plaines d'Abraham, peu avant 1620 et donna naissance au premier Français né au Québec en 1621. 78% des Québécois en descendent (contre 79% pour sa sœur, donc à la limite de la précision des données disponibles). Sa sœur Françoise, épouse de Pierre Desportes, est peut-être retournée en France après la prise de Québec en 1629. La marraine de sa fille Hélène serait Hélène Boullé, la femme de Samuel de Champlain selon le testament de ce dernier, document qui permet également de déduire que les sœurs Langlois sont bien des sœurs.

---

<sup>42</sup> Voir la carte des distributions de A10 à <http://miroise.org/GparADN/catherine-pillard-amerindian/> et l'article de PILIPENKO et al. 2015. MtDNA Haplogroup A10 Lineages in Bronze Age Samples Suggest That Ancient Autochthonous Human Groups Contributed to the Specificity of the Indigenous West Siberian Population. PLoS ONE 10(5): e0127182. <https://doi.org/10.1371/journal.pone.0127182>

## Le chromosome MT (lignée maternelle)

Les deux sœurs ont bien sûr le même ADNmt et les descendants de l'une et l'autre se voient lors de la recherche de correspondants. Elles semblent avoir le monopole de leur haplogroupe H44a1. Sur la quarantaine de trousseaux avec cette étiquette, les 3/4 sont identifiées comme provenant d'une des deux sœurs. La présence d'une telle quantité d'informations permet de mieux évaluer si les mutations sont nombreuses à partir d'une même ancêtre (la mère des sœurs Langlois) afin de savoir si on peut prédire quand un haplogroupe s'est formé. Le résultat est renversant car 12 des trousseaux ont une mutation et 3 en ont deux. Sur une période d'environ 400 ans, les mutations sont trop nombreuses pour être prévisibles<sup>43</sup>. On voit donc que l'apparition des haplogroupes MT annoncée sur le site YFULL n'ont pas de fondement réel et qu'il faudra plutôt se baser sur les mitochondries trouvés sur les sites archéologiques pour calibrer la variation de ces haplogroupes.

### 6.10 Les sœurs Aucoin et l'hétéroplasmie

---

Sur 20 trousseaux de descendants des deux sœurs au projet French Heritage, 5 ont la mutation T16093Y<sup>44</sup>, 11 sont T16093C et 4 sont T16093T. En d'autres mots, parmi les descendants des sœurs Aucoin, filles de Barbe Minguet, la valeur lue à la position 16093 n'est pas la même selon le descendant. Cette valeur est normalement T mais dans 11 des 20 trousseaux, la lecture donnait C. Dans 5 cas, une partie des tests donnait la valeur C, et une autre la valeur T.

L'hétéroplasmie est un phénomène qui n'est pas fréquent. Lors du test ADNmt complet, on examine plusieurs fois l'ensemble des 16569 valeurs du chromosome. Il arrive que les valeurs lues soient différentes pour la même position. Cela signifie que certaines mitochondries ont subi une mutation et d'autres n'ont pas cette mutation. Si la quantité est suffisamment élevée (au moins 20%), le résultat portera une lettre autre que les quatre du vocabulaire de base de l'ADN, soit A, C, G et T. Ici, la lettre Y indique un mélange entre C et T.

Les sœurs Aucoin sont des pionnières acadiennes. Michelle serait née entre 1618 et 1625 selon les recensements anciens. Quant à Jeanne, son baptême en 1630 a été découvert à La Rochelle. Les tests ADN de descendants sont répartis entre les deux sœurs et l'hétéroplasmie trouvée dans deux tests de chacune des sœurs. Cela nous dit que l'hétéroplasmie peut survivre sur 400 ans, mais aussi que les deux sortes de mutations peuvent devenir la principale.

---

<sup>43</sup> Bien qu'il y ait un grand nombre de conceptions si on multiplie le nombre de générations par le nombre de tests, le produit est surévalué car il y a beaucoup de lignées qui se recoupent.

<sup>44</sup> Y est la notation pour C ou T, ou en d'autres mots, le test contient au moins 20% de T16093C et 20% de T16093T.

## Le chromosome MT (lignée maternelle)

L'haplogroupe des sœurs Aucoin est réduit à la lettre H. Formera-t-il une nouvelle catégorie dans l'avenir avec la poignée de variations par rapport au protocole rCRS? Espérons que oui.

## 7 L'ADN autosomal

---

L'ADN autosomal, trouvé dans les 44 chromosomes appariés humains, permet de trouver des cousins.

### 7.1 Aperçu

---

Le test ADN le plus utilisé examine l'ADN autosomal. Alors que l'ADN Y et l'ADNmt sont uniparentaux, transmis selon les lignées paternelles ou maternelles, l'ADN autosomal est reçu de chaque parent. Il permet, par comparaison avec d'autres personnes, d'évaluer un lien de parenté. En utilisant les plus grosses bases de données, c'est un outil pour retrouver des parents biologiques, parfois à plusieurs générations de distance. Une autre utilisation fortement utilisée pour commercialiser ces tests est l'aspect ethnique, bien qu'il soit impossible d'obtenir la précision parfois annoncée.

### 7.2 Les 23 paires de chromosomes

---

L'ADN humain nucléaire, c'est-à-dire contenu dans le noyau des cellules, comprend 23 paires de chromosomes. 22 paires sont similaires et comprennent chacune un chromosome paternel et un autre, maternel. La 23<sup>e</sup> paire (X et Y) est différente et sera expliquée en dernier.

Les chromosomes reçus par l'enfant sont formés avant la conception. Les 23 chromosomes du père sont dans un spermatozoïde alors que les 23 de la mère sont dans une ovule. La conception consiste alors à réunir ces 46 chromosomes. L'ADN reçu par l'enfant est la combinaison des chromosomes parentaux qui sont eux-mêmes des combinaisons de chromosomes grand-parentaux.

Dans un test ADN généalogique, les valeurs sont lues à l'unité, détectées par une certaine séquence (ou amorce). La puce de lecture ne peut pas, à ce moment, savoir si la valeur vient du chromosome paternel ou du maternel.

Par ailleurs, les chromosomes ont une longueur variable et sont numérotés en fonction de leur longueur. Le numéro 1 est le plus long et le 22, le plus court. Avec une centrifugeuse, on peut trier les chromosomes en fonction de cette longueur. Le processus complet est toutefois plus complexe et pour les fins de généalogie, les détails techniques seront peu précisés.

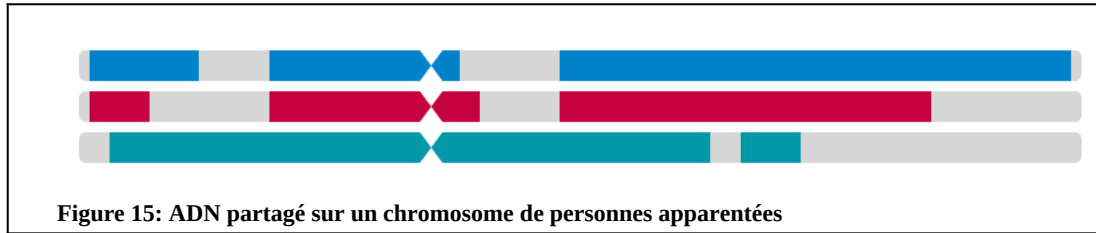
### 7.3 La combinaison des chromosomes

---

Chaque chromosome parental est une combinaison de deux chromosomes. Ce n'est pas un mélange. Il faut voir le chromosome comme formé par un premier segment d'un

## L'ADN autosomal

des chromosomes de la paire, puis par le reste de l'autre chromosome. Voici un exemple d'ADN partagé entre des personnes dont la parenté est proche:



Les deux personnes du haut sont des frères. Bien que les points de coupure semblent identiques, l'ADN partagé ne débute pas exactement au même endroit. Par ailleurs, plus on s'éloigne et plus les segments partagés deviennent petits ou se fractionnent. Pour évaluer la parenté entre deux personnes, on tiendra compte de l'ADN partagé au total, du nombre de segment et de leur longueur, en éliminant éventuellement les segments plus petits.

L'ADN n'est pas distribué de façon uniforme à chaque génération. Il peut arriver qu'un segment soit transmis complètement à la génération suivante, ou encore qu'un segment soit fractionné en plusieurs morceaux. L'ADN reçu d'un grand-parent peut varier entre 13 et 33%, au lieu de 25%.

De plus, à cause de l'endogamie (certains ancêtres apparaissent plus d'une fois parmi les ancêtres plus récents partagés par deux personnes), la relation réelle peut être plus éloignée que celle prédite par un test. Il est même possible que l'ADN commun proviennent de plusieurs couples d'ancêtres. Les algorithmes de calcul des relations sur les sites web d'ADN ne tiennent pas compte, en général, de cette endogamie.

### 7.4 Le test ADN autosomal

Le test ADN personnel se fait avec de la salive. Le laboratoire FTDNA envoie des bâtonnets que l'on frotte dans la joue alors que la plupart des autres laboratoires utilisent des éprouvettes où le client dépose un crachat. Pour améliorer les résultats, il est conseillé de ne rien manger dans l'heure précédant la prise de l'échantillon parce les aliments contiennent eux aussi de l'ADN qui viendrait corrompre l'analyse. De plus, le client peut frotter les joues pour dégager un peu plus d'ADN.

Certains laboratoires font une seule analyse et le résultat viendra donc de cette analyse. D'autres ont une panoplie de tests supplémentaires (autosomal, Y et MT, mais aussi des tests de différentes précisions ou avec un panneau de résultats).

Dans tous les cas, le test autosomal se fait de façon similaire, mais en définissant des positions différentes. Si le test d'une compagnie est comparé avec celui d'un autre

## L'ADN autosomal

laboratoire, il pourrait donc y avoir une inférence des données différentes ou encore l'omission des données différentes. Dans certains cas extrêmes, on pourrait même obtenir des résultats impossibles. Il y a aussi une perte de précision si l'on importe d'un laboratoire à un autre. Si la population de comparaison a fait peu de tests et qu'on accepte un partage faible, il est préférable d'utiliser le laboratoire préféré pour ce type de population.

Voici quelques exemples de tests ADN populaires<sup>45</sup>.

Compagnie	23andme	AncestryDNA	Family Finder (FTDNA)	myHeritage
SNP autosomaux	630 132	637 639	612 272	576 157
SNP Y	3 733	1691	Aucun, test séparé	3495
SNP MT	4 318	263	Aucun, test séparé	Aucun
Segments partagés	Si le client le veut	Non	Oui	Oui
Image des segments partagés	Oui	Non	Oui	Oui
Origines colorées	Oui	Oui	Oui	Oui
Clients (juillet 2023)	14M	23M	1,5M	6,8M
Arbre en ligne	Non (lien possible) <sup>46</sup>	Oui	Oui	Non
Importation	Non	Non	Oui	Oui
Méthode	Salive	Salive	Grattage de joue	Grattage de joue
Contact	Messagerie interne	Messagerie interne	Courriel	Messagerie interne

**SNP autosomaux** - Le nombre de positions varie. De plus, environ 2% des valeurs ne sont pas lues. Un tableau de l'ISOGG<sup>47</sup> permet de comparer le nombre de positions similaires. Ainsi, la version 2 de FTDNA a 538 746 positions communes avec la version 5 de 23andme et 191 769 avec la version 2 d'AncestryDNA. En plus des positions examinées, il peut y avoir un changement des paramètres pour trouver des matches. Ainsi, depuis 2016, FTDNA exige pour un match de moins de 20 cM, au moins un segment d'au moins 9 cM.<sup>48</sup> En effectuant le test de nouveau, on pourrait obtenir de nouveaux résultats avec la même base de données, mais pour des cousins éloignés.

<sup>45</sup> [https://isogg.org/wiki/Autosomal\\_DNA\\_testing\\_comparison\\_chart](https://isogg.org/wiki/Autosomal_DNA_testing_comparison_chart)

<sup>46</sup> 23andme génère un arbre en ligne à partir des cousins les plus proches mais ne le partage pas avec les autres clients.

<sup>47</sup> [https://isogg.org/wiki/Autosomal\\_SNP\\_comparison\\_chart](https://isogg.org/wiki/Autosomal_SNP_comparison_chart)

<sup>48</sup> <https://cruwys.blogspot.com/2016/05/new-match-thresholds-for-family-tree.html>

## L'ADN autosomal

**SNP Y et MT** - Certains laboratoires ajoutent des données sur l'ADN Y et MT. Avec 23andme, les haplogroupes sont disponibles directement dans les résultats des clients et pour environ la moitié des correspondants (ceux qui ont fait les versions les plus récentes). Avec AncestryDNA, il faut utiliser un site extérieur pour connaître l'haplogroupe Y. Avec FTDNA, c'est l'haplogroupe des clients qui ont fait des tests Y ou MT que le client verra parmi ses correspondances.

**Segments partagés** - La position exacte des positions partagée permet parfois de découvrir la relation entre deux correspondances et, par la suite, de documenter la lignée concernée. Chez FTDNA, cet outil est appelé Chromosome Browser. Si on ne choisit aucun client, on peut obtenir la liste des segments de tous les correspondants. Avec 23andme, cette liste est limitée aux 1500 plus proches, et quand un nouveau match s'ajoute, le plus loin disparaît mais est encore visible si on a conservé les liens (dans le fichier *Relatives*). Il n'y a que Ancestry qui n'offre pas cette information.

**Origines colorées** - Le *Chromosome Painter* montre pour chaque segment d'ADN l'origine ethnique supposée suite à la comparaison avec les autres clients du laboratoire.

**Clients** - Le généalogiste qui essaie de résoudre une énigme avec l'ADN autosomal aura plus de chances de trouver quand il y a plus de clients. En France, toutefois, le premier choix serait myHeritage qui a fait de la publicité et expédié des trousseaux de 2018 à 2022.

**Arbre en ligne** - Certains laboratoires permettent aux clients la saisie directe d'un arbre généalogique avec des limitations selon le cas, par exemple si un ancêtre apparaît 2 fois (endogamie) ou si on veut saisir un grand nombre d'informations. L'importation d'un fichier Gedcom est souvent permise. 23andme et FTDNA permettent toutefois un lien vers un arbre extérieur. L'accès à cet arbre est parfois réservé aux correspondants ou aux abonnés ou encore selon le bon vouloir du client.

**Importation** - Quelques sites permettent d'importer des données ADN autosomales. En plus de FTDNA et myHeritage, on peut ajouter GedMatch et Geneanet. Si les correspondances sont faibles, elles peuvent disparaître suite à l'importation, tout comme d'autres peuvent apparaître. Cela ne devrait pas affecter les plus gros *matches*. Gedmatch et Geneanet peuvent aussi conserver les haplogroupes, alors que FTDNA n'importe que la partie autosomale.

**Méthode** - Le test de myHeritage se fait chez le laboratoire FTDNA, ce qui explique qu'ils utilisent la même façon de recueillir de la salive en grattant l'intérieur des joues. Cette méthode peut s'avérer utile avec une personne qui n'arrive pas à fournir assez de salive. FTDNA offre aussi un grand nombre de tests qui peuvent être effectués à partir du

même échantillon. Si on a testé le Y d'un oncle décédé, par exemple, on peut demander le test Family Finder ou celui sur l'ADNmt.

**Contact** - Avec certains laboratoires, il faut utiliser la messagerie interne pour communiquer et le client fournit peu d'informations. Parfois, cela peut prendre des mois avant d'avoir une réponse.

### 7.5 La recherche de cousins

---

La plupart des généalogistes utilisent un test ADN pour rechercher des cousins ou débloquer des lignées. Le test ADN autosomal livre alors au client une liste de correspondants partageant de l'ADN. Selon le laboratoire, le résultat est affiché de différentes façons. De plus, si vous trouvez le même cousin dans deux laboratoires différents, l'ADN partagé pourrait être différent.

Deux mesures sont utilisées pour déterminer l'ADN commun : les cM et le pourcentage.

Le cM ou centiMorgan est utilisé par FTDNA, Ancestry et plusieurs outils. C'est une mesure dont la valeur absolue peut changer selon le laboratoire. Ainsi, la relation entre un parent et un enfant est de 3564 chez FTDNA (en 2023), 3489 chez Ancestry et 3720 chez DNA Painter<sup>49</sup> (un site qui évalue les relations entre les personnes).

23andme montre seulement le pourcentage, mais on peut calculer les cM (ou se servir de DNA Painter qui utilise les mêmes chiffres). Si on lit le fichier "Relatives"<sup>50</sup> dans un chiffrier, on remplace les points par des virgules (avec un chiffrier fonctionnant en français) et on ajoute une colonne comme :

$$=7439*T1$$

Il est souvent préférable d'utiliser les cM plutôt que les pourcentages. Le risque d'une erreur est plus faible si on travaille avec des entiers qu'avec des fractions. Une bonne partie de la littérature utilise aussi les cM pour expliquer les relations possibles entre deux personnes. Ainsi, DNA Painter situe une relation entre deux frères ou sœurs à une valeur entre 1613 et 3488 cM, et pour des demi-frères, à 1160 et 2436. 1613 cM vaut 21,68% sur ce site web.

Le test autosomal donnera donc une liste de cousins et leur ADN partagé, ainsi que d'autres informations variées selon le site utilisé, selon le bon vouloir du gestionnaire des résultats ou selon la version du test. Le défi consiste alors à utiliser ces informations pour reconstruire la généalogie du cousin découvert de cette façon. Nous verrons plus loin des exemples en fonction des différents laboratoires.

---

<sup>49</sup> <https://dnapainter.com/tools/sharedcmv4>

<sup>50</sup> <http://francogene.com/adn/23andme.html>

## 7.6 Le type de cousins trouvés par ADN

---

La relation entre deux personnes dépend de plusieurs facteurs. L'année de naissance permet de distinguer entre un parent ou grand-parent et un enfant ou petit-enfant. Il est possible qu'un oncle soit le dernier enfant de sa famille et qu'il ait 30 ans de moins que l'aîné qui pourrait avoir eu plusieurs enfants avant la naissance de cet oncle. De même, un cousin éloigné pourrait se trouver deux générations plus haut ou plus bas dans un arbre généalogique. L'ADN seul ne permet pas de distinguer directement ces relations. Selon le contexte, on peut préférer tel scénario pour expliquer une relation, mais parfois, il y a une autre possibilité. Ainsi, il peut s'avérer un cousin est en fait un demi-cousin, que son père légal n'est pas son père biologique. Il est aussi possible d'avoir un nombre de cM ne permettant pas de trancher une relation; ainsi, il y a une certaine plage de résultats où on ne sait pas si nous avons des sœurs ou des demi-sœurs. Il faut savoir qu'il s'agit de statistiques. L'ADN est hérité au hasard et il se peut que l'on tombe sur un cas marginal.

Voyons quelques cas en fonction des cM partagés, tirés de résultats réels de tests faits chez FTDNA. La valeur proposée par DNA Painter est aussi indiquée. Dans ce cas, les valeurs extrêmes proviennent de données réelles après élimination du maximum et du minimum. Plus on est au centre des résultats et plus ceux-ci sont probables. Par contre, si on est à la limite, il est possible mais peu probable d'avoir une valeur plus éloignée à cause d'une endogamie éventuelle, si un ancêtre commun apparaît à une faible distance dans la généalogie des deux personnes comparées. Le lecteur trouvera sur le site DNA Painter un tableau assez complet des valeurs prédites (en cM) pour différentes relations de cousinage entre deux personnes et en cliquant sur une case, sur des blocs de 100 valeurs avec le nombre de données pour ce bloc.

**Parent et enfant** - 3551 à 3564 cM (FTDNA, données réelles). Selon DNA Painter, la plage serait 2376 à 3720. On a reçu 2412 résultats, dont 2221 entre 3400 et 3700, et 142 en 3800 et 3899. Il est probable que les résultats utilisés viennent surtout d'un seul site avec plus de clients et d'un second qui utilise une valeur différente pour calculer l'ADN commun. Les autres valeurs découlent en fait des courbes de statistiques. Les 14 plus basses pourraient être des cas où le père légal n'est pas le père biologique mais où le client ne le sait pas, ou encore de résultats ADN contaminés suite à une greffe de moelle ou pour une autre raison médicale, ou bien de l'importation d'un test ADN partiel.

**Frères ou sœurs et demis** - 2450 à 2840 cM (FTDNA, données réelles pour des frères). Selon DNA Painter, la plage serait 1613 à 3488. La plage pour une demi-relation (avec seulement le père ou la mère en commun) serait 1160 à 2436. Dans les deux cas, les valeurs extrêmes proviennent de données réelles.

**Demi-relation** - Une demi-relation signifie que la parenté se fait par seulement l'homme ou la femme d'un couple et non par les deux conjoints. C'est le cas par exemple pour un enfant né hors-mariage.

**1C1R** - Les relations simples sont données par un terme simple : oncle ou tante, neveu ou nièce, cousin ou cousine. Quand les relations sont plus éloignées, les termes utilisés montrent plutôt le niveau de cousinage (1C pour des grands-parents en commun, 2C pour la génération suivante) suivi d'un décalage pour une des personnes (1R pour une génération de plus ou de moins, 2R pour deux générations). DNA Painter montre les valeurs prédites en fonction du type de relation. Zéro est utilisé comme minimum puisqu'on ne peut pas mesurer moins que zéro cM. Dans ce graphique, une relation 3C (on remonte et on descend 4 fois) est près d'une relation 2C2R (on remonte 3 fois et on redescend 5 fois), ce qui indique qu'en fait, l'ADN va donner le nombre de générations séparant deux personnes. Les chiffres affichés montrent que le taux d'erreur (le nombre de cM obtenus par le test par rapport à la moyenne mathématique), pour une relation 3C ou 2C2R, peut être de deux fois la valeur lue. On arrive donc rapidement à des chiffres où il est impossible de prédire exactement une relation.

**1st Cousin - 3rd Cousin** - Les laboratoires donnent souvent une plage de résultats au lieu d'une seule prédiction. Ainsi, FTDNA proposera « *1st Cousin - 3rd Cousin* » alors que 23andme suggérera une relation plus précise « *2nd Cousin Once Removed* » même si la vraie relation est légèrement différente. Ancestry, ayant une version française, donnera les résultats dans cette langue, également en proposant une plage de relations : « *Cousin issu de germain – Petit-cousin* ».

Les données de DNA Painter viennent d'une sollicitation publique et n'ont pas été validées individuellement. Pour cette raison, il peut y avoir un certain décalage par rapport à ce que le test ADN donnera comme résultat. En plus, dans une population endogamique, avec un nombre relativement petit de fondateurs (comme en Nouvelle-France), le résultat peut être altéré parce qu'il y a plusieurs ancêtres communs, parfois apparaissant sur plusieurs générations.

De plus, le nombre de segments et le segment commun le plus long peuvent modifier la prédiction donnée par un test ADN. Ainsi, FTDNA aura des résultats comme les suivants :

Relation	cM partagé	Segment le plus long
2nd Cousin - 4th Cousin	31	31
3rd Cousin - 5th Cousin	82	19

Même si l'ADN commun est plus grand, FTDNA éloignera une relation si le plus long segment commun est plus petit, afin de réduire l'effet de l'endogamie. Pour la même raison, les plus petits segments partagés ont été éliminés pour le calcul de la relation. D'autres laboratoires pourraient en tenir compte et la relation entre deux personnes changera selon le laboratoire.

### 7.7 Confirmer un cousinage avec l'ADN

---

L'ADN permet de comparer la relation biologique entre deux personnes alors que la généalogie traditionnelle, à base de documents, peut trouver un lien différent. Ainsi, si l'on a trouvé deux couples homonymes vivant dans la même région à une époque récente, l'ADN peut aider à identifier le bon couple. Il devrait aussi identifier les cousins adoptés par rapport aux autres. Certains ont aussi découvert que leur frère était un demi-frère.

Il est préférable de faire des tests de comparaison avec le même laboratoire. Certaines situations donnent une zone grise entre par exemple un frère et un demi-frère. En important les données depuis deux laboratoires vers une base de données commune, on perd alors un peu de précision. Mais le point le plus important est évidemment d'avoir les résultats des deux personnes au même endroit.

#### 7.7.1 Comparer des frères et sœurs

---

Selon le tableau de DNA Painter, la relation entre des frères et sœurs est entre 1613 et 3488 cM, alors que pour des demi-frères et demi-sœurs, c'est plutôt entre 1160 et 2436 cM. Ce sont des données statistiques basées sur les contributions de milliers de personnes. Il est possible mais peu probable que les résultats soient un peu plus élevés ou faibles que ces chiffres. De plus, les laboratoires ne définissent pas les cM de la même façon. Si les résultats sont nettement à l'intérieur de la zone indiquée et conforme à la documentation, alors, la relation est confirmée, par exemple avec 2600 cM ou plus pour une fratrie.

Si on est dans la zone grise (ici, entre 1600 et 2440 cM), il faut utiliser une autre méthode que le nombre de cM. On regarde les plus grosses correspondances trouvées dans la base de données de comparaison, ceci au même moment. Si l'une des personnes a fait son test en 2015 et l'autre en 2020, il se peut qu'un cousin proche ait fait son test en 2018 et qu'il ne soit donc pas trouvé en 2015. Il faut donc faire la comparaison. Les cousins trouvés avec plus de 100 cM devraient apparaître pour les deux personnes souvent avec des valeurs similaires. En bas de 100 cM, l'effet du hasard dans l'ADN hérité commence à se faire sentir. Un match de 70 cM chez un frère pourrait être omis chez l'autre frère.

Si deux frères ont un résultat pour l'ADN Y, l'haplogroupe obtenu devrait être cohérent. L'haplogroupe est en fait le SNP Y le plus récent parmi ceux examinés par le

laboratoire. Il est possible que les deux frères aient des SNP pour des étapes différentes sur la ligne chronologique. Si les haplogroupes sont non compatibles (par exemple, R-M269 et G-M201), alors ils n'ont pas le même père biologique. Les femmes n'ayant pas de chromosome Y, cette comparaison n'est possible que pour des frères.

De même, si des frères ou sœurs ont des résultats pour l'ADNmt, l'haplogroupe MT devrait être cohérent. Dans cette comparaison, la précision du test est un facteur important. FTDNA a offert 3 tests MT appelés *MT*, *MT Plus* et *FMS (Full MT Sequence)*. Les deux premiers vont souvent donner une seule lettre qui devrait être la lettre initiale du test FMS. 23andme et Geno2 font un échantillonnage de l'ADNmt et il faut alors comparer avec le début d'un FMS de FTDNA. Si par hasard on utilise le tableau de Yfull, une variation est encore possible. C'est pour toutes ces raisons qu'on parle de résultats cohérents. Le test de 23andme peut aussi donner un résultat faux, si une mutation importante n'a pas été détectée.

Le chromosome X peut être utilisé dans cette comparaison. L'homme reçoit le X de sa mère alors que la femme reçoit des X de ses deux parents. Normalement, deux frères devraient avoir un segment d'ADN X partagé (mais il semble que dans environ 6% des cas, ils puissent avoir reçu des X différents de leur mère).

### 7.7.2 Comparer des cousins

---

Plus la parenté est éloignée et plus le nombre de cM en commun peut varier. L'ADN n'est pas distribué uniformément au fil des générations.

Pour des cousins partageant deux grands-parents, la relation attendue est entre 396 et 1397 cM. Pour une demi-relation (un seul grand-parent), ce serait 156 à 979 cM. Ici aussi, il y a une zone grise alors que les deux relations sont possibles. Dans un tel cas, la méthode est similaire à celle utilisée pour des frères et sœurs avec une différence importante : des matches de plus de 100 cM pourraient disparaître.

Plus la relation s'éloigne et plus la variation de l'ADN commun est importante. Il faut donc faire attention avant de conclure trop hâtivement en étudiant une telle relation. Il est sans doute préférable d'identifier quelques personnes parmi les matches pour voir quelle est la relation attendue. De plus, l'endogamie peut fausser les résultats. Si les parents d'un match ont épousé le frère et la sœur de vos parents, l'ADN calculé pourrait être le même qu'avec votre frère ou votre sœur. La même remarque s'applique quand on recule de plusieurs générations.

### 7.7.3 Confirmer un grand-parent

---

La relation calculée avec un grand-parent est similaire à celle d'un demi-frère (984 à 2462 cM). Comme l'ADN diminue de moitié environ à chaque génération, on devrait ici aussi retrouver les mêmes matches importants et une partie des matches plus éloignés.

## 7.8 Résoudre une énigme généalogique avec l'ADN

---

L'ADN généalogique est un outil pouvant servir à résoudre des énigmes généalogiques.

Les adoptés (« nés sous X » en France) l'utilisent pour retrouver des parents biologiques. En allant vers le passé, après environ cinq générations vers le passé ou dix en tout, les possibilités deviennent trop nombreuses et il est pratiquement impossible, sauf exception, de confirmer ou découvrir des parents avec l'ADN autosomal.

### 7.8.1 Recherche des parents biologiques

---

La méthode est assez simple mais demande beaucoup de travail.

Il faut étudier les correspondances les plus proches pour reconstruire leur arbre généalogique. Ensuite, en recoupant les couples trouvés chez plusieurs cousins, on trouve des aïeuls potentiels. Puis, on fait la recherche inverse en tentant d'identifier leurs descendants. Selon les circonstances, on peut trouver de la parenté de plus en plus proche. Il est même possible, parfois, de convaincre le descendant d'un de ces couples de faire un test ADN pour se rapprocher des parents biologiques.

La façon de construire ces arbres varie beaucoup selon le laboratoire utilisé.

**AncestryDNA** est le préféré chez les Québécois parce qu'il fait aussi le lien avec l'arbre généalogique en ligne de ces derniers, parce que c'est un site de généalogie. Le lien est toutefois fait à partir des arbres généalogiques personnels communs. Il est possible de construire un arbre théorique et de trouver des cousins via le système Thrulines qui nécessite d'avoir un arbre en ligne et qui indique combien de clients ont les mêmes ancêtres (avec des détails similaires, sinon on aurait une situation comme 100 descendants d'un homme et aucun de sa femme). Si un adopté ignore qui sont ses parents, il faut plutôt étudier les arbres des correspondances trouvées et trouver ceux qui se répètent. Il est aussi permis de faire un arbre théorique et de voir après quelques jours quels ancêtres ont des descendants partageant de l'ADN.

**myHeritage** devrait être similaire à Ancestry avec toutefois moins de clients en Amérique du Nord, mais beaucoup plus en France (où l'entreprise a fait de la publicité de 2018 à 2022). Avec la version débloquée ou l'abonnement, on peut former une matrice de triangulation (voir plus bas) permettant de regrouper les cousins partageant entre eux des ancêtres communs.

**23andme** n'a pas d'arbres généalogiques en ligne mais propose un arbre généalogique établi à partir des cousins trouvés. Auparavant, il faut permettre le partage des segments. Dans le profil, il faut cliquer « Yes » dans la section « *Would you like to display your ancestry results and matching DNA segments to DNA Relatives participants?* ». Par la suite, il faut demander la liste avec « View Relatives » qui est d'abord une demande de création de la liste (si la liste a été créée auparavant, il faudra éventuellement attendre la nouvelle version), ce qui donnera à la fois la liste des 1500 plus proches et qui fera un arbre en ligne « Family Tree » avec la position théorique des cousins les plus proches. Comme l'ADN ne donne pas un positionnement exact, on peut éditer l'arbre obtenu et ajouter des générations. Il est possible d'identifier plusieurs cousins avec l'arbre en ligne (par exemple, si une personne a fait le test ADN avec un de ses parents), ou avec les indices laissés par le client (âge, ville de résidence, noms de famille des ancêtres et lieux de naissance, etc.). Les segments font partie de l'information du fichier «Relatives», ce qui permet aussi d'identifier les personnes proches entre elles (dont un segment serait positionné au même endroit). On peut aussi faire un tableau des personnes partageant de l'ADN avec un cousin, ce qui donnera le pourcentage commun.

23andme montre un haplogroupe pour environ la moitié des matches. Il est intéressant de comparer l'haplogroupe obtenu par ce test (avec les versions les plus récentes) et la parenté ayant des haplogroupes similaires. Dans certains cas, si l'autre personne n'a pas utilisé la même version, l'haplogroupe peut être un peu différent. Ce n'est pas parce qu'un client a le même haplogroupe paternel qu'il est le père du sujet, mais si l'haplogroupe n'est pas compatible (R- et G- par exemple), c'est certain que ce n'est pas le père. Cette caractéristique est aussi disponible avec le Family Finder qui suit.

FTDNA offre plusieurs tests ADN dont le test **Family Finder**. C'est ce dernier qui devrait être utilisé pour la recherche de parents biologiques en plus du test ADN Y si c'est un homme, Le test YSTR-37 devrait donner un nom de famille pour environ la moitié des Québécois et Acadiens de vieille souche française. Le test *Family Finder* pourrait aider à compléter la généalogie familiale. Il est aussi possible de trouver de la parenté présumée du même nom en utilisant les autres laboratoires et de les inviter à soit importer leurs données génétiques, soit faire eux aussi un test ADN Y. Comme 23andme, la liste des matches et des segments partagés est aussi disponible. Si le match a aussi fait les tests Y et MT, des haplogroupes pourraient apparaître comme autres informations. Avec l'existence de projets, FTDNA permet aussi de choisir des matches à l'intérieur d'un projet (via *Advanced Matching* pour le *Family Finder*). Un administrateur de projet a parfois des informations supplémentaires comme la ville de résidence ou l'arbre en ligne avec moins de noms cachés comme privés (si la personne n'a pas de données de décès et selon la configuration du client). Par ailleurs, le même échantillon de salive permet de faire un autre test. Donc, avec le test ADN Y d'un cousin, par exemple, on

peut demander plus tard le *Family Finder*. Si un parent (oncle, tante, cousin, etc.) paternel ou maternel est identifié dans l'arbre, le système fera la distinction entre les cousins paternels et maternels.

FTDNA offre aussi la possibilité de comparer les relations entre cousins pour voir qui est cousin avec qui, et identifier des ancêtres plus fréquents en choisissant une liste de 10 cousins. Ce système est toutefois assez rudimentaire.

### 7.8.2 La triangulation

Alors qu'avec l'ADN Y et l'ADNmt, la triangulation consiste à documenter deux lignées à partir de personnes testées présentant des résultats similaires, elle prend une forme très différente avec les tests d'ADN autosomal.

Le concept repose sur une matrice où des cousins sont énumérés sur chaque axe, en indiquant qui est parent avec qui à chaque interception.

Chez FTDNA, le concept est appelé « Matrix » mais ne semble pas très perfectionné. Toutefois, on peut identifier qui est apparenté avec tel cousin grâce à l'outil « *In Common With* ».

Avec myHeritage, l'outil « AutoGroupes » choisit les meilleurs candidats et forme une matrice avec la liste de ces candidats sur chaque axe, et une case colorée pour ceux partageant une parenté suffisante avec d'autres candidats, formant ainsi des nuages ou « *clusters* ». Il faut ensuite explorer les généalogies disponibles pour trouver l'ancêtre commun dans chaque nuage.

Chez Gedmatch, parmi les outils Tier 1, l'outil de triangulation compare les 500 premiers candidats pour déterminer ceux qui ont une relation entre eux.

23andme et Ancestry ne semblent pas avoir de système utilisant une matrice mais il est possible de savoir qui est parent avec qui parmi les cousins trouvés. La liste de ces cousins ne peut toutefois pas être enregistrée.

### 7.8.3 Recherche des grands-parents biologiques

La recherche de grands-parents est similaire à celle de parents avec quelques remarques intéressantes.

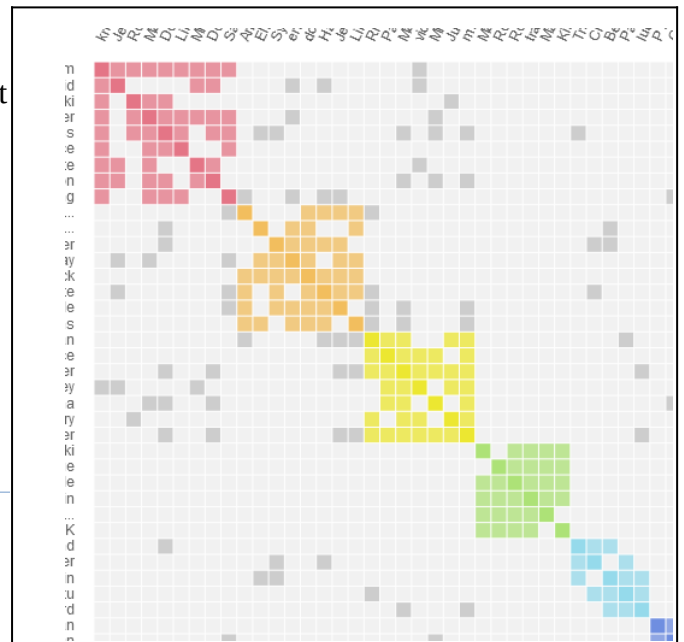


Figure 16: Autogroupes de myHeritage

Si vous connaissez déjà un des parents (par exemple, mère connue et père inconnu), plusieurs laboratoires permettent d'isoler les parentés paternelle et maternelle quand on a identifié un cousin assez proche. Par exemple, on pourrait tester la mère ou un cousin maternel, et les autres cousins proches devraient être du côté paternel.

Plus on remonte loin et plus la recherche est difficile à cause de l'augmentation du nombre d'ancêtres communs avec vos matches. Il y a une exception, c'est d'une lignée isolée des autres. Par exemple, un Américain ont l'ascendance a une seule lignée québécoise sur 6 générations pourrait débloquer cette lignée avec quelques indices comme une année approximative de naissance et en recherchant parmi la parenté trouvée les membres de la famille dont la trace est perdue.

### 7.8.4 Contacter un cousin biologique

---

L'adoption peut provenir d'une fille-mère ou d'un enfant adultérin ou d'un autre secret de famille. Il faut alors faire usage de beaucoup de diplomatie pour entrer en contact avec des cousins biologiques. Par exemple, il ne faut pas dire immédiatement que l'on recherche ses parents biologiques. Dans plusieurs cas, les données ont disparu d'un arbre public après l'envoi d'un tel message.

Il faut d'abord recopier l'information disponible avec une capture d'écran ou la saisie dans un logiciel de généalogie. Ensuite, il est préférable de contacter d'abord la parenté plus éloignée, parce que le facteur émotionnel est moins fort quand on s'éloigne des personnes concernées. Dans plusieurs cas, le premier contact a été positif parce qu'on parlait seulement de généalogie.

### 7.9 L'ADN longue durée des zones d'empilement

---

Même si l'ADN est combiné à chaque génération, une séquence peut subsister durant plusieurs générations, ce qui vient fausser les résultats et annoncer une relation à 5 générations alors que cette séquence d'ADN dure depuis 10 générations ou plus. Il y a des sections d'ADN qui sont plus propices à cette survie à long terme, ce qui nous donne des cousins proches alors qu'ils ne le sont pas. Quand on découvre ce phénomène pour la première fois, on recherche parmi nos ancêtres celui qui a laissé son ADN en dehors du mariage, alors qu'en fait, il y a bien un ancêtre commun très éloigné.

Selon un blogue sur le sujet<sup>51</sup>, ces segments seraient personnels (les miens seraient différents de ceux de mon voisin, par exemple) mais il s'agit peut-être des valeurs lues et non des positions. Selon un article scientifique, les zones populaires seraient différentes en Europe et en Asie. Pour référence, voici les empilements trouvés<sup>52</sup> selon la norme

---

<sup>51</sup> <https://cruwys.blogspot.com/2018/01/small-segments-and-pile-ups.html>

<sup>52</sup> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3907355/>

## L'ADN autosomal

Hg19 alors que Hg38 est plus récent. La norme du génome humain Hg38 étant plus récente, les positions sont plus grandes.

Chromosome	Début (Hg19) <sup>53</sup>	Fin	Total (Hg38) <sup>54</sup>	Longueur (cM)	Family Finder	23andme
1	118 434 520	153 401 108	248 387 328	9,95	Omis	Omis
2	85 304 243	99 558 013	242 696 752	6,53	Omis	Omis
2	132 695 025	141 442 636	242 696 752	9,16	Omis	Trouvé
2	192 352 906	198 110 229	242 696 752	5,04	Omis	Trouvé
8	10 428 647	13 469 693	146 259 331	7,96	Trouvé	Trouvé
9	38 293 483	72 605 261	150 617 247	8,15	Omis	Omis
10	44 555 093	53 240 188	134 758 134	7,58	Omis	Trouvé
15	20 060 673	25 145 260	99 753 195	10,46	Omis	Trouvé
15	27 115 823	30 295 750	99 753 195	9,29	Omis	Trouvé
16	19 393 068	24 031 556	96 330 374	6,18	Trouvé	Trouvé
17	59 518 083	64 970 531	84 276 897	6,23	Omis	Trouvé
17	77 186 666	78 417 478	84 276 897	5,66	Omis	Trouvé
21	16 344 186	19 375 168	45 090 682	6,91	Omis	Trouvé
22	16 051 881	25 095 451	51 324 926	20,82	Trouvé	Trouvé

En examinant les matches dans les tests de l'auteur, il est possible de savoir si des zones d'empilement ont été trouvées ou omises dans les résultats. Ainsi, la zone dans le chromosome 1 est omise en se basant sur les résultats de ces tests, alors que celle du numéro 10 est omise dans le Family Finder mais trouvée chez 23andme. Certaines zones sont parfois omises parce que difficiles à détecter et il est possible qu'un laboratoire détecte une zone d'empilement qui n'était pas dans l'étude initiale.

Que doit-on faire avec cette information? Si vous tentez d'identifier les personnes dont le seul segment se trouve dans cette zone, il y a une forte probabilité que vous ne pourrez pas les placer dans votre arbre généalogique parce que l'ADN commun a voyagé depuis un grand nombre de générations. Si vous voulez un défi personnel avec la génétique, essayez de trouver l'ascendance complète de toutes vos correspondances dans un de ces segments, puis de trouver un couple unique dont l'ADN serait à l'origine de ce segment.

<sup>53</sup> La liste des zones d'empilement trouvées se base sur la version Hg19 du génome humain (article de référence) alors que la longueur des chromosomes vient de la version Hg38 ainsi que les résultats de l'auteur.

<sup>54</sup> Selon wikipedia.

Si vous avez des racines profondes en Amérique du Nord, il est probable que vos ancêtres immigrants viennent d'un grand nombre de régions et qu'avec beaucoup de travail, vous puissiez identifier les ancêtres à l'origine d'une zone d'empilement. Par contre, si vos ancêtres dans les trois ou quatre dernières générations sont dans l'Ancien Monde, cette identification devrait être impossible.

### 7.10 Les origines ethniques

---

Plusieurs font un test ADN dans le but de découvrir l'origine de leurs ancêtres sans avoir recours à l'ADN. La publicité de certains laboratoires vante ces résultats qui deviennent la raison d'acheter de nombreux clients et aussi la déception des adoptés qui pensaient trouver leur parenté biologique à partir de clients qui n'ont pas d'intérêt pour la généalogie.

L'ADN n'est pas distribué de façon uniforme. Si votre ADN vient à 50% de chacun de vos parents, ce n'est pas à 25% de chaque grand-parent. De plus, les parties héritées de chaque parent varient selon l'enfant. Il s'en suit que peu importe la technologie, l'ADN reçu de 64 ancêtres n'est pas le 1/64 de l'ADN de chacun. Il est même possible de n'hériter d'aucun morceau d'ADN de certains de ces 64 ancêtres.

Par ailleurs, il y a plusieurs façons d'analyser l'ADN pour en découvrir des origines régionales. Ces méthodes reposent sur des statistiques et non sur de l'ADN qui serait uniquement présent dans certaines populations, un peu comme les groupes sanguins. Ainsi, telle combinaison est plus fréquente chez les Européens du centre et moins chez les Asiatiques. Elle n'est pas exclusive à l'une ou l'autre région. La comparaison se fait avec des populations de référence qui sont maintenant la clientèle de chaque gros laboratoire alors que dans la documentation plus ancienne, l'on faisait un échantillonnage diversifié de différentes populations. À cause des coûts des tests par rapport au salaire moyen de la population locale ainsi que de lois qui limitent la livraison de trousse de prélèvement, certaines régions sont plus testées que d'autres. Les Américains ayant sans doute acheté beaucoup plus de tests ADN que les habitants d'autres pays, mais partageant une bonne variété d'origines surtout britanniques, ils ont aussi influencé la catégorisation des populations de référence. La France, au contraire, est sous-représentée et il s'en suit qu'un Français (ou un descendant ayant surtout des ancêtres en France) aura souvent, selon ces tests, des origines britanniques, espagnoles, allemandes ou italiennes, les voisins de la France.

Il ne faut donc pas s'attendre à des résultats précis et si cela arrive, ce n'est pas parce que le laboratoire donne des meilleurs résultats mais par hasard, et quand le même laboratoire fera une mise à jour de vos résultats, ceux-ci changeront.

À cela s'ajoute un 2% d'ADN inconnu, parce qu'un test autosomal ne trouve pas toutes les valeurs recherchées. Pour cette raison, si des jumeaux identiques ou si la même personne faisant en même temps deux fois le même test ADN de la même version chez le même laboratoire, le résultat montrera deux tableaux d'origines en contradiction l'un avec l'autre. Ceci est encore plus évident si l'on compare des frères et sœurs ou encore les résultats de deux laboratoires différents.

Et pour rendre les résultats encore moins exacts, la géographie a changé lors des grandes guerres et d'invasions, et la diversité des populations aujourd'hui n'est plus celle d'il y a 100 ou 200 ans. Les bases de données génétiques rattachent l'origine de leur clientèle aux déclarations de chaque client. Si ma grand-mère est née dans un autre pays mais qu'elle s'est mariée et a vécu au même endroit que ses parents, son origine basée sur son lieu de naissance n'est pas valide.

Il faut donc retenir que les origines ethniques selon des tests génétiques ne sont pas précises, qu'elles ne sont pas des « origines » puisque basées sur la clientèle moderne, et qu'elles ne sont pas « ethnique » mais géographique, également en se basant sur la géographie moderne.

En ce qui concerne les origines amérindiennes, pour la plupart des gens, c'est une ancêtre ayant vécu il y a huit générations ou plus, ce qui donnerait 0,4% de l'ADN à condition que l'ADN soit distribué uniformément. Il faut aussi tenir compte des inventions de certains qui transforment un ancêtre d'origine inconnue en autochtone ou qui changent complètement la signification d'un test ADN faute d'en comprendre toutes les nuances, ainsi que d'adoptions possibles qui peuvent modifier les résultats de l'élément ethnique.

### 7.11 Le génome complet pour un usage autosomal

En théorie, l'analyse du génome complet devrait permettre d'en extraire des données compatibles avec les tests des principaux laboratoires (AncestryDNA, 23andme, myHeritage, FTDNA) et donc leur importation vers les sites permettant la comparaison des données de plusieurs laboratoires (myHeritage, FTDNA, Gedmatch, Geneanet). En pratique, ce genre de service n'est pas encore annoncé de façon officielle en 2023. Nebula l'avait annoncé mais ce projet a été abandonné.

Plusieurs obstacles limitent la génération d'un fichier de données brutes à des fins généalogiques. Le principal est peut-être un manque d'intérêt de la clientèle qui va plutôt se tourner vers les tests déjà disponibles et associés à des bases de données dépassant le million de clients.

### 7.11.1 Peut-on aller plus loin?

---

Par ailleurs, serait-il possible de détecter un cousinage plus éloigné en utilisant un génome complet?

Un des obstacles vient de l'endogamie. Si on remonte cinq générations, par exemple, et qu'on recherche les ancêtres communs entre deux personnes, il est bien possible qu'on trouve un seul couple. Mais dans un tel cas, on s'attend à en trouver un grand nombre après six ou sept générations. En bref, quand la parenté est trop éloignée, parce que nos ancêtres ne vivaient pas chacun sur une île loin des autres, il y a souvent plusieurs ancêtres communs possibles. Donc, même en utilisant des séquences plus longues (on pourrait mesurer 2 millions de positions avec le test le plus courant en utilisant 3 puces programmées de façon complémentaire, au lieu de 650 000), on aurait le problème d'isoler les ancêtres communs possibles, mais on pourrait peut-être ajouter une ou deux générations. Le reste du génome humain, avec 3 milliards de bases, est assez stable et c'est cette stabilité qui est utilisée pour identifier les 650 000 valeurs lues dans un test ADN autosomal.

Une façon de contourner ce problème est d'utiliser l'ADN Y et MT qui sont encore identifiables après 20 ou 30 générations, mais cela suit un parcours uniparental, de père en fils ou de mère en fille, et sans qu'on puisse identifier précisément l'ancêtre commun le plus proche sans avoir de documentation. Par exemple, bien qu'ils aient des résultats Y compatibles, on ne peut pas savoir exactement combien de générations séparent les pionniers Gagnon ou Boucher venus au Québec, bien qu'on puisse avoir une idée approximative.

Pour ce qui est de l'ADN ancien, les études montrent plutôt les relations entre des personnes ayant vécu au même endroit durant un certain nombre d'années que leurs relations avec des personnes vivantes. On ne pourrait pas déduire, par exemple, que telle personne vivant à proximité d'un site archéologique descend d'une personne dont les ossements sont dans ce même site. Même si la séquence commune était longue, on ne pourrait pas savoir si elle existait longtemps avant l'ancêtre trouvé.

### 7.12 L'information médicale

---

23andme est le seul laboratoire populaire à offrir ouvertement les prédispositions à certains problèmes d'origine génétique. Selon le tableau de l'ISOGG, la version courante de son test donne 630 000 valeurs. Selon un expert du domaine, son test ADN contiendrait 250 000 valeurs servant seulement à trouver des cousins. Il y aurait donc 380 000 valeurs servant à la fois à trouver des cousins et à découvrir ces prédispositions, sur les 3 milliards de valeurs du génome humain complet.

## L'ADN autosomal

23andme offre deux tests. Le test de base ne livre que les résultats généalogiques et les origines, en plus des 1500 cousins les plus proches. En payant un supplément, la partie médicale est aussi analysée et avec un abonnement, on pourrait en plus recevoir les données sur environ 3000 cousins supplémentaires. Certaines personnes présument que les données médicales sont comprises dans le fichier de données brutes et proposent parfois d'autres sites pour faire l'analyse du contenu. L'auteur ne s'intéressant pas à ce sujet, il n'a pas été étudié.

Les tests WGS donnant un génome complet (la couverture réelle varie selon le laboratoire) peuvent aussi offrir une analyse médicale plus raffinée. Ainsi, **Nebula** offre des tests WGS pour 99\$, 249\$ et 899\$, et exige une adhésion (*membership*) de 200\$ ou 275\$ à leur groupe afin de recevoir des rapports médicaux à mesure que la recherche avance. **Dante** détecte le risque de plus de 50 maladies et suggère des habitudes de vie ou une diète en fonction du génome.

## 8 Le laboratoire FTDNA

---

Le laboratoire FamilyTreeDNA ou FTDNA est le premier à vendre des tests ADN individuels à partir de l'an 2000. Ce laboratoire offre une grande variété de tests Y et MT, certains maintenant retirés mais qui peuvent apparaître parmi les résultats. Il a aussi, en 2023, la plus importante base de données génétiques sur l'ADN Y et MT.

On y trouve également des projets, qui sont des regroupements de résultats afin d'en faciliter l'étude et l'affichage de résultats. Le plus gros projet concernant la francophonie est sans doute *French Heritage* ou Héritage français, lancé en 2005 par *The French-Canadian Heritage Society of California and the Southern California Genealogical Society*, est géré à partir de 2008 environ par le Québécois Jacques Beaugrand, puis depuis 2018 par Denis Beaugard et Denis Savard.

### 8.1 Introduction

---

FTDNA a offert au fil des années plusieurs sortes de tests ADN Y, MT et autosomal. Ces tests peuvent souvent être effectués à partir d'un seul échantillon de salive prélevé en grattant l'intérieur des joues. Ainsi, le client peut effectuer le test Y-37, puis faire une mise à niveau vers le test Y-111 et le Big Y. Entre-temps, il a pu ajouter un test ADNmt et un test autosomal.

### 8.2 Les tests disponibles

---

#### 8.2.1 Tests ADN Y

---

La série de tests offerts pour l'ADN Y, réservés aux hommes, est la suivante :

Année	2000	2001	2003	2006	2011	2013	2018	2019
Test	Y-12	Y-25	Y-37	Y-67	Y-111	Big Y	Big Y-500	Big Y-700
Distance	0 ou 1	2	4	7	10			
Marqueurs similaires	11	23	33	60	101			
Disponible?	Non	MAN	Oui	MAN	Oui	Non	Non	Oui
Max haplotypes	203	18	8	7	7			
2 <sup>e</sup> rang	97	10	7	6	3			

Le niveau de précision du test dépend du nombre de marqueurs. Plus le nombre de marqueurs semblables est élevé et plus l'ADN comparé est proche. Ainsi, un test Y-12

exige d'avoir 11 valeurs similaires à l'intérieur d'un même projet ou 12 pour l'ensemble des résultats. Un test Y-111 exige 101 valeurs similaires (ou permet 10 différences). Ce concept n'est plus le même avec les tests Big Y.

Plusieurs tests ne sont plus offerts mais leurs résultats apparaîtront comme matches lors d'une vérification. Ainsi, le test Y-12, longtemps offert comme test le moins cher, puis en participant à un projet, n'apparaît plus. Les tests Y-25 et Y-67 sont disponibles comme mises à niveau (MAN) de tests moins précis. Les tests Big Y fonctionnent de façon différente en examinant les SNP plutôt que des marqueurs STR. La première génération n'offrait que les SNP connus en soi, sans analyse poussée. Le laboratoire russe YFULL ayant commencé à extraire directement ses 400 marqueurs STR à partir des résultats bruts, FTDNA a répliqué avec sa propre extraction de 500, puis 700 marqueurs STR. Quand le Big Y 700 a été offert, les résultats de Big Y 500 ont été ajustés pour les marqueurs ajoutés.

### 8.2.2 Tests ADNmt

La série de tests offerts pour l'ADNmt, pour tous les clients mais suivant la lignée maternelle, est la suivante :

Test	MT	MT Plus	FMS
Régions analysées	HVR1	HVR1+HVR2	HVR1+HVR2+CR
SNP analysés	558	558+576 (1134)	16 569
Disponible?	Non	Non	Oui

Le moteur interne de recherches offre d'abord les résultats où l'haplogroupe est identique et jusqu'à 3 différences, puis ceux où les parties HVR1 et HVR2 sont identiques, et finalement seulement HVR1.

Si vous avez peu de matches avec c'est probablement parce que votre branche a été peu testée, ou en d'autres mots, que peu de clients ont des résultats similaires aux vôtres. Si un FMS donne une seule lettre comme haplogroupe (ou certaines combinaisons comme H1 ou U5), c'est également parce qu'à l'époque où le regroupement s'est fait, il y avait peu d'échantillons. À l'avenir, il est possible que par exemple V soit redéfini comme V123.

Le résultat offert par défaut suit le protocole RSRS (comparaison avec l'Ève africaine). rCRS est aussi disponible (Ève européenne). Cela comprend la liste des différences par rapport à la norme. De plus, dans plusieurs cas, il y a un affichage séparé des mutations. Une mutation MT a la forme A123G qui signifie que la position 123 vaut normalement A mais qu'elle vaut G dans vos résultats. Les lettres autres que A, C, G ou

T sont de l'hétéroplasmie, quand au moins 20% des résultats contiennent une de ces valeurs.

### 8.2.3 Test Family Finder

---

FTDNA offre un seul test autosomal appelé Family Finder mais a changé les positions échantillonnées. En théorie, cela ne devrait pas influencer les matches trouvés s'ils sont importants. De même, la méthode pour accepter un match a été modifiée avec la longueur requise en cM si le total est en bas de 20 cM. Un test fait il y a plusieurs années aura donc un certain nombre de petites correspondances qui ne sont plus affichées aujourd'hui.

À la suite de l'acquisition de FTDNA par l'entreprise australienne myDNA, d'autres tests autosomaux ont été ajoutés, le plus souvent avec une portée médicale. Ces autres tests ne devraient pas influencer la liste des correspondances trouvées.

À la fin de 2023, FTDNA a débuté un programme d'extraction d'ADN Y à partir de tests Family Finder. Comme il y a eu plusieurs versions, cette opération commence avec la version la plus récente. Certaines troussees auront donc un haplogroupe Y sans qu'il y ait de test ADN Y. Ces résultats sont affichés seulement sur le profil du client.

## 8.3 Visite guidée

---

Les écrans d'information de FTDNA ont été refait vers 2020 et la présentation a changé en conséquence. Les écrans suivants sont ceux disponibles en 2023. Le site est seulement en anglais, Si vous utilisez un navigateur qui traduit le contenu, il faudra faire attention avec les noms de personnes ou de projets qui pourraient aussi être traduits.

À des fins d'illustration, les écrans de résultats sont tirés du projet *French Heritage* (Héritage français).

Le client de FTDNA notera qu'il est inutile de donner son mot de passe à un administrateur de projets et que l'exiger va à l'encontre des règles que doivent suivre les administrateurs. Celui-ci peut avoir accès à tous les résultats si vous rejoignez son projet et que vous lui donnez l'accès requis (limité ou avancé). Si vos données ne sont plus dans un projet que vous avez rejoint, c'est parfois parce qu'un administrateur a retiré vos données de ce projet. Si cela est arrivé, il est préférable de changer son mot de passe et de choisir soi-même les projets intéressants.

### 8.3.1 Panneau d'entrée (*Dashboard*)

---

Après avoir commandé un test ou importé le résultat d'un autre laboratoire, un compte est créé. Il y a un seul compte par trousse. Si vous testez deux personnes, chacun aura un compte différent mais il est permis d'utiliser le même courriel et le même mot de

## Le laboratoire FTDNA

passer pour les deux. La première étape est la saisie du numéro de trousse et du mot de passe depuis la page d'entrée au site :

<https://www.familytreedna.com/sign-in>

Le premier écran affiché est le panneau principal.

Une série de menus dans le haut de la page donne accès aux données principales, tout comme les sections dans le reste de la page. L'ordre de présentation peut changer selon les tests effectués.

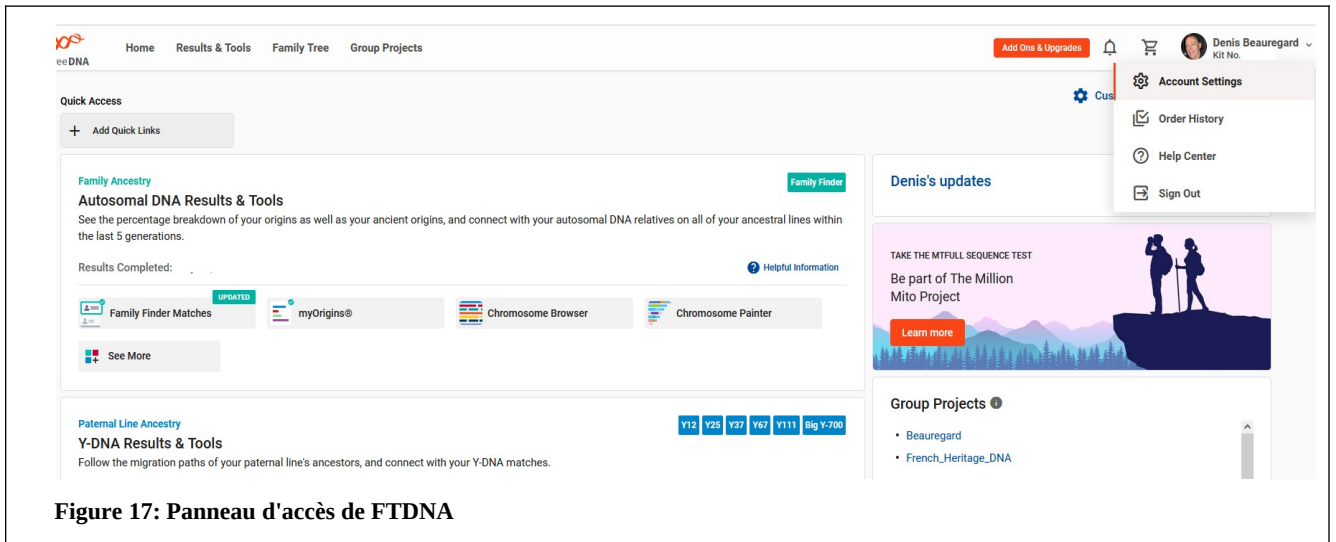


Figure 17: Panneau d'accès de FTDNA

Plus loin à droite, un lien permet d'acheter des mises à niveau et d'autres tests. Une cloche montre les notifications venant des différents projets reliés au compte. Un panier montre les achats en cours, le cas échéant. La photo optionnelle du client et son nom permettent de voir les paramètres du compte [Account Settings], un historique des commandes depuis le premier achat, un centre d'aide et un lien pour fermer l'accès (qui arrête après deux heures d'inactivité).

Les sections [Y-DNA Results & Tools], [Big Y], [mtDNA Results & Tools] et [Autosomal Results & Tools] donnent accès aux différents types de résultats que nous verrons plus bas, en gros les matches, le chemin théorique de migration de la lignée (Migration Maps), d'une carte des ancêtres les plus anciens des matches (Matches Maps), le pays d'origine des matches (Ancestral Origins)

D'autres outils sont disponibles au bas, comme votre arbre généalogique personnel si vous l'avez ajouté à votre profil [Family Tree], un arbre phylogénétique public [Public

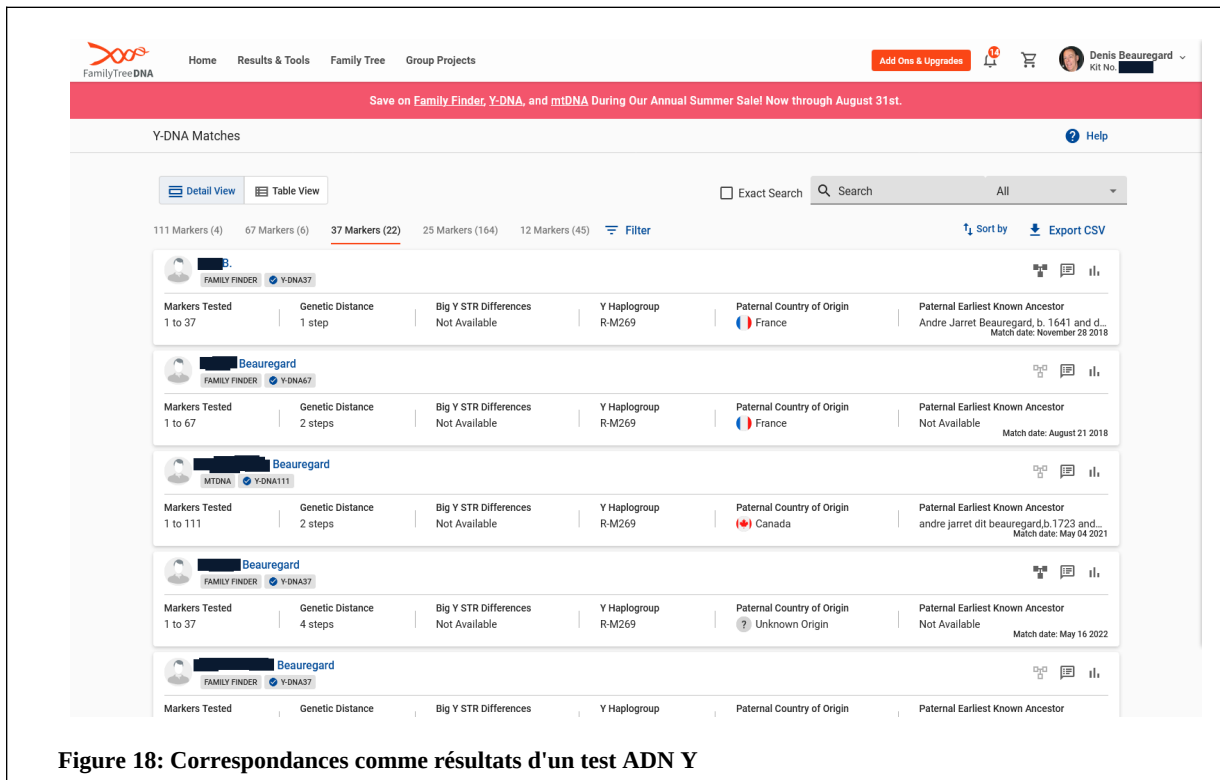
*Haplotrees*], des rapports sur la fonction *Discover* de l'ADN-Y, un accès avancé à des résultats en croisant les tests, et à d'autres outils secondaires.

La colonne droite montre des liens vers un sondage (*Survey*) sur vos origines, la liste de vos projets et des symboles (*Badge*) de vos haplogroupes selon vos tests.

### 8.3.2 La section [*Y-DNA Results & Tools*]

Cette section montre d'abord une liste des correspondances, la généalogie ou arbre phylogénétique des haplogroupes Y (*Haplotree & SNPs*) aussi disponible avec le lien « *Badge* » de la colonne à droite, des rapports *Discover* selon votre résultat, et d'autres informations sous le lien *See More*.

Les correspondances ont la forme suivante :



The screenshot displays the 'Y-DNA Matches' interface on the FamilyTreeDNA website. At the top, there is a navigation bar with 'Home', 'Results & Tools', 'Family Tree', and 'Group Projects'. A red banner promotes a 'Save on Family Finder, Y-DNA, and mtDNA' sale. Below the banner, the page title is 'Y-DNA Matches' with a 'Help' link. The main content area features a table of matches. The table has columns for 'Markers Tested', 'Genetic Distance', 'Big Y STR Differences', 'Y Haplogroup', 'Paternal Country of Origin', and 'Paternal Earliest Known Ancestor'. The first match shows 1 to 37 markers tested, a genetic distance of 1 step, and a paternal country of origin of France. The second match shows 1 to 67 markers tested, a genetic distance of 2 steps, and a paternal country of origin of France. The third match shows 1 to 111 markers tested, a genetic distance of 2 steps, and a paternal country of origin of Canada. The fourth match shows 1 to 37 markers tested, a genetic distance of 4 steps, and a paternal country of origin of Unknown Origin. The fifth match shows 1 to 37 markers tested, a genetic distance of 4 steps, and a paternal country of origin of Unknown Origin. The page also includes a search bar, a filter button, and a 'Sort by' dropdown menu.

**Figure 18: Correspondances comme résultats d'un test ADN Y**

Par défaut, les détails sont affichés, avec en haut la sélection du nombre de marqueurs et le nombre de correspondances pour chacun, des filtres dont la sélection d'un projet commun (appelé *Group*), et d'autres critères. Le lien « *Sort* » permet de trier les colonnes, sauf par le nom du client (il faut plutôt choisir la vue par table et cliquer sur le titre « *Name* »).

Puis, pour chaque match, une image éventuelle du client, son client avec un lien pour voir son courriel et des détails additionnels, les tests ADN (l'haplogroupe MT est affiché en cliquant sur le nom du client), le nombre de marqueurs, le nombre de différences ou distance génétique, le nombre de STR différents si vous avez fait le test Big Y tous les deux, l'haplogroupe du client, le pays d'origine de son ancêtre (\*), l'ancêtre le plus éloigné (\*) et la date où le match a été trouvé par FTDNA. Un arbre symbolique indique si le client a fait un arbre sur FTDNA (\*) (aucun arbre si les carrés sont blancs et arbre privé s'il est biffé). Vous pouvez ajouter une note (qui sera visible par les administrateurs de vos projets). Un dernier lien permet d'estimer le nombre de générations vers l'ancêtre commun mais cette évaluation est jugée peu fiable selon la plupart des experts.

(\*) certaines informations doivent être ajoutées par le client

L'haplogroupe mentionné ne précise pas s'il s'agit de l'évaluation à partir des STR ou d'un test précis (sauf si c'est un *Big Y*). Les STR donnent des haplogroupes comme les suivants, mais un client peut avoir testé le même SNP.

C-M216, E-M2, E-M35, E-M96, G-M201, I-M170, I-M223, I-M253, I-P37, J-M172, J-M267, K-M9, L-M20, N-M231, O-M175, Q-M242, R-M173, R-M198, R-M207, R-M269, R-M512, T-M70
---

Un match avec zéro différence ne signifie pas qu'il s'agit d'une personne proche. D'un côté, un père et un fils peuvent avoir deux différences. De l'autre côté, l'ancêtre commun peut avoir vécu il y a 300 ans et donner zéro différence. Il faut utiliser le test autosomal pour valider une relation père-fils. Et si le test autosomal montre la bonne valeur (autour de 3560), il faut encore comparer l'ADN X (qui doit être assez différent) pour s'assurer que ce n'est pas le même test qui a été importé depuis un autre laboratoire il y a longtemps.

Les lignées paternelles peuvent être comparées si l'arbre de l'autre client est disponible ou s'il a laissé assez d'indices (si on clique sur son nom). Si c'est le cas, la signature est dite « triangulée » jusqu'au premier ancêtre présent dans les deux lignées, l'ACPR (ancêtre commun le plus rapproché, ou *MRCA* pour *Most Recent Common Ancestor*). Si les lignées sont documentées, la signature peut être ajoutée au catalogue collectif de signatures ADN<sup>55</sup>.

Dans certains cas, un match peut suggérer un nom de famille, ce qui permet parfois de débloquent une lignée paternelle.

Si vous avez importé un test « *Geno 2* » de *National Geographic* ou un test ADN Y d'un autre laboratoire, les fonctions disponibles sont limitées au minimum. Le test « *Y*

<sup>55</sup> <http://www.francogene.com/triangulation/>

« Geno 1 » est en fait un test « Y-12 » et l’affichage est similaire à celui d’un test fait directement avec FTDNA, mais sans la possibilité de mise à niveau.

### 8.3.3 La section [Big Y]

Cette section a une structure différente des autres, le test Big Y étant une étape de plus dans l’étude du chromosome Y.

Le « *Block Tree* » est un diagramme en blocs de vos résultats et de ceux des autres résultats *Big Y* ayant un chromosome Y voisin du vôtre. Les rapports « *Discover* » donnent une vue d’ensemble à vos résultats. La section « *Big Y Matches* » affiche vos cousins les plus près. Finalement, les résultats bruts sont donnés sous forme de SNPs (présents ou non dans vos résultats, mais aussi s’ils sont affichés dans les résultats d’ensemble (donc, significatifs), avec la valeur du chromosome Y de référence et la valeur lue dans votre cas.

Le « *Block Tree* » montre les résultats sous forme de blocs. Voici un exemple :

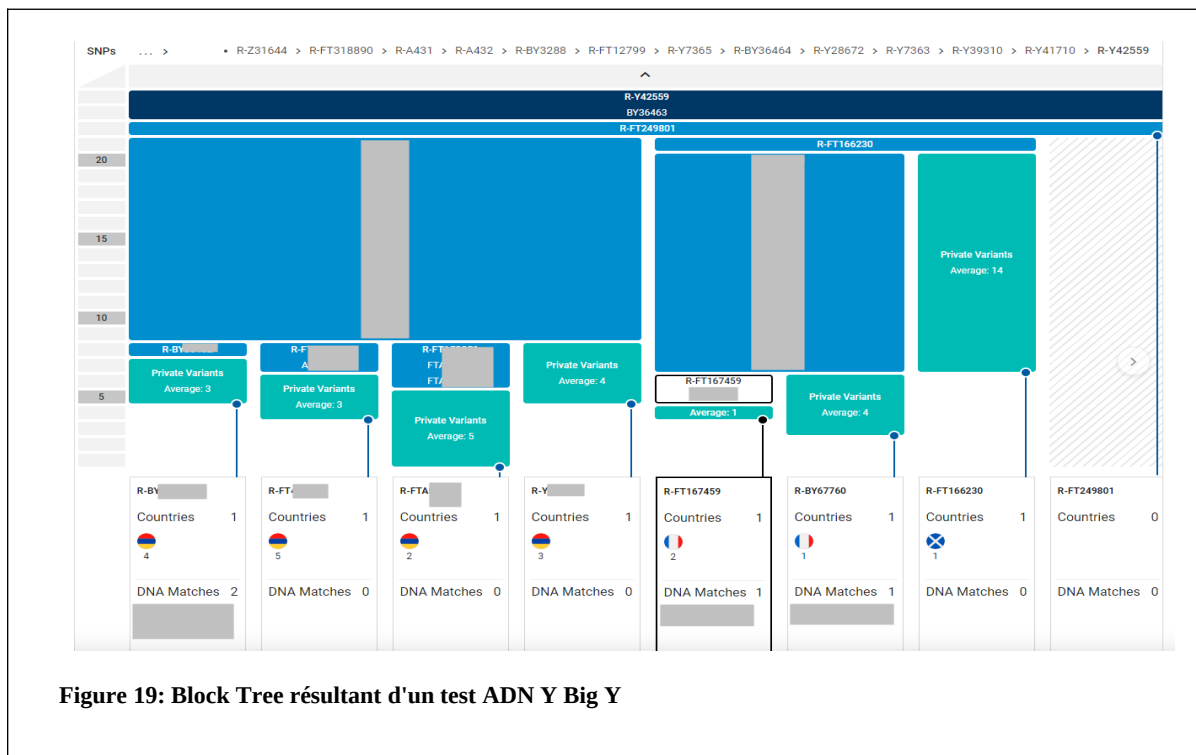
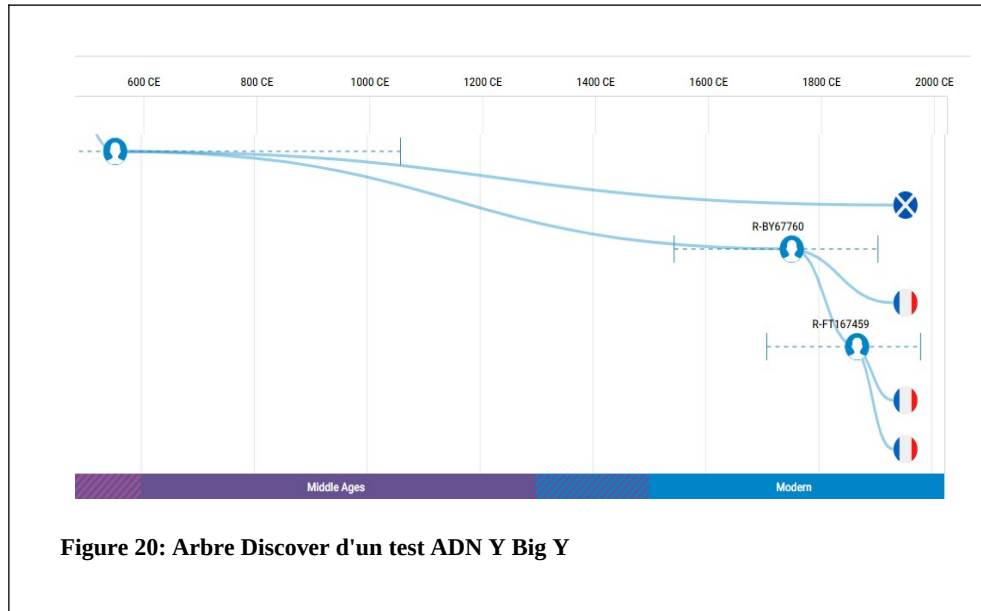


Figure 19: Block Tree résultant d’un test ADN Y Big Y

Chaque bloc bleu ou blanc est une suite de SNP trouvés parmi les descendants de ce bloc. Un bloc blanc ne concerne que vos résultats et contient un premier groupe de SNPs partagés avec les autres clients identifiés ou anonymes au-dessous du bloc. Le premier

SNP identifie le bloc. Les blocs bleus couvrent tous les tests situés au-dessous. Les blocs verts, appelés « *Private Variants* » contiennent des SNPs qui ne sont, pour le moment, pas partagés avec les cousins les plus proches. Ils peuvent avoir un nom s'ils apparaissent ailleurs, chez d'autres clients qui auraient, par hasard, un même SNP mais dans un contexte différent. Des drapeaux identifient souvent le pays d'origine de l'ancêtre le plus ancien.

La ligne au-dessus des blocs montre la séquence des SNPs précédents. Ces SNPs sont aussi affichés parmi les résultats « *Discover* ».



*Discover* montre la position de vos SNP sous une forme différente et sans les listes de SNPs du *Block Tree*. Ce sont donc, essentiellement, des nœuds avec, à gauche les nœuds ou SNPs plus anciens, et à droite, les plus récents, avec au-dessus des années approximatives de formation et un écart calculé d'après le nombre de SNPs trouvés. Parmi les autres résultats, l'on trouve :

- L'histoire de votre haplogroupe (« *Haplogroup Story* », affichée par défaut), soit ceux qui vous précèdent et le nombre de tests qui suivent.
- Le compte par pays (« *Country Frequency* ») permet de situer votre haplogroupe (ou les haplogroupes plus anciens) sur une carte géographique.
- Les connexions célèbres (« *Notable Connections* ») montrent des célébrités dont on connaît l'haplogroupe (d'après leur généalogie et non à partir de tests), en indiquant un haplogroupe commun et son année moyenne de formation.

- La carte de migration (« *Migration Map* ») montre le chemin suivi par votre lignée paternel depuis son début en Afrique.
- Les anciennes connexions (« *Ancient Connections* ») se font avec des sites archéologiques.
- L'arbre temporel (« *Time Tree* », illustré plus haut) place vos résultats parmi l'ensemble des clients partageant des haplogroupes similaires.
- Le chemin ancestral (« *Ancestral Path* ») donne la séquence des SNPs principaux et leur date de formation estimée.
- Les projets suggérés (« *Suggested Projects* ») sont ceux où l'on trouve normalement des clients ayant l'haplogroupe affiché. La priorité est donnée aux projets par haplogroupes.
- Les détails scientifiques (« *Scientific Details* ») montrent une courbe de probabilité de l'année de naissance de celui qui a eu la mutation indiquée par le SNP.
- La page « À propos » (« *About* ») détaille l'origine des informations utilisées.

Il existe aussi un système pour afficher cet arbre à l'intérieur d'un projet, ce qui montre aussi le nom de l'ancêtre fourni par le client.

#### 8.3.4 La section [*mtDNA Results & Tools*]

---

Dans la page d'accueil, plusieurs liens sont disponibles comme indiqué plus haut. Pour l'ADNmt, le lien « *See More* » ajoute les origines de l'haplogroupe, une vidéo faite à partir des résultats, l'origine des autres clients et la liste des mutations. Cette liste est une comparaison avec les normes RSRS (Ève africaine) et rCRS (Ève européenne), avec les 3 sections HVR1, HVR2 et CR, selon le test effectué. Pour un administrateur, la section CR n'est affichée que si le client le permet. L'écran affiche aussi les mutations différentes par rapport à l'haplogroupe trouvé. Ainsi, 309.1C signifie qu'à la position 309, une valeur « C » a été ajoutée par rapport à l'haplogroupe. Le fichier de données brutes « FASTA » est aussi disponible depuis cet écran. Le fichier FASTA montre la liste des 16569 valeurs sous la forme d'une chaîne avec les lettres A, C, G et T.

La liste des matches utilise une présentation différente de l'ADN Y. Une section « *Filter Matches* » permet de choisir toutes les données de FTDNA ou un projet en particulier, puis le niveau de précision du test<sup>56</sup>, le nombre de résultats par écran, le choix d'un nom de famille du client et une date si le résultat est très fréquent.

---

<sup>56</sup> Test complet ou FMS avec HVR1, HVR2, Coding Region; ancien MT Plus avec HVR1, HVR2; ancien MT avec HVR1

Chaque correspondance est ensuite affichée. Le critère change selon le test. Pour le test complet (FMS), on accepte jusqu'à 3 différences (ou distance génétique), alors que pour HVR1 et HVR2, le résultat doit être identique. On trouve, de gauche à droite, la distance génétique, le nom du client (et certains détails si on clique sur ce nom), des icônes pour envoyer un courriel, prendre une note, voir un arbre en ligne et le niveau de test de l'autre client), puis l'ancêtre maternelle la plus ancienne, l'haplogroupe MT et la date de la découverte de la correspondance dans la base de données de FTDNA.

Les correspondances ont la forme suivante :

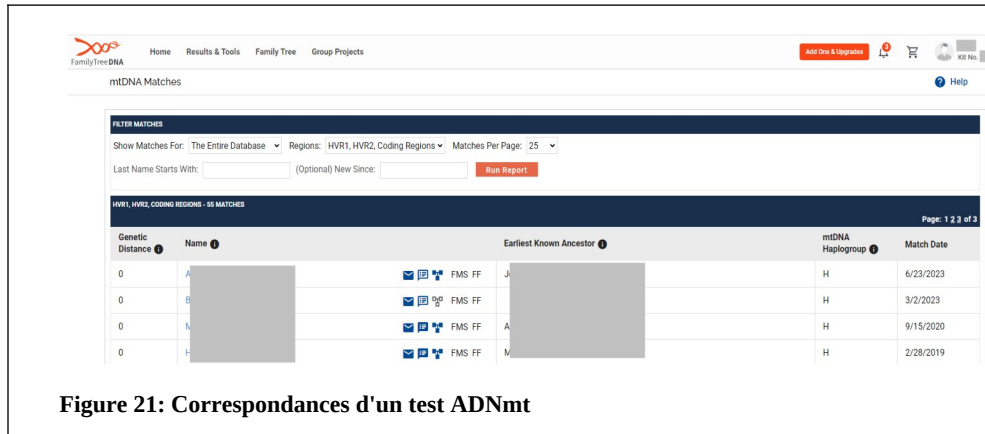


Figure 21: Correspondances d'un test ADNmt

Si vous avez importé un test « *Geno 2* » de *National Geographic* d'un autre laboratoire, les fonctions disponibles sont limitées au minimum car la partie ADNmt est basée sur environ 1500 SNP répartis à travers le génome MT. Le test « *MT Geno 1* » est en fait un test HVR1 et l'affichage est similaire à celui d'un test fait directement avec FTDNA, mais sans la possibilité de mise à niveau.

### 8.3.5 La section [Autosomal DNA Results & Tools]

Le test ADN autosomal de FTDNA est appelé « *Family Finder* » et a pour but d'explorer la famille dans toutes les directions au lieu de suivre uniquement les lignées uniparentales paternelle ou maternelle. Ce test devrait donc livrer un certain nombre de cousins en fonction des autres clients ayant fait un test semblable ou qui ont importé un test fait par un autre laboratoire. Quand on regarde les résultats, on ne peut pas différencier l'origine de ces données, sauf si la personne a utilisé un nom différent.

À cause de la nature même du test autosomal, les résultats offerts prennent une forme différente. Dans cette section, on trouve de gauche à droite les correspondances « *Family Finder Matches* », un estimé des origines « *myOrigins* », un navigateur de chromosome « *Chromosome Browser* », le détail des origines selon les segments du chromosome « *Chromosome Painter* », et sous « *See More* » une matrice de comparaison « *Matrix* »,

des origines anciennes« *anciensOrigins* » et finalement un lien pour télécharger« *Data Download* » les résultats bruts. Voyons cela plus en détails.

Les correspondances « *Family Finder Matches* » prennent la forme d'un tableau.

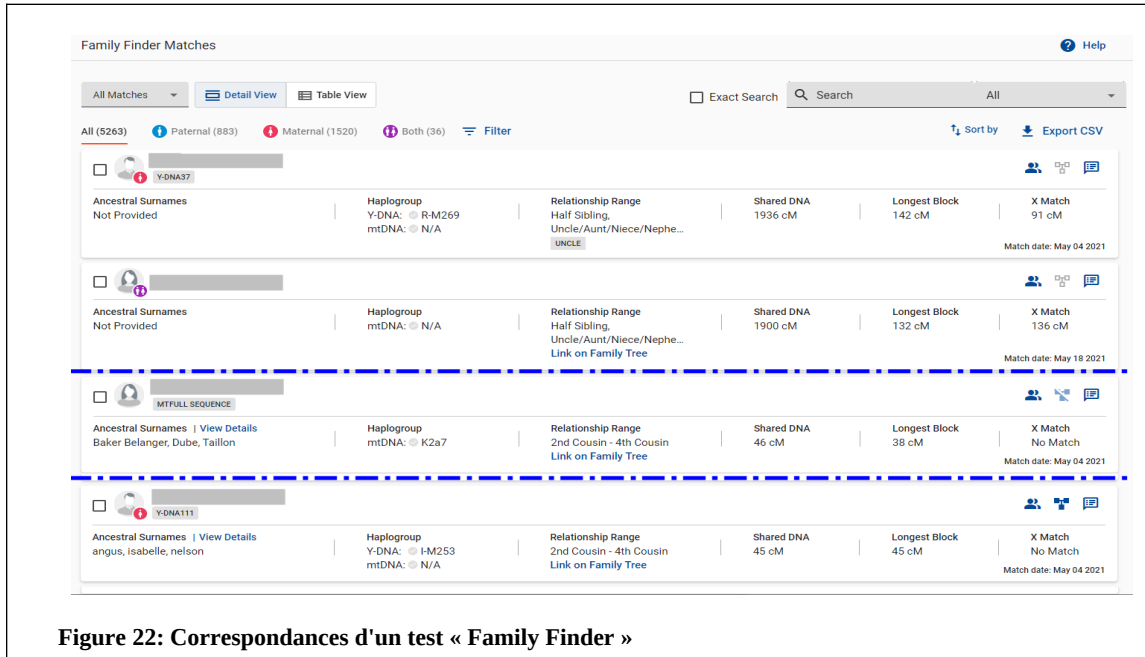


Figure 22: Correspondances d'un test « Family Finder »

Dans le haut du tableau, des liens permettent d'afficher tous les liens ou choisir le rapprochement (parenté immédiate, proche, éloignée, très éloignée, ADN X) et avec un choix entre deux formats (Détails et Tableau), ainsi que de rechercher des noms ou un haplogroupe. Au-dessous, d'autres filtres permettent de n'afficher que la parenté paternelle ou maternelle, de ne voir que ceux qui ont (ou n'ont pas) un arbre en ligne, ainsi que de trier de différentes façons et d'enregistrer les données dans un fichier CSV qui sera lisible avec un chiffrier. Au bas de la page, le lien « *Advanced Matches* » permet de choisir les correspondances dans un même projet.

Pour chaque correspondance, les données affichées peuvent comprendre le nom du client, les surnoms qu'il a inscrit dans son profil, ses haplogroupes s'il a fait des tests ADN Y ou MT, le lien familial calculé par FTDNA et au-dessous, le lien réel selon la position du match dans votre arbre en ligne, l'ADN partagé total, le plus long segment, un lien pour afficher les autres correspondances apparentées (ou non) avec ce match, une icône indiquant s'il y a un arbre en ligne, et finalement une note qui sera visible par vous et les administrateurs de vos projets.

L'estimé des origines « *myOrigins* » (version 3) est un tableau donnant des origines en pourcentages avec une carte illustrant les régions correspondantes en utilisant les couleurs. Dans le haut du panneau gauche, un lien permet de télécharger les résultats de

la version 2 pour comparaison. Dans le bas, le lien « *Show All* » permet voir toutes les populations de référence.

Le navigateur de chromosome « *Chromosome Browser* » montre les segments de chromosomes partagés avec entre un et sept autres clients. Par défaut, la liste de tous les matches est affichée avec un lien « *Download All Segments* » permettant d'avoir la position exacte des segments partagés. Pour chaque correspondance, on peut voir le nombre de segments partagés ainsi que le total et le segment le plus long. Il est aussi possible de voir les matches qui correspondent entre eux (« *In Common With* »).

Un autre tableau intéressant montre le détail des origines selon les segments du chromosome « *Chromosome Painter* », soit la région d'origine en fonction des autres clients du laboratoire. Il y a une différence entre les « régions » et les « continents ». On peut aussi choisir, avec le filtre, de n'afficher que les résultats d'une région choisie.

La matrice de comparaison « *Matrix* » montre les relations entre vos matches, dans le but de vous aider à trouver des personnes apparentées entre elles.

Les origines anciennes « *anciensOrigins* » se basent sur l'ADN archéologique et associe la génétique à trois groupes principaux, les chasseurs-cueilleurs, les fermiers et les envahisseurs de l'Âge de Fer.

Le lien pour télécharger « *Data Download* » permettra d'importer vos données sur d'autres sites comme myHeritage, Gedmatch, Geneanet et CopaQ, ainsi que des sites d'analyse. Il s'agit d'un fichier de plus de 600 000 lignes avec des valeurs comme A, C, G, T et « - » et le but n'est pas d'en lire le contenu ligne par ligne.

FTDNA permet l'importation de tests ADN autosomaux depuis certains laboratoires. Il faut lire la liste des versions permises et les télécharger depuis le site d'origine sans dézipper le fichier. Certaines fonctions sont disponibles pour un coût supplémentaire de 19 \$US.

### 8.4 Les projets

---

FTDNA offre une fonction très intéressante sous la forme de projets (ou de groupes, selon la page consultée). Un projet est une série de troupes. Les administrateurs regroupent les résultats Y et MT et les visiteurs peuvent alors voir si un client a déjà des résultats à leur nom (ou au nom de leurs ancêtres) et ensuite, essayer de trianguler (ou de confirmer la signature si elle n'est pas encore triangulée) ou faire un test pour s'ils sont de la même famille.

Un projet comme Héritage français (*French Heritage*) contient en 2023 des résultats venant de plus de 11 000 troupes. Si quelqu'un a une incertitude sur ses propres

résultats, il suffit de rejoindre le projet et de les placer avec ceux d'un autre test pour comparaison.

Depuis les années 2010, les données s'accumulent pour former un catalogue de signatures. Dans sa version actuelle, le catalogue donne le nom des ancêtres communs dont deux clients descendent selon une lignée documentée et par deux fils (ADN Y) ou deux filles (ADNmt). La signature est alors validée jusqu'à cet ancêtre. Le catalogue utilise des couples afin qu'il n'y ait pas d'ambiguïté sur le pionnier en question. S'il y a de la descendance par deux pionniers, les deux noms seront affichés.

Rejoindre un projet peut se faire lors de l'achat d'un test ou plus tard. Depuis son profil personnel, plusieurs chemins permettent de choisir un projet.

- Menu du haut, **Group Projects -> Join A Project**
- Au bas de la page, sous **COMMUNITY -> Group Projects**
- Dans la colonne droite, **Join a Project**

Le projet **French Heritage** se trouve plus bas, dans la section

#### **DUAL (Y-DNA AND MTDNA) GEOGRAPHICAL PROJECTS**

sous la lettre F. Une fois que vous avez rejoint le projet, vous devez donner un accès, de préférence « limité » afin que vos données soient examinées pour être classées dans un des sous-groupes ou éventuellement dans un nouveau sous-groupe.

Il faut certaines conditions pour être regroupé dans les projets de l'auteur. D'autres projets peuvent avoir d'autres exigences (les projets par haplogroupe, par exemple, demande que l'haplogroupe corresponde au projet).

- Il doit s'agir de résultats Y ou MT (sauf un ancien test Geno 2 qui aurait été avant la fin des opérations);
- Les résultats doivent être prêts (il faut commander le test, retourner la trousse et faire l'analyse)
- Le regroupement doit être possible. Il faut identifier l'ancêtre ancien dans son profil ou avoir un arbre en ligne, ou avoir un nom de famille pour lequel il y a déjà une signature similaire.

Il est préférable d'organiser d'avance ses résultats pour être regrouper rapidement. Sinon, vous pouvez aussi contacter un administrateur de projet avec les informations nécessaires. C'est aussi utile d'indiquer que vous recherchez le père biologique d'un grand-père, par exemple, afin de ne pas être classé avec une famille biologique autre que la vôtre.

Pour bien configurer votre profil, il faut aller dans votre profil (*Account Settings*), sous votre nom ou photo en haut à droite, puis :

- Sous **GENEALOGY - Earliest Known Ancestors**, le nom de vos ancêtres connus les plus anciens, Paternel et Maternel. Attention, l'ancêtre maternelle est la mère de la mère de la mère etc. de votre mère, uniquement en ligne féminine. Au Québec et en Acadie, je suggère le nom de l'ancêtre immigrant, son conjoint et l'année du mariage pour éviter toute ambiguïté. En Europe, le nom de l'ancêtre le plus lointain, son conjoint, l'année du mariage et le pays ou département français (comme FR-38). Cette information sera aussi affichée dans les résultats publics et sera visible par vos matches.
- Sous **GENEALOGY - Surnames**, les noms de famille des ancêtres, ce qui aide parfois au regroupement.
- Sous **PRIVACY & SHARING**, il faut accepter les matches [**Opt in to Matching**]. Le paragraphe **IGGM** concerne l'utilisation de vos données par la police. Plus bas, il faut aussi choisir les bons niveaux pour les matches, avec [**Match Levels**]. Certains ont trop de matches avec le test Y-12 et cachent donc ces résultats, mais en général, on devrait tout accepter car il est aussi possible qu'un seul client ait un résultat compatible avec le vôtre et qu'il ait fait son test avant le développement des tests plus précis d'aujourd'hui.
- Sous **PROJECT PREFERENCES**, vous verrez la liste de vos projets, de leurs administrateurs, et des permissions que vous donnez à chacun.
  - « **Minimum** » signifie que vous n'avez pas confiance dans cet admin, « **Limité** » est le fonctionnement normal permettant de vous regrouper plus facilement ou d'étudier vos résultats, et « **Avancé** » permet aux admins de changer certains éléments critiques (vous retirer d'un autre projet ou modifier votre généalogie en ligne). Le but du mode « **Avancé** » est en fait de permettre à un projet de payer un test, parfois suite à la commandite d'un autre participant. Cela ne donne rien d'envoyer votre mot de passe à un admin.
  - **Opt in to Sharing** a pour but de permettre l'affichage de vos résultats dans les pages publiques des projets. Il y a un seul endroit pour tous les projets.
  - **Coding Region Sharing** vise les tests ADNmt seulement. Cela a pour effet de montrer la partie centrale du génome MT aux administrateurs de projets.
- L'onglet **NOTIFICATION PREFERENCES** montre les options concernant les messages envoyés par courriel quand il y a de nouveaux matches. Selon le test,

vous pouvez opter de ne pas recevoir, par exemple, les avis de nouveaux matches pour le niveau HVR1 d'ADNmt ou le niveau Y-12 d'ADN Y. Ceci est pratique pour certains résultats dans des catégories très populaires. Si vous avez deux courriels dans votre profil, seul le premier recevra ces notifications.

#### 8.4.1 Résultats publics

---

Si vous avez coché les bonnes cases, vos résultats devraient apparaître parmi les résultats publics du projet choisi. Dans mes projets, il y a souvent trois sections pour les nouveautés.

La section **Ungrouped** (trouvée dans tous les projets) réunit les résultats avant leur regroupement.

Une section comme **Zz Observer** sert à classer ce qui n'est pas directement relié au sujet du projet tout en conservant la trousse. Le Y et le MT sont indépendants, alors certains ont un résultat Y avec une origine française et un MT avec une origine différente. Pour les noms de famille, une telle section peut réunir les noms qui ne ressemblent pas aux principaux noms. Ils peuvent se retrouver dans une telle section tout en pouvant utiliser les autres fonctions reliées aux projets. Dans tous les cas, il est possible de contacter un admin et de lui demander d'être ajouté à tel sous-groupe pour comparer les résultats. Certains font un test ADN pour trouver leurs parents biologiques ou pour débloquer une impasse documentaire. Les admins peuvent alors donner un coup de pouce.

Finalement, selon la configuration du client ou les informations disponibles, il n'est pas évident de classer un résultat. Il peut donc y avoir des sections temporaires, avant une étude plus poussée (par exemple, « **ZyH002 Not grouped H2\* to H5\*** » reçoit les haplogroupes MT commençant par H2, H3, H4 et H5 en attendant leur regroupement après étude ou si le client ajoute les informations permettant de mieux le classer.

Dans *French Heritage*, les catégories non-groupées de l'ADNmt utilisent les préfixes

- **Zw (HVR1)**,
- **Zx (HVR1+HVR2)** et
- **Zy (FMS)**.

Les catégories non-groupées de l'ADN Y sont :

- **Zx ADD documented paternal ancestor** (il faut documenter pour regrouper)
- **Zx GIVE Limited or Advanced access to admins** (les admins doivent voir les résultats)



## Le laboratoire FTDNA

BARBIER, Jeanne m 1670 Julien PLUMEREAU dit LATREILLE (GFAN 1876)(TR00320)						
Mi41205	Jeanne BARBIER m 1670 Julien PLUMEREAU	France	H	A16129G, T16187C, C16189T, C16222T, T16223C, G16230A, T16278C, C16311T	G73A, C146T, C152T, C195T, A247G, T319C, 522.1A, 522.2C, 309.1C, 309.2C, 315.1C	
BARBIER, Marie m 1637 Nicolas MARSOLET (GFAN 3613)(TR00095)						
B49249	Marie Le Villain, b. abt 1585 Rouen, FR	France	H1ag	A16129G, T16187C, C16189T, T16223C, G16230A, T16278C, C16311T	G73A, C146T, C195T, A247G, 522.1A, 522.2C, 309.1C, 315.1C	
257459	Marie Levillain c1582 Rouen	France	H1ag	A16129G, T16187C, C16189T, T16223C, G16230A, T16278C, C16311T	G73A, C146T, C195T, A247G, 522.1A, 522.2C, 309.1C, 315.1C	
BARDOU BARDEAU, Marie m 1669 Mathurin DROUET dit GRANDMAISON (GFAN 877)(TR00709)						
980127	Elisabeth Leclerc, b. Abt 1617 and d. Bef 1669	France	H3k	A16129G, T16187C, C16189T, T16223C, G16230A, T16278C, C16311T	G73A, C146T, C195T, A247G, 522.1A, 522.2C, 309.1C, 315.1C	
BASILE, Perrine m 1685 Andre CELESTIN (GFAN 86143)						
567203	Perrine Basile	France	H	A16129G, T16187C, C16189T, G16230A, T16278C, C16311T, C16519T	G73A, C146T, C152T, C195T, A247G, 522.1A, 522.2C, 315.1C	
BASTINAUD dit PELLETIER, Louise m 1684 Louis SAULNIER (GFAN 96985)(TR00487)						
912492	Louise Bastineau dit Peltier b1665 and d1709	France	H1aa	A16129G, A16183c, T16187C, T16223C, G16230A, T16278C, C16311T	G73A, C146T, C152T, C195T, A247G, 522.1A, 522.2C, 315.1C	
411130	Louise Bartineau dit Peltier b ca 1668 d ca 1740	France	H1aa	A16129G, A16183c, T16187C, T16223C, G16230A, T16278C, C16311T	G73A, C146T, C152T, C195T, A247G, 522.1A, 522.2C, 315.1C	

**Figure 24: Résultats publics d'un test ADNmt**

### 8.4.4 Les numéros de trousse

Les numéros de trousse permettent d'identifier un test de façon anonyme. Il n'est toutefois pas possible de voir la trousse d'un autre client sans qu'il ait rejoint un projet. C'est donc via les projets que le visiteur pourra savoir quels résultats sont associés à telle trousse. Un administrateur peut aussi voir certains résultats en passant par une autre trousse. Ainsi, si une trousse contient le test ADNmt d'un descendant de telle pionnière, on pourra voir le nom des autres clients qui ont accepté de participer aux correspondances (*matching*) mais on ne pourra pas voir leur numéro de trousse.

Quant aux numéros en soi, ils sont composés d'un préfixe et d'un numéro séquentiel. L'ISOGG a publié une liste des préfixes<sup>57</sup> dont voici un aperçu :

Préfixe	Signification
(rien)	Achat régulier via FTDNA
AM	Achat via Amazon
B	Transfert depuis Ancestry ou 23andme
BP	Achat en vrac
E	Achat via iGenea
IN	Achat international (en dehors des États-Unis, à partir de novembre 2017)

<sup>57</sup> [https://isogg.org/wiki/FTDNA\\_kit\\_prefixes](https://isogg.org/wiki/FTDNA_kit_prefixes)

	environ)
MH	Achat via myHeritage
MI	Achat international de plusieurs trouses
MK	Achat de plusieurs trouses
N	Test Geno 1 ou Geno 2 via National Geographics

## 8.5 Exporter et importer des données

---

### 8.5.1 Exporter des données

---

FTDNA offre plusieurs sortes de tests et donc de données qui peuvent être exporter vers d'autres sites.

#### 8.5.1.1 Données Family Finder

---

Avec ce test autosomal, les données brutes peuvent être exportées, ainsi que des listes de correspondants pour analyse (en général, il est préférable de rejoindre un projet pour demander de l'aide pour comprendre les résultats).

Pour les données brutes, depuis l'écran d'accueil, section [**Autosomal DNA Results**], lien [**See More**], lien [**Data Download**]. "Build 37" signifie que les données sont identifiées selon la version 37 du génome humain. La version 38 est parfois utilisée par d'autres sites et la conversion sera automatique. La version de 2023 est appelée T2T et ne change pas la partie autosomale. Les options sont :

- Build 37 Autosomal : chromosomes 1 à 22
- Build 37 X Chromosome : chromosome X (ou XX pour les femmes)
- Build 37 Concatenated : tous les chromosomes, sauf X et MT. En version GZIP, c'est le fichier à exporter.
- GZIP : version compressée, à utiliser pour exporter vers d'autres sites
- CSV : version en texte, lisible par un humain (mais plus de 600 000 lignes)

#### 8.5.1.2 Données ADNmt

---

L'ADN des mitochondries est une liste d'environ 16569 valeurs. Comme des insertions et retraits sont possibles, ce nombre peut varier.

Pour les données brutes, depuis l'écran d'accueil, section [**mtDNA Results**], lien [**See More**], lien [**Mutations**]. Les onglets [RSRS] et [rCRS] montrent les différences entre la

référence (RSRS est l'Ève africaine, et rCRS, l'Ève européenne) et vos résultats. Le lien [FASTA] au bas de l'écran donne le fichier avec les 16569 valeurs lues.

### 8.5.1.3 Données ADN Y

---

Il existe deux tests différentes pour l'ADN Y : STR et SNP.

Pour les STR, depuis l'écran d'accueil, section [Y DNA Results], lien [See More], lien [Y STR Results]. Vous obtenez dans un premier temps, les données brutes par panneau, soit les valeurs lues qui sont des chiffres. "-" signifie que la valeur n'a pu être lue. Au bas de l'écran, un lien est affiché : [Download CSV].

Les SNP peuvent prendre deux formes. Le test Big Y montre les résultats d'une recherche de milliers de SNP alors que les paquets de tests (*SNP Packs*) se concentrent sur une centaine de SNP seulement, en fonction de ce que vous avez déjà comme résultats.

Pour le Big Y, depuis l'écran d'accueil, section [Big Y], lien [Results]. Le tableau obtenu contient différents filtres pour afficher les résultats disponibles. Dans le haut de la page, la section intéressante est appelé [Private Variants] et montre les SNP les plus récents, à moins d'avoir un match proche. La section [Matching] affiche les SNP partagés avec les matches les plus proches. Si le SNP est une série de chiffres (sans préfixe alphabétique), il s'agit d'un variant non nommé, souvent parce qu'il n'a pas été découvert chez plusieurs clients et qu'il n'est donc pas assez fiable.

Un variant non détecté est appelé [No Call]. Dans les autres cas, ce sera la valeur lue.

Dans le haut de l'écran se trouve un lien [Download Files]. Il permet d'exporter sous forme de fichier CSV plusieurs sortes de résultats (*Matches*, SNP dérivés ou recherchés, SNP ancestraux, SNP non détectés), ainsi que les fichiers VCF et BAM. Le fichier VCF contient la différence entre votre chromosome Y et la référence, alors que le fichier BAM contient votre génome Y complet, sous forme de segments lus et avec un format assez complexe. Depuis quelques années, il y a des frais de 99 \$US pour générer le fichier BAM, ceci compensant la baisse importante du prix du test Big Y.

### 8.5.2 Importer des données

---

Autrefois, FTDNA permettait l'importation de données d'autres laboratoires mais cette pratique est abandonnée depuis longtemps. Aujourd'hui, les importations sont limitées aux tests d'ADN autosomal de certains laboratoires.

Ceci se fait en exportant d'abord les données depuis l'autre laboratoire. Il faut conserver le fichier reçu qui sera sous une forme compressée (ZIP par exemple).

## Le laboratoire FTDNA

Ensuite, il faut créer un compte FTDNA pour chaque fichier qui sera importé. Si on importe des résultats AncestryDNA et 23andme, il faudra donc créer deux comptes FTDNA. Finalement, il faut suivre les instructions pour importer les données brutes compressées.

### 8.6 Versions

---

Le tableau suivant montre les différentes versions du test fait par FTDNA.

Version	Début <sup>58</sup>	Fin	Lignes <sup>59</sup>
1		Février 2011	725 276
2	Février 2011	Avril 2019	627 076
2a	Janvier 2011		725 291
2b	Janvier 2015		708 167
2c	Janvier 2016		715 977
2d	Avril 2016		720 350
3	Avril 2019		

---

<sup>58</sup> Source: <https://h600.org/wiki/Microarray+File+Formats>

<sup>59</sup> Source: [https://isogg.org/wiki/Autosomal\\_SNP\\_comparison\\_chart](https://isogg.org/wiki/Autosomal_SNP_comparison_chart)

## 9 Le laboratoire 23andme

---

Le laboratoire 23andme (pour 23 paires de chromosomes et moi) est surtout utilisé pour sa partie médicale mais le test généalogique est tout de même assez intéressant en soi et sa base de données est en 2023 la seconde importance pour ce qui est du nombre de clients.

### 9.1 Introduction

---

23andme a offert, depuis ses débuts, 5 versions de son test ADN. Depuis la version 4, les haplogroupes Y et MT des clients sont disponibles, ainsi que les positions des segments dans les chromosomes, ce qui permet parfois d'identifier le lien de parenté entre deux matches et de compenser ainsi la rareté des informations généalogiques.

Ce laboratoire n'a qu'une version anglaise.

À la fin de 2023, une grande quantité de résultats ADN ont été volé chez ce laboratoire, ce qui a entraîné la suspension de plusieurs services ainsi qu'une fragilisation des finances. Ainsi, il n'est plus possible de télécharger les listes de cousins et la position des segments partagés avec les autres clients. FTDNA a aussi cessé l'importation des données de ce laboratoire. 23andme a aussi ajouté la double authentification.

### 9.2 Les tests disponibles

---

#### 9.2.1 Le test généalogique

---

Le test de base de 23andme est essentiellement généalogique. Il comprend, dans sa version 5, un échantillonnage de 682 549 ou 630 132 valeurs (selon la page de ISOGG.ORG) dont 16 521 dans le chromosome X, 3 733 dans le Y et 4 318 dans l'ADNmt. Les haplogroupes Y, réservé aux hommes, et MT, pour tous, sont affichés directement dans les résultats, dans le profil d'un correspondant ayant fait une version assez récente et dans un fichier appelé «*Relatives*». Le test de base trouve jusqu'à 1500 cousins et lorsque de nouveaux cousins plus proches sont découverts, les plus anciens sont retirés de la liste, mais si le pointeur a été sauvegardé (dans le fichier «*Relatives*»), ces données sont accessibles.

Le test permet aussi l'affichage d'un arbre généalogique avec les cousins les plus proches, en utilisant les relations entre les cousins pour distinguer les ancêtres communs, sans toutefois nommer ces ancêtres communs. Il est possible d'éditer cet arbre généalogique avec des outils rudimentaires qui permettent de tenir compte de l'endogamie (pour rattacher un ancêtre à deux endroits dans l'arbre).

### 9.2.2 Le test médical

Ce sujet n'est pas abordé en profondeur dans la présente publication. Il faut toutefois retenir que, lors de l'achat, la partie médicale est offerte pour environ le double du prix d'achat. Si vous ne voulez que la partie généalogique, il faut donc choisir le test de base. De plus, le test médical comprend un abonnement annuel payant qui permet d'identifier environ 4500 cousins en tout, au lieu de 1500.

### 9.3 Visite guidée

L'auteur a commencé à utiliser 23andme seulement en 2022. Certaines fonctions ne sont pas évidentes à trouver et il faut parfois donner une permission pour avoir un accès complet au contenu.

#### 9.3.1 Panneau d'entrée

La première page affichée sur le site 23andme est un tableau des différents rapports et études disponible avec un menu latéral et un autre dans le haut de l'écran pour voir les informations généalogiques. Les pages affichées sont différentes dans ces deux menus. Commençons avec le menu latéral situé à droite (*Quick Links*).

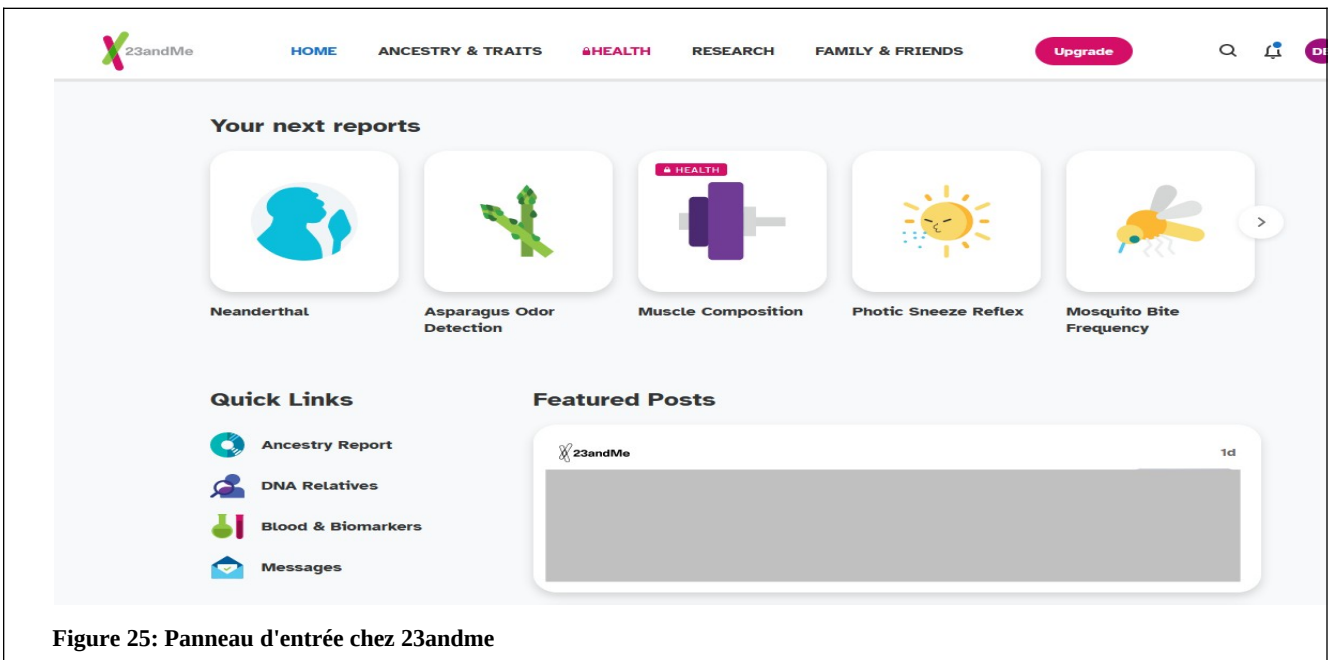


Figure 25: Panneau d'entrée chez 23andme

#### 9.3.2 Ancestry Report

Il s'agit d'un aperçu de la composition ancestrale du client, soit le pourcentage estimé de chaque région mentionnée avec un découpage régional accompagné d'une carte

illustrant les régions concernées. La France est divisée en 13 régions et selon les résultats de l'auteur, dont l'ascendance est presque entièrement française, selon 23andme, 73,3% de son ADN serait surtout français (d'après les régions indiquées), 25,4% espagnol ou portugais (il doit s'agir en réalité de l'ouest de la France, sous-testée), avec des traces d'ADN amérindien (0,3%), chypriotes (0,3%), chinois du Nord et tibétains (0,2%) et Asie centrale (0,2%). Il faut toujours garder en tête que ces tests se font avec une comparaison du reste de la clientèle et non de l'étude de sites archéologiques.

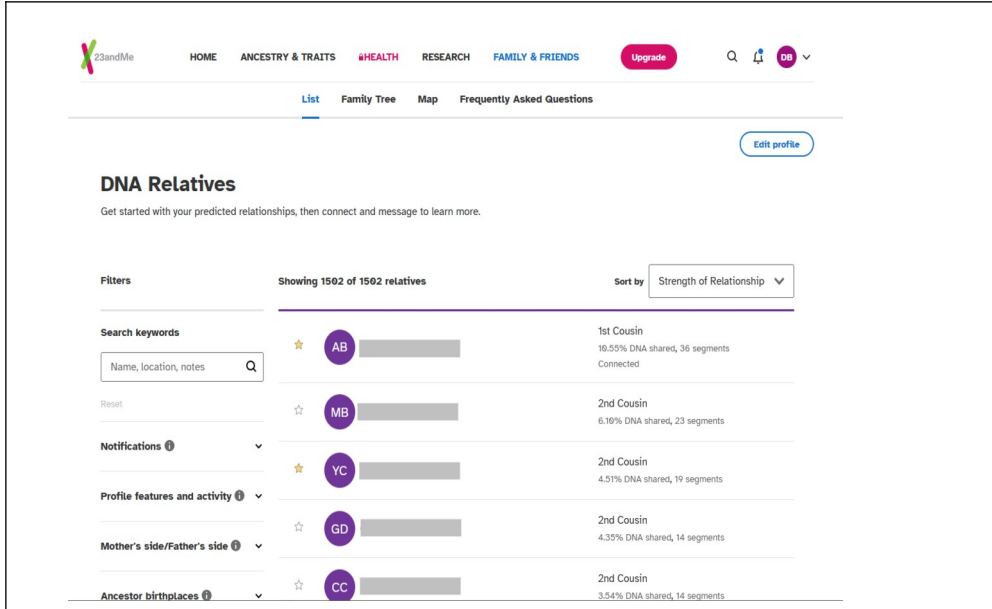
### 9.3.3 DNA Relatives

Les cousinages sont affichés si le client le permet. Ce système fonctionne dans les deux sens avec tous les laboratoires (vous pouvez voir chez les autres ce que ces autres peuvent voir à votre sujet). Avec 23andme, toutefois, il faut l'activer vu que le laboratoire offre aussi un test médical. Dans le profil, il faut donc répondre **YES** à cette question:

Would you like to display your ancestry results and matching DNA segments to DNA Relatives participants?

**(X) Yes, show my ancestry results and matching DNA segments.**

Cela vous donnera un écran similaire au suivant:



The screenshot shows the 23andMe DNA Relatives interface. At the top, there's a navigation bar with 'HOME', 'ANCESTRY & TRAITS', 'HEALTH', 'RESEARCH', and 'FAMILY & FRIENDS'. Below that, a secondary navigation bar includes 'List', 'Family Tree', 'Map', and 'Frequently Asked Questions'. The main heading is 'DNA Relatives' with a subtext: 'Get started with your predicted relationships, then connect and message to learn more.' There are filters on the left for search keywords, notifications, profile features, and birthplaces. The main area shows a list of relatives, sorted by 'Strength of Relationship'. The first five relatives are:

Name	Relationship	DNA Shared	Segments
AB	1st Cousin	16.55%	36 segments
MB	2nd Cousin	6.16%	23 segments
YC	2nd Cousin	4.51%	19 segments
GD	2nd Cousin	4.35%	14 segments
CC	2nd Cousin	3.54%	14 segments

Figure 26: Liste des correspondants chez 23andme

L'élément principal de cette page est constitué par la liste de vos 1500 «relatives» les plus proches avec leur nom, un lien de parenté suggéré par les résultats, ainsi que l'ADN

partagé (en pour-cent) et le nombre de segments. On peut se connecter avec un autre client et avoir un peu plus d'informations à son sujet. Une étoile permet d'annoter certains cousins. En cliquant sur le nom du cousin, on peut voir des informations supplémentaires à son sujet.

- Des données personnelles saisies par le client sont affichées (année de naissance, niveau d'activité sur le site, situation, texte «à propos»)
- Une demande de connexion est possible.
- On peut envoyer un message via la messagerie interne
- On peut voir la relation génétique calculée selon l'ADN partagée. En cliquant sur «View DNA details», on voit l'ADN partagé ou les segments partagés si la personne est connectée. Il est aussi possible de comparer ce tableau avec d'autres cousins (selon les permissions données par ceux-ci). Fonction désactivée en partie suite au vol de données.
- Si la personne est assez proche et apparaît dans l'arbre généalogique du site (voir plus loin), on peut voir cet arbre mais il n'est pas positionné sur cette personne si on clique.
- L'environnement familiale (*Family Background*) comprend le lien de naissance des grands-parents, les patronymes familiaux et un lien vers un arbre généalogique externe.
  - Une autre section permet de comparer les origines génétiques (*Ancestry Composition*) si la personne est connectée. Fonction désactivée suite au vol de données.
  - Les haplogroupes maternels et paternels sont ensuite comparés et commentés. Selon l'ADN échantillonné, ces haplogroupes sont plus ou moins précis.
  - Les Européens ont une quantité faible d'ADN néandertalien. Un chiffre résume votre position par rapport aux autres clients.
- La fonction suivante permet de comparer la parenté en commun. Ce tableau donne, pour vous-même et le cousin choisi, l'ADN partagé, le niveau de parenté estimé et si l'ADN se recoupe (*DNA Overlap*). Un «oui» signifie que le cousin choisi et le match affiché ont de l'ADN en commun, alors qu'un «non» suggère que la relation n'est pas du même côté de la famille (paternelle au lieu de maternelle, par exemple). Fonction désactivée suite au vol de données.

- La page se termine avec une note pour vos propres besoins et qui n'est pas visible par l'autre client.

En marge gauche, des filtres permettent de trouver plus rapidement un cousin, un endroit ou une information saisie dans une note. On peut aussi ne voir que ceux qui ont des données généalogiques, les favoris (ceux avec une étoile jaune), ceux qui ont envoyé ou reçu des messages. Les côtés maternels et paternels demandent qu'un test ait été fait par un parent. Un cousin maternel ou paternel n'est pas suffisant, bien que ces côtés seront visibles chez vos cousins proches avec l'arbre en ligne.

D'autres statistiques sont également disponibles.

#### 9.3.4 Liste des cousins

---

Fonction désactivée suite au vol de données.

Tout en bas de la page, un lien permet de demander la liste des cousins «*Request DNA Relatives Data Download*». Il est préférable de demander cette liste de temps en temps afin de conserver le lien vers les cousins les plus lointains. Même si cela n'est pas utile directement pour identifier ces cousins et trouver leur généalogie, un lien éloigné donne parfois un indice pour identifier un autre cousin plus proche et ainsi faire sa généalogie.

Vous recevrez ensuite un message disant que la liste est prêt et le lien au bas sera renommé comme: «*Download DNA Relatives Data* ». Cela donnera une page permettant de télécharger plusieurs informations différentes, dont la section «*DNA Relatives Data*». Le fichier indique, pour chaque segment partagé, le nom du cousin, son patronyme (*Surname*), le chromosome concerné, le début et la fin du partage, une distance génétique, le nombre de SNP dans le segment, un lien vers le profil permettant de consulter les données des cousins devenus trop éloignés), l'année de naissance, les relations choisies (en éditant l'arbre en ligne si-dessous), prédite et une plage, l'ADN partagé et le nombre de segments, les côtés maternel et paternel si un parent est testé, les haplogroupes MT et Y (selon la version du test), les noms de famille du client et positions géographiques donnés par le client, les pays de naissance des grands-parents, les notes, la connexion, l'affichage des résultats ancestraux (les origines) et finalement un lien vers la généalogie en ligne sur un autre site.

Vous pouvez partager ce fichier avec l'auteur, ce qui permet parfois d'identifier de nouvelles lignées et des signatures dans certaines conditions.

Avec la version de base, en 2023, cette liste de cousins ou *relatives* est limitée à 1500 et certains rapportent 1501 ou 1502, un nombre qui est constant : quand de nouveaux

## Le laboratoire FTDNA

cousins sont découverts, les plus anciens sont enlevés. Avec la version médicale, la liste semble plutôt limitée à 5000. Le contenu de cette liste est le suivant :

Étiquette	Description
Display Name	Nom complet du client
Surname	Patronyme du client
Chromosome Number	Numéro du chromosome
Chromosome Start/ End Point	Début et fin de ce segment
Genetic Distance	Distance génétique (signification inconnue)
# SNPs	Nombre de valeurs partagées dans ce segment
Full IBD	Segment identique (pour <i>Identity by descent</i> ) <sup>60</sup>
Link to Profile Page	Lien vers le profil sur 23andme <sup>61</sup>
Sex	Genre du client (champ vide)
Birth Year	Année de naissance saisie par le client
Set Relationship	Relation selon votre arbre modifié
Predicted Relationship	Relation prédite selon vos résultats (mais peut être modifiée selon votre arbre modifié)
Relative Range	Plage de relations prédites
Percent DNA Shared	ADN partagé en pourcentage. On peut multiplier par 7439 pour la valeur en cM (dans un chiffrier en français, remplacez le point par une virgule)
Segments Shared	Nombre de segments partagés
Maternal Side/Paternal Side	Lien avec les parents si le père ou la mère a fait le test ADN de 23andme
Maternal Haplogroup	Haplogroupe MT ou maternel, prédit, pour les versions 4 et 5
Paternal Haplogroup	Haplogroupe Y ou paternel ancien, pour les versions 4 et 5 et si le client est un homme
Family Surnames	Patronymes saisis par le client
Family Locations	Lieux saisis par le client, parfois récents, parfois anciens
Maternal/Paternal Grandmother/Grandfather Birth Country	Code du pays de naissance des grands-parents
Notes	Notes saisies par vous
Sharing Status	Statut de connexion (si vous ou l'autre client l'avez demandé)
Showing Ancestry Results	Affichage des résultats ancestraux (semble lié aux versions 4 et 5)

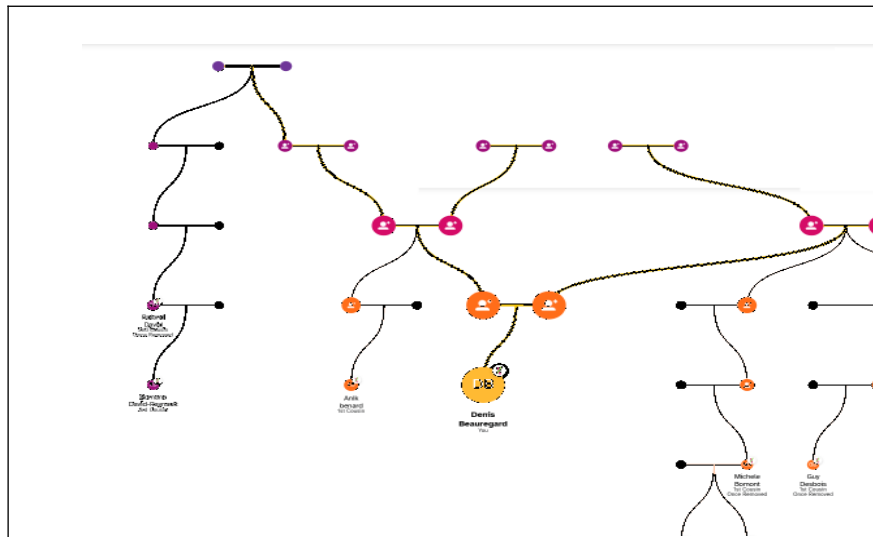
<sup>60</sup> Si le segment commun n'est pas au début du chromosome

<sup>61</sup> Ce lien permet de revoir le profil d'un cousin éloigné, lorsqu'il n'est plus dans la liste des *relatives*

Family Tree URL	Lien vers l'arbre généalogique extérieur
-----------------	--

### 9.3.5 Arbre en ligne (lien «*Family Tree*»)

Un arbre en ligne est généré automatiquement avec le lien «*Family Tree*» en haut de la page des *relatives*. La version brute donne la position théorique des cousins les plus proches, les uns par rapport aux autres. L'ADN n'étant pas partagé de façon uniforme et l'endogamie étant possible, certaines relations peuvent être un peu différentes de la réalité.



Cet arbre peut être édité pour ajouter les noms de votre parenté et déplacer parfois des branches pour ajouter ou retirer des générations. Si vous décidez de refaire l'arbre à zéro (en passant par la FAQ), vous perdrez toutes vos corrections et il faudra les refaire au complet.

### 9.3.6 Section *Ancestry & Traits*

Le reste de la page d'accueil contient surtout de l'information médicale et un lien vers la messagerie.

Dans le haut de la page, le lien «*Ancestry & Traits*» ajoute d'autres fonctions:

- La section «*Ancestry & Trait Overview*» montre différents rapports sur votre ADN regroupés sous «*All Ancestry Reports*» et «*Ancestry Composition*».
- La section «*DNA Relatives*» est la même que dans le menu latéral.

- La section «*Traits*» contient des statistiques sur plus de 30 caractéristiques que vous pourriez avoir, comme la probabilité d'avoir des pellicules, du poil dans le dos, de perdre vos cheveux, ou encore la couleur des yeux.

## 9.4 Utilisation pour la généalogie

---

Le test de 23andme peut-il aider à valider sa généalogie et à savoir si notre génétique correspond à notre généalogie documentée ?

### 9.4.1 Les haplogroupes Y

---

Le catalogue de signatures de comparaison permet d'avoir une signature de référence pour presque la moitié de la population du Québec ou avec des racines acadiennes paternelles. Le SNP donné par 23andme n'est pas souvent identique à celui affiché au catalogue mais un expert pourra facilement vérifier si votre résultat est cohérent avec celui du catalogue si vous fournissez votre lignée paternelle et votre fichier *Relatives*. Cette liste est limitée aux 1500 cousins les plus proches, à moins d'avoir fait aussi le test médical. Cela se limite toutefois à 4500 matches et donc une partie seulement de la population possible. Ces limites sont augmentées si vous êtes connectés à d'autres.

Le test sur l'ADN Y comprend la recherche de 3 733 SNPs dans le chromosome Y. En moyenne, il vous donnera le SNP d'un ancêtre ayant vécu il y a environ 5000 ans. Dans certains cas, le nombre de pionniers du Québec ou d'Acadie ayant ce SNP est élevé et dans d'autres cas, plutôt faible. Par exemple, avec R-P312, le nombre d'immigrants est très élevé et cela ne suffira pas pour distinguer un grand nombre de familles. Par contre, si le pionnier est R-A431 et que 23andme trouve I-M223, alors, il est certain que ce n'est pas cohérent.

Un adopté verra une série possible de patronymes parmi ceux ayant le même haplogroupe, ou en communiquant avec l'auteur, d'autres possibilités, mais avec deux limitations. D'abord, une partie seulement des pionniers a fait l'objet d'une triangulation et les autres ont parfois un ancêtre adopté entre la personne testée et l'immigrant. Ensuite, le père biologique peut n'avoir aucune origine française et donc ne pas être concerné par un catalogue visant surtout l'immigrant d'avant 1800.

### 9.4.2 Les haplogroupes MT

---

Le catalogue de signatures de comparaison permet d'avoir une signature de référence pour environ la moitié des descendants de nos pionnières venues sous le régime français. Le test de 23andme n'est toutefois pas aussi précis que celui de FTDNA et il faut en tenir compte lors de comparaisons.

D'un côté, l'haplogroupe donné par 23andme n'a pas toujours une valeur identique. Ainsi, un haplogroupe comme K2b1a avec un test FMS devient K avec le test HVR2,

K2b1 avec le test Geno 2 et K2b avec celui de 23andme. Il s'agit d'une même pionnière mais d'analyses incomplètes et donc moins précises par rapport au test FMS.

D'un autre côté, le test de 23andme comprend les valeurs brutes de l'ADNmt, soit les SNP MT qui ont été détectés. Un expert peut parfois évaluer l'haplogroupe d'un résultat et faire une comparaison si le résultat de 23andme se limite à une lettre et que la signature de comparaison comprend d'autres valeurs mais avec également une seule lettre.

### 9.4.3 Les autres lignées

---

Il y a plusieurs façons d'analyser les résultats du test ADN de 23andme pour vérifier des lignées.

Si la personne testée a un arbre généalogique déjà documenté et connaît ses ancêtres, on peut examiner les matches. L'information reçue des correspondants varie beaucoup. Dans le meilleur des cas, un lien vers un arbre généalogique extérieur, le lieu de résidence, la section «À propos» ou les autres informations fournies par le client donnent assez de données pour faire l'arbre généalogique de la personne en question. Parfois, c'est un avis de décès qui permettra de compléter l'information, ou encore le nom d'un parent en examinant l'ADN partagé entre cette personne et vos autres cousins.

Il faut donc faire un arbre généalogique sur 5 ou 6 générations des personnes trouvées. Ensuite, on peut valider si cet arbre est bon en regardant leurs haplogroupes, comme expliqué plus haut. Finalement, on regardera si l'ADN partagé est cohérent avec la parenté trouvée en comparant les arbres généalogiques. Une cohérence n'est pas toujours parfaite. Si les cousins viennent de la même région, il se peut que l'ADN partagé soit plus élevé que si on se base sur un seul couple d'ancêtres communs parce que d'autres couples peuvent aussi contribuer au cousinage. Au Québec, en Acadie, en Louisiane, dans les régions où un petit nombre de familles sont à l'origine de la plupart des gens, l'effet fondateur a comme résultat que la parenté estimée selon l'ADN est plus forte que si on se limite à la documentation, même si les autres cousinages sont plus éloignés. De plus, l'ADN n'est pas réparti de façon uniforme et l'ADN de certains ancêtres a laissé une trace plus forte.

Dans le cas de l'auteur, toutes les personnes mentionnées dans l'arbre proposé par 23andme ont pu être identifiées et reliées à leur position, avec 2 exceptions (une personne probablement adoptée et une autre ne donnant que l'initiale de son prénom). Deux autres lignées semblent à première vue problématiques parce que le rattachement fait par 23andme les place à un endroit différent de l'arbre et il est difficile de savoir s'il s'agit d'une erreur en faisant leur ascendance, d'une mauvaise identification (un anonyme par exemple), d'une adoption donnant un lien familial différent, etc.

#### 9.4.4 Débloquer la lignée d'un adopté

---

Peut-on utiliser les données d'un test 23andme pour débloquer une personne ?

La méthode consiste à construire les arbres généalogiques des cousins trouvés, puis à trouver les ancêtres qui apparaissent dans plusieurs arbres. On compare également entre eux les cousinages des matches. Il est aussi utile de valider leurs haplogroupes respectifs afin de valider notre travail.

Parfois, il est possible de contacter une autre personne et de comparer les matches de cette personne afin de mieux cibler les ancêtres communs.

Dans tous les cas, le concept est le même : identifier de la parenté par recoupement de la généalogie des autres.

### 9.5 Exporter et importer des données

---

#### 9.5.1 Exporter des données

---

Cliquez sur vos initiales, dans le haut à droite de votre écran d'accès, puis sur [Download Raw Data]. Il faudra ensuite cocher l'énoncé au bas de la page disant que vous reconnaissez les risques en envoyant vos données vers d'autres sites. Les données seront prêtes après un petit délai de préparation.

Il est aussi possible d'exporter la liste des cousins trouvés (*Relatives*) à condition de permettre de les trouver (*matching enabled*), en cliquant sur le lien au bas de l'écran [*DNA Relatives*].

Par contre, on ne peut pas enregistrer facilement la parenté de la parenté qui est affiché par pages de 10 clients. Si vous voulez conserver ces informations pour les comparer entre clients, il est préférable d'utiliser le navigateur Chrome plutôt que Firefox qui ne recopie pas certaines informations.

#### 9.5.2 Importer des données

---

23andme ne permet pas d'importer les données d'un autre laboratoire. Il est toutefois possible de gérer plusieurs troussees 23andme depuis un même compte.

### 9.6 Versions

---

Le tableau suivant montre les différentes versions du test fait par 23andme.

## Le laboratoire FTDNA

Version	Début <sup>62</sup>	Fin	Lignes <sup>63</sup>
1		fin 2007	
2	fin 2007		570 609
3	nov. 2010	nov. 2013	956 363
4	nov. 2013	août 2017	604 913
5	août 2017		629 983

---

<sup>62</sup> Source: <https://h600.org/wiki/Microarray+File+Formats>

<sup>63</sup> Source: [https://isogg.org/wiki/Autosomal\\_SNP\\_comparison\\_chart](https://isogg.org/wiki/Autosomal_SNP_comparison_chart)

## 10 Le laboratoire AncestryDNA

---

Le laboratoire AncestryDNA, rattaché aux nombreux sites régionaux d'Ancestry, a la plus grosse base de clients ayant fait des tests ADN. C'est aussi un site web de généalogie avec de nombreux arbres généalogiques de clients, ce qui est son point fort. Par contre, l'ADNmt trouvé est inutilisable en généalogie et il faut utiliser un site extérieur pour connaître un haplogroupe Y assez ancien.

### 10.1 Introduction

---

AncestryDNA a fait des tests ADN de plusieurs façons. Au début, l'entreprise a acheté les résultats de Sorensen afin d'avoir une bonne base. Par la suite, on a permis aux clients de saisir leur haplogroupe, même si l'information ne venait pas de leur test. Finalement, ayant le vent dans les voiles, AncestryDNA n'héberge plus que des données ADN venant de leurs tests et vérifiant la parenté entre les clients en comparant les arbres généalogiques.

Ce laboratoire a une interface fonctionnant dans plusieurs langues mais ne livre pas en France.

### 10.2 Le test disponible

---

#### 10.2.1 Le test généalogique

---

AncestryDNA offre un seul test ADN. Celui-ci donne une série de résultats touchant surtout les origines ethniques et des communautés plus récentes, ainsi que des données sur les autres clients, comme leur arbre généalogique ou leur lieu de résidence. La fonction Thrulines montre, pour chacun de vos ancêtres documentés, le nombre d'autres clients qui en descendent. Il est ensuite possible de choisir un de ces ancêtres, puis de voir comment sont reliés les autres descendants.

### 10.3 Visite guidée

---

L'auteur a commencé à utiliser AncestryDNA (et Ancestry tout court) seulement en 2022. Certaines fonctions nécessitent un abonnement.

#### 10.3.1 Panneau d'entrée

---

Ancestry est un site de généalogie. Pour voir les résultats des tests ADN, il faut cliquer sur le lien ADN dans le haut de l'écran. Cela ouvre un menu avec ces choix:

- **Résumé des résultats de votre ADN** - Ce résumé est un moyen d'accès aux autres résultats

- **Histoire de mon ADN** - Il s'agit de l'estimation des **origines ethniques** de votre ADN. L'information est présentée comme de la ressemblance avec l'ADN de régions et non comme la composition de vos ancêtres. Dans le cas de l'auteur, cela donne 67% en France (même si AncestryDNA n'a jamais livré officiellement de trousse ADN en France), 28% dans les îles britanniques, 4% en Allemagne et 1% d'ADN amérindien. L'ADN britannique et allemand est sans doute une vision de l'ADN français dans ses régions limitrophes. Cette histoire semble mise à jour à chaque année, en donnant parfois des résultats plus près de la documentation trouvée. Ancestry distingue les apports géographiques paternel et maternel, identifiés comme « Parent1 » et « Parent 2 » si aucun lien n'est fait vers un parent ou un cousin paternel ou maternel.
- **Correspondances ADN** - la liste des correspondances peut être consultée au moyen de plusieurs filtres. Le premier filtre, exclusif, montre l'ensemble des cousins, ceux des côtés paternel et maternel (expérimental en 2023), par ancêtre (voir Thrulines plus bas) et par lieu à l'aide d'une carte. Le second filtre, sélectif, affiche les cousins dont on n'a pas encore consulté le profil, ceux ayant des ancêtres communs, ceux ayant fait l'objet d'une note, ceux ayant ou pas un arbre lié (à l'ADN) ou pas, en fonction de l'ADN partagé (en cM), et en utilisant des groupes personnalisables.
- **Thrulines** - cette fonction sera expliquée plus bas.
- **Traits** - pour un supplément, 40 caractéristiques sont analysées.
- **Questionnaires ADN** - ces questionnaires semblent destinés à recueillir des données pour leur utilisation par le volet «traits».
- **Activer un test** - c'est l'activation d'un nouveau test, éventuellement pour un autre membre de votre famille et à l'intérieur d'un compte commun.
- **Acheter un autre test** - le but est de vendre un test ADN à un autre membre de votre famille.

### 10.3.2 Thrulines

---

Cette fonction permet de connaître le nombre de clients de Ancestry descendant de chacun de vos ancêtres, ont fait leur test ADN et ont publié leur généalogie sur ce site.

La fonction est activée une fois que votre arbre, relié à votre test ADN, contient suffisamment d'ancêtres. Par la suite, de temps en temps (apparemment aux 2 ou 3 jours), Ancestry va comparer votre arbre à celui de vos matches ayant un arbre et indiqueront

## Le laboratoire AncestryDNA

pour chacun le nombre d'autres clients qui descendent de cet ancêtre. Il y a aussi un schéma de la relation en question. Par exemple:

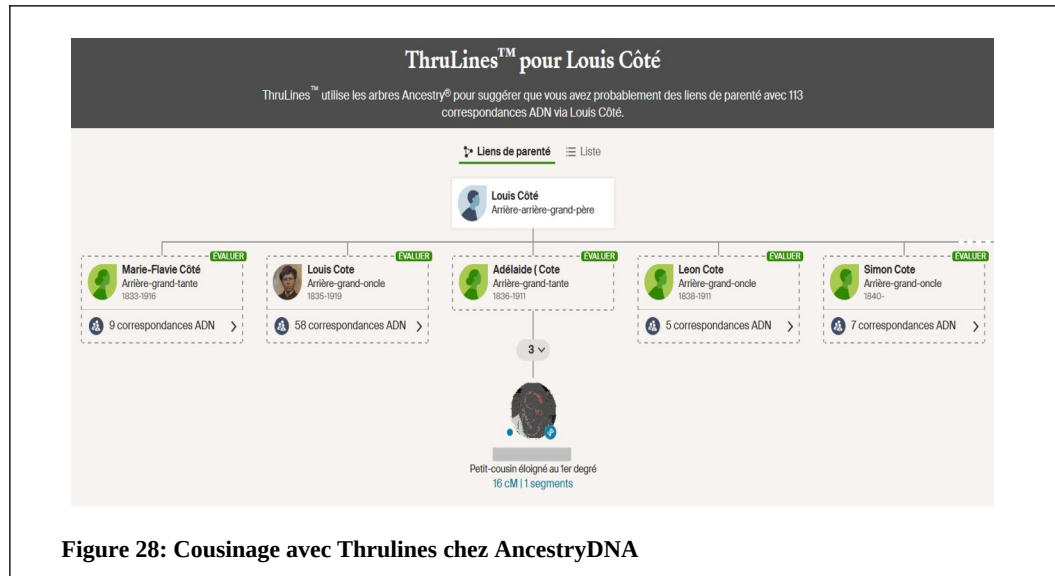


Figure 28: Cousinage avec ThruLines chez AncestryDNA

Parmi les nombreux enfants de cet ancêtre, 9 descendent de sa fille Marie Flavie et ont fait un test ADN sur ce site. En cliquant sur son nom, 3 descendants apparaissent pour continuer à explorer sa descendance. Par la suite, on atteint les personnes testées et le nombre de cM partagés.

Il va de soi qu'il est possible qu'une des personnes (ou plusieurs) peut être adoptée et descendre de cet ancêtre par un ancêtre différent. L'idée est alors que s'il y a beaucoup de descendants, il est peu probable qu'aucun ne descende de cet ancêtre commun.

La correspondance se fait à partir des arbres disponibles sur ce site. Si le nom est écrit différemment, par exemple un ancêtre est appelé parfois Joseph, parfois Joseph Louis et parfois Joseph Toussaint, le système ThruLines ne le verra pas dans tous les arbres disponibles. Ainsi, une de mes ancêtres a 98 descendants et son unique conjoint en a un seul, ma nièce qui a écrit son nom de la même façon que moi. Il y a donc des anomalies possibles. Dans mon cas, une nièce ayant fait un test au même endroit, ainsi qu'un cousin maternel, tous mes ancêtres avoir au moins un descendant, et tous les maternels au moins deux, mais certains n'en ont aucun, parfois pour une graphie différente, parfois parce que le cousin n'a pas mis de nom dans son arbre.

### 10.4 Utilisation pour la généalogie

Le test de AncestryDNA peut-il aider à valider sa généalogie et à savoir si notre génétique correspond à notre généalogie documentée ?

#### 10.4.1 Les cousins trouvés

---

La liste des correspondances est limitée aux autres clients ayant au moins 6<sup>64</sup> cM en commun. Pour l'auteur, les résultats sont les suivants en 2023 :

ADN → Correspondances → Toutes les correspondances → ADN partagé

- Toutes les correspondances (69 109)
- Correspondances proches (6 662)
- Correspondances éloignées (62 447)

ADN → Correspondances → Par parent

- Côté maternel - 23 942 correspondances (il n'est pas nécessaire de tester un des parents, le côté étant identifié à partir de l'arbre généalogique)
- Côté paternel - 35 048 correspondances
- Les deux côtés - 67 correspondances
- Aucun côté assigné - 10 052 correspondances

Ce site ne permet pas d'enregistrer la liste des correspondances. Si vous voulez le faire, le navigateur Chrome permet la saisie des détails (mais pas Firefox) mais il faut descendre le bas de la page à de nombreuses reprises pour avoir une liste plus complète. Il est suggéré de limiter les résultats à une plage en cM car l'affichage est très lent. L'auteur a pu ainsi extraire une liste de 23 000 cousins avec, par exemple, la relation suggérée, le nombre de cM, la présence d'un arbre relié ou non.

#### 10.4.2 Les haplogroupes Y

---

L'haplogroupe Y des hommes peut être déterminé à partir d'un site externe, comme la routine de Morley. Ce test va examiner les SNP Y du test pour proposer un SNP trouvé (affiché en vert) et des SNP possibles (ceux qui ne sont pas négatifs). Le SNP est souvent vieux de 5000 ans et partagé avec de nombreuses familles. Ce n'est pas suffisant pour confirmer un pionnier mais cela discrimine les pionniers avec un haplogroupe trop différent.

#### 10.4.3 Les haplogroupes MT

---

Les SNP MT recherchés par AncestryDNA ne permettent pas de trouver un haplogroupe MT.

#### 10.4.4 Les autres lignées

---

En utilisant Thrulines, on peut avoir une certitude sur l'identité de certains ancêtres qui ont un grand nombre de descendants, mais après plusieurs générations, ce n'est

<sup>64</sup> Si on saisit de 6 à 7 cM, aucun client n'est trouvé. Avec les autres plages, le nombre de cousins n'apparaît pas.

qu'une comparaison d'arbres généalogiques et l'ADN partagé peut venir par d'autres chemins, notamment à cause de l'effet-fondateur.

#### 10.4.5 Débloquer la lignée d'un adopté

---

Peut-on utiliser les données d'un test AncestryDNA pour débloquer une personne ?

Comme pour le test de 23andme, la méthode consiste à construire les arbres généalogiques des cousins trouvés, puis à trouver les ancêtres qui apparaissent dans plusieurs arbres. Comme Ancestry est d'abord un site de généalogie avec les arbres de ses clients, la recherche des ancêtres communs est plus facile.

Dans tous les cas, le concept est le même : identifier de la parenté par recoupement de la généalogie des autres.

### 10.5 Exporter et importer des données

---

#### 10.5.1 Exporter des données

---

Cliquez sur le lien ADN dans le haut de votre profil, puis sur [Paramètres]. En bas de la page, on clique ensuite sur le lien [**Télécharger les données ADN**]. Le système demandera ensuite de confirmer votre intention.

Il n'est pas possible d'exporter directement la liste des cousins trouvés. Cela peut toutefois se faire pour une partie des correspondants trouvés. À partir de la page [Correspondances ADN], on peut choisir une plage avec le filtre [ADN partagé], aller au bas de la page à plusieurs reprises, le temps que les données apparaissent, puis faire une copie de la page (avec ctrl-A (sélectionner la page) et ctrl-C (copier), suivi de ctrl-V (coller) dans un éditeur de texte). Il est préférable d'utiliser le navigateur Chrome plutôt que Firefox qui ne recopie pas certaines informations. L'auteur a recopié vers Seamonkey qui est un éditeur de pages HTML avec ces plages et résultats :

Plages	Correspondances trouvées
100-3000	43
40-100	288
30-39	972
25-29	1506
22-24	1970
20-21	2204
19	1386

## Le laboratoire AncestryDNA

18	1656
17	2100
16	2511
15	3174
14	3864
8-13	1335

### 10.5.2 Importer des données

---

AncestryDNA ne permet pas d'importer les données d'un autre laboratoire mais dans une version précédente, certaines données ADN pouvaient être ajoutées. Cela pourrait changer car au début de 2024, les règlements internes ont changé en fonction de tests importés.

### 10.6 Versions

---

Le tableau suivant montre les différentes versions du test fait par AncestryDNA.

Version	Début <sup>65</sup>	Fin	Lignes <sup>66</sup>
1	Janvier 2012	Mai 2016	700 153
2	Mai 2016		661 800
2a	Mai 2016	Janvier 2018	668 942
2b	Janvier 2018	Mai 2018	650 410
2c	Mai 2018	Février 2019	664 429
2d	Février 2019		677 864

---

<sup>65</sup> Source: <https://h600.org/wiki/Microarray+File+Formats>

<sup>66</sup> Source: [https://isogg.org/wiki/Autosomal\\_SNP\\_comparison\\_chart](https://isogg.org/wiki/Autosomal_SNP_comparison_chart)

## 11 Les autres laboratoires

---

Nous allons aborder de façon légère les autres laboratoires utilisables pour de la généalogie génétique.

### 11.1 myHeritage

---

Le site web de généalogie myHeritage utilise les services de FTDNA pour ses tests ADN. Il a livré en France de 2018 à la fin de 2022 et est le préféré dans ce pays (et en Europe). En Amérique du Nord, c'est plutôt Ancestry qui a cette position dominante.

Le test de myHeritage est similaire au Family Finder de FTDNA auquel il ajoute l'ADN Y. Il permet aussi l'importation de données depuis d'autres laboratoires ADN.

L'importation des données depuis un autre site donne accès à quelques informations et un paiement supplémentaire ajoute d'autres fonctionnalités.

Les fonctionnalités de base comprennent, après une importation :

- La liste des correspondances avec l'ADN partagé (en% et en cM), le nombre de segments et le segment le plus long;
- Une boîte de messagerie interne (le courriel demeure secret);
- Le nombre d'individus dans l'arbre généalogique de l'autre client, sans y avoir accès;
- La possibilité d'importer plusieurs fichiers ADN dans le même compte.

En débloquant (y compris par l'offre gratuite quand elle est disponible), on ajoute :

- Un accès à l'arbre généalogique de l'autre client;
- Les origines ethniques;
- D'autres outils comme l'auto-groupe (regroupant les correspondants qui se trouvent l'un l'autre) et le comparateur de chromosomes (situe l'ADN partagé dans les chromosomes).

L'auteur n'a pas acheté le test ADN de myHeritage. Il semble que certaines fonctions soient disponibles si on importe un test ADN et qu'on le débloque, mais qu'il faudrait un abonnement pour avoir toutes les fonctions. Cela donne l'impression qu'il est préférable de tester ailleurs, d'importer et de débloquer les fonctionnalités, ou encore de tester avec

myHeritage, d'importer vers FTDNA, puis d'exporter vers myHeritage (comme si on avait un frère jumeau identique) lorsque l'offre gratuite des outils est annoncée.

### 11.1.1 Exporter et importer des données

---

#### 11.1.1.1 *Exporter des données*

---

myHeritage permet l'exportation des données ADN.

#### 11.1.1.2 *Importer des données*

---

myHeritage permet l'importation des données ADN et de gérer plusieurs troussees (et les données de plusieurs personnes) dans un même dossier. Pour cela, il faut d'abord exporter les données depuis d'autres laboratoires, puis suivre les instructions pour les importer.

Lorsque myHeritage annonce le déblocage gratuit lors de l'importation des données (ce qui n'arrive pas souvent), il est possible d'enlever les données déjà téléchargées depuis un autre laboratoire, puis d'importer de nouveau ces données. Il semble qu'il soit aussi possible de télécharger de myHeritage vers FTDNA, puis de FTDNA vers myHeritage (en prenant soin de ne pas effacer le test initial), comme si vous aviez un jumeau identique.

### 11.2 LivingDNA

---

Ce laboratoire britannique a fait beaucoup de publicité sur Facebook autour de l'année 2020. Son test est similaire à celui de 23andme mais concerne plutôt la clientèle britannique.

### 11.3 Yseq

---

Ce petit laboratoire allemand peut produire des tests ADN à la pièce à un prix moindre par rapport à FTDNA. Il offre par exemple des SNPs Y à la pièce ou par panneau, des tests ADNmt et un test du génome complet ou par panneau, ainsi qu'un génome complet et l'identification des SNP s'ils sont déjà connus chez Yfull. L'exportation vers Yfull permet de placer les résultats dans un arbre global. Sa base de données de comparaison est toutefois assez petite en comparaison avec celle de FTDNA.

### 11.4 Les génomes complets

---

Cela a coûté des milliards et une décade au gouvernement américain pour décoder le génome humain. Avec les tests maintenant disponibles, le décodage d'un génome personnel peut se faire en une journée.

Le but d'un génome complet est en général médical. Ainsi, Nebula demande un abonnement à son info-lettre avec ses tests. Dante est un autre laboratoire de ce groupe

avec Yseq et FGC. Ce dernier offre aussi le génome complet des chromosomes Y et MT. Dans ce cas, grâce à une centrifugeuse, il est possible d'isoler les chromosomes analysés et le résultat est essentiellement le chromosome Y, bien que le fichier soit assez gros.

Les laboratoires ont quelques formats principaux pour livrer leurs résultats. Le fichier BAM (souvent, de plus de 50 Go) contient le résultat complet, sous forme de chaînes d'environ 100 valeurs (le génome complet totalise 3 milliards de valeurs en deux copies). Le fichier VCF est une version abrégée avec plutôt la différence entre le résultat et le génome de référence. Le fichier FASTA sert à l'ADNmt qu'il décrit comme une série de lettres parmi A, C, G et T, les quatre bases de l'ADN.

Les tests de première génération utilisent une méthode similaire aux tests SNP, soit la recherche de certaines séquences. Ceux de deuxième génération (NGS) lisent des segments de 100 à 400 valeurs contiguës. La troisième génération est en train d'arriver et permettrait des segments d'un million de valeurs. D'ailleurs, en 2023, le consortium T2T a permis de compléter l'étude du chromosome Y dont la longueur est passée de 59 à 62 millions de bases. Le Y est difficile à analyser car il comprend plusieurs longues séquences répétitives.

En 2023, aucun laboratoire n'offre la transformation d'un génome complet en fichier autosomal. Toutefois, certains sites peuvent offrir des outils destinés à cette fin. Voir par exemple WGSEextract.

### 11.5 Le site YFULL

---

YFULL ne fait pas d'analyse ADN mais intègre les résultats des génomes complets ou des chromosomes Y ou MT complets dans son arbre phylogénétique. Il reçoit donc les fichiers BAM pour en extraire les SNPs Y et les incorporer à son arbre YTREE. De même, il analyse des fichiers FASTA de l'ADNmt pour son arbre MTREE.

Ce site ne travaille qu'avec des génomes complets. Ainsi, les données peuvent provenir du Big Y de FTDNA et de la plupart des laboratoires livrant un génome complet, mais pas des tests populaires comme ceux de AncestryDNA ou 23andme, ou les tests STR de FTDNA.

Suite à l'invasion de l'Ukraine par la Russie, ce laboratoire s'est déplacé de la Russie vers l'Espagne.

#### 11.5.1 Importer des données

---

Yfull permet d'importer les fichiers VCF et BAM utilisés pour des génomes complets du chromosome Y (comme leBig Y de FTDNA, les Y Elite de FGC, et les tests de Nebula, Dante, Yseq, etc.), ainsi que les fichiers FASTA. Ce site fera ensuite une analyse

du positionnement relatif des résultats ADN Y dans un arbre couvrant l'ensemble de l'humanité. Un autre arbre est aussi utilisé pour le chromosome MT.

## 11.6 Gedmatch

---

Gedmatch n'est pas non plus un laboratoire. Il reçoit les données autosomales de plusieurs laboratoires et offrent plusieurs outils. Il peut faire une comparaison ciblée («*One-to-one*») ou encore par rapport aux autres données déjà reçues («*One-to-many*»). Un des résultats intéressants est la présence des haplogroupes si le test est importé depuis 23andme. Dans l'importation vers Gedmatch, les haplogroupes sont saisis par le client. Ces haplogroupes sont visibles avec certains outils.

Un autre outil, dans le bloc payant [**Tier 1**] s'appelle Lazarus. En combinant l'ADN de deux ou trois enfants (ou plus), et d'un parent du côté paternel ou maternel (oncle, tante, cousin, par exemple), il réussit à isoler jusqu'à 80% de l'ADN du défunt. Cela est parfois suffisant pour aider à débloquer une génération avec un parent inconnu.

Gedmatch offre aussi plusieurs outils analysant les origines ethniques appelées Admixture.

### 11.6.1 Importer des données

---

Gedmatch permet d'importer les résultats de tests autosomaux de la plupart des laboratoires. Il est aussi permis d'ajouter à la main les haplogroupes trouvés. La version gratuite est limitée à 5 troussees.

## 11.7 Geneanet

---

Le site de généalogie Geneanet, situé en France, a ajouté une fonction génétique à ses arbres en ligne. Ces opérations ont toutefois cessé à la fin de 2023.

## 11.8 Comparaison

---

L'auteur a fait plusieurs tests ADN autosomaux et les a importés sur quelques sites spécialisés. Le tableau suivant compare le nombre de cousins trouvés selon le laboratoire ou le site d'importation. Le tableau montre le nombre total de correspondances, celles d'au moins 100 cM (en théorie, au Québec, assez pour identifier les ancêtres communs) et d'au moins 50 cM. Ce sont des résultats personnels et non des moyennes à partir de plusieurs troussees.

Test	FF (FTDNA)	23andme <sup>67</sup>	AncestryDNA
Original (Total)	5 294	1 502	69 353

---

<sup>67</sup> 23andme limite les correspondances à 1500 par défaut, et 5000 avec le test médical. Ce chiffre est parfois un peu plus élevé, comme 1502 au lieu de 1500. 100 cM devient 1,34%.

## Les autres laboratoires

Importé (FTDNA)	5 294	6 177	5 128
Gedmatch <sup>68</sup>	3 000	3 000	3 000
myHeritage	11 708	8 323	11 660
Geneanet	419		
Original (100 cM et plus) <sup>69</sup>	3	15	49
FTDNA	3	3	3
Gedmatch	3	3	3
myHeritage	4	4	
Geneanet	0		
Original (50 cM et plus)	16	287	140
FTDNA	16	16	13
Gedmatch	18	29	16
myHeritage	75	75	
Geneanet	0		

Voici les valeurs minimales requises pour avoir un match. Les laboratoires ont d'autres critères pour décider s'il y a une correspondance, comme le plus long segment partagé ou la longueur des segments utilisés pour le calcul. À titre d'exemple, FTDNA a changé son algorithme en janvier 2021. Les 10 plus grosses correspondances de l'auteur dans la liste de 2017 comprenaient des valeurs entre 65 et 81 cM. Avec le nouvel algorithme de 2021, les valeurs pour ces mêmes personnes variaient de 16 à 43 cM. La différence vient du fait qu'on ne tient plus compte des segments les plus petits dans le calcul du total. Donc, à partir des mêmes données, selon l'algorithme, l'ADN partagé sera différent.

Laboratoire	cM minimal
23andme	8,9 cM (0,12%) si on n'atteint pas la limite de 1500 correspondances

<sup>68</sup> La version gratuite de Gedmatch est limitée à 3000 résultats. La version payante (Tier 1) prévoit jusqu'à 100 000.

<sup>69</sup> Les trois cousins les plus proches avec le test Family Finder et Gedmatch sont les mêmes. myHeritage en ajoute un quatrième. Les quatre cousins font aussi parties de ceux trouvés avec Ancestry. Un seul a testé avec 23andme.

## Les autres laboratoires

AncestryDNA	8 cM selon leur outil de recherche ou 6 cM selon les indications du site
Family Finder (FTDNA)	7 cM (15 cM avant 2021)
GedMatch	Faible (avec 100 000 correspondances)
Geneanet	7 cM
myHeritage	8 cM

### 11.9 Pour en savoir plus

Les laboratoires ADN ont de la documentation sur leurs propres services afin d'aider le client à choisir un produit. Le site de l'ISOGG<sup>70</sup> donne une longue liste de laboratoires et d'autres informations pour le lecteur qui voudrait tout savoir sur le sujet. Les seuls sites ayant une interface ou de la documentation en français sont Ancestry, myHeritage et Geneanet.

---

<sup>70</sup> [https://isogg.org/wiki/Choosing\\_a\\_DNA\\_testing\\_company](https://isogg.org/wiki/Choosing_a_DNA_testing_company)

## 12 Vocabulaire

---

Voici quelques termes utilisés en généalogie génétique.

### 12.1 Termes francophones

---

**23andme** - laboratoire d'analyse génétique spécialisé surtout dans l'ADN médical. Voir le chapitre 9 qui lui est dédié.

**ACPR** - Ancêtre commun le plus rapproché. Dans les triangulations Y et MT, le premier ancêtre de toutes les personnes testées selon la documentation et ayant des résultats cohérents

**ADN** - nom générique pour désigner les données génétiques humaines extraites dans le cas d'un test d'analyse en vue d'une utilisation en généalogie. L'ADN est formé de 4 bases appelées A, C, G et T.

**ADNau** ou autosomal ou autosomique - ADN formé des 22 paires non-sexuelles de l'ADN humain, utilisé pour évaluer la relation entre deux personnes

**ADNmt** - ADN des mitochondries dont la longueur est de 16569 valeurs plus des insertions moins des retraits.

**ADN X** - chromosome sexuel humain. L'homme possède des chromosomes X et Y. Son X est reçu de sa mère. La femme possède deux chromosomes X. L'un est celui de son père sans recombinaison alors que l'autre est une combinaison des deux chromosomes de sa mère

**ADN Y** - chromosome sexuel de l'homme, reçu de son père et transmis à ses fils. Ce chromosome a une longueur d'environ 63 millions de bases.

**Adoption** - explication diplomatique pour expliquer une brisure génétique. Il peut aussi s'agir d'une infidélité, d'un viol, de l'enfant d'une fille en donnant le nom de sa grand-mère au lieu de celui de sa mère, etc.

**AncestryDNA** - laboratoire associé au site de généalogie Ancestry. Voir le chapitre 10 qui lui est dédié.

**Arbre phylogénétique** - représentation graphique des données génétiques en fonction de la chronologie

**BAM** - Format utilisé pour exporter le résultat d'analyse de génomes complets (WGS). Ce format est assez complexe et comprend entre autres des séquences d'environ 100 à 400 valeurs de suite dans un chromosome (sur 6 milliards).

**Chromosome** - molécule utilisée pour la fabrication d'un humain ou de tout autre être vivant sur terre. L'humain possède 23 paires de chromosomes nucléaires (trouvés dans le noyau des cellules) ainsi que celui des mitochondries (ADNmt).

**cM** - centiMorgan, unité de comparaison de l'ADN partagé. Voir le chapitre 7

**Dante** - laboratoire produisant une analyse du génome humain complet

**DG** - distance génétique. Il s'agit du nombre de différences entre les données brutes ou haplotypes de deux tests ADN Y ou MT. Avec l'ADN Y, on compare les marqueurs STR. Avec l'ADNmt, on examine les SNP différents de la référence. Cette distance est indépendante du nombre de générations séparant deux personnes. Ainsi, un père et un fils

## Vocabulaire

pourraient avoir une distance génétique de 3 (mais c'est très rare) et deux cousins éloignés avec un ancêtre commun à 10 générations pourraient avoir une distance de zéro.

**ÉNP** - événement non parental ou brisure génétique, lorsque le père biologique n'est pas le père légal

**FGC** - laboratoire produisant une analyse du chromosome Y complet, maintenant fermé

**French Heritage** - voir Héritage français

**FTDNA** (FamilyTreeDNA) - premier laboratoire d'analyse génétique à offrir des tests au grand public. Voir le chapitre 8 qui lui est dédié.

**Génome complet (WGS)** - extraction de l'ensemble de l'ADN humain. Un génome humain comprend environ 6,2 milliards d'informations réparties sur 23 paires de chromosomes en plus de l'ADNmt.

**Haplogroupe MT** - résumé d'une signature ADNmt en vue de son regroupement. Il suit un format comme : A1b2c3 où A représente une des grandes familles humaines et 1b2c3 des sous-divisions consécutives. Un tel haplogroupe représente un regroupement pour fins de classement. Il existe un grand nombre de variations comme H qui n'est pas sous-divisé, H1-C16239T qui est une division de H1 sans le SNP C16239, alors que H1-T152C! est une division de H1 avec en plus le SNP T152C.

**Haplogroupe Y** - résumé d'une signature ADN Y en vue de son regroupement. Il suit un format comme : R1b-A431 où R1b représente une des grandes familles humaines et A431 une mutation de référence. Un tel haplogroupe représente une étape sur une échelle chronologique de plus de 100 000 ans.

**Haplotype** - les détails d'une signature ADN, comme les marqueurs STR de l'ADN Y ou les SNP de l'ADNmt

**Héritage français (French Heritage)** - principal projet regroupant des données génétiques chez FTDNA et couvrant la francophonie, le Québec et l'Acadie

**ISOGG.org** - Organisme centralisant des informations sur les tests ADN

**Mitochondries** - organelles transmises par la mère à ses enfants et dont l'analyse permet de confirmer une lignée maternelle

**myHeritage** - site de généalogie offrant des tests ADN faits par le laboratoire FTDNA. Voir la section 11.1 qui lui est dédié.

**Nebula** - laboratoire produisant une analyse du génome humain complet

**Police** - l'utilisation de votre test ADN autosomal par la police se fait si vous l'autorisez en fonction de vos paramètres personnels

**Projet** - regroupement de données chez les laboratoires FTDNA et YFULL et éventuellement d'autres. Un projet implique en général une participation volontaire par les clients qui donnent à un administrateur de projet un accès à leurs résultats et à certaines informations qui sont souvent confidentielles ou réservées à ceux qui partagent de l'ADN

**rCRS** - génome de référence des mitochondries appelé l'Ève européenne et basé sur le chromosome des mitochondries d'une femme faite à l'université Cambridge. CRS est la version initiale (Cambridge Reference Sequence) et rCRS la version corrigée, une position ayant été omis

## Vocabulaire

**Recombinaison** - lors de la reproduction, le fait que deux chromosomes d'une même paire soient recombinaison ensemble et non transmis tels quels à la génération suivante

**RSRS** - génome de référence des mitochondries appelé l'Ève africaine et représentant la valeur théorique de l'ADNmt humain le plus ancien

**SNP autosomal** - dans l'ADNau, valeur lue à une position particulière

**SNP MT** - dans l'ADNmt, valeurs différentes par rapport à la référence (RSRS ou rCRS). Le format est C16239T, soit que la position 16239 du génome de référence contient C alors qu'on a lu T chez le client

**SNP Y** - dans l'ADN Y, valeur lue à une position précise et conservée par les générations suivantes

**STR Y** - dans l'ADN Y, longueur d'une séquence lue dans le chromosome et formant un marqueur

**Triangulation (autosomale)** - survient lorsque deux personnes partagent le même segment d'ADN paternel ou maternel qu'un probant et se confirme quand ces deux personnes partagent entre elles ce même segment

**Triangulation (Y et MT)** - survient lorsque deux lignées paternelles (ADN Y) ou maternelles (ADNmt) documentées mènent à des personnes testées ayant des résultats cohérents

**YFULL** - site d'analyse de génomes complets ayant son propre arbre phylogénétique

**YSEQ** - laboratoire allemand d'analyse d'ADN Y et MT

### 12.2 Termes anglophones

---

**DNA** - voir ADN

**GD** - *Genetic Distance*, voir DG

**MRCA** - *Most Recent Common Ancestor*, voir ACPR

**NPE** - *Not the parents expected* ou *Non parental event*, voir ÉNP

Section vide

