# Comparaison de deux arbres d'ascendants.

Jacques Paris

Montréal, mars 2022

L'idée de vouloir comparer deux arbres d'ascendants ne devrait pas surprendre mais avant même de se lancer dans une comparaison, il faut s'assurer que les contenus de ces arbres répondent à certaines conditions explicites. Celles-ci ne sont pas prédéterminées dans les détails, une condition pouvant avoir plusieurs variantes acceptables et leur choix final faisant l'objet de conventions passées entre les « propriétaires » des arbres. Une fois ces conventions établies, la comparaison peut commencer.

Il s'agit avant tout que les sujets des deux arbres appartiennent à la même génération. Les conventions portent ensuite sur l'identification des individus pris en compte et sur la longueur des branches. Un arbre se termine par une multiplicité de branches, les ancêtres occupant ces extrémités, les têtes de lignée, devraient être identifiés proprement (par exemple par un nom et un prénom); que peut-on accepter? La longueur des branches est très variable; accepterait-on un horizon fixe (une génération butoir pour toutes les branches) ou certains principes envisageables dans certaines situations (par exemple au Québec les « fondateurs » de lignée, nouveaux immigrants), ou une combinaison des deux?

La comparaison peut alors commencer selon un plan préétabli et avec des outils acceptés. Car il est de nombreuses « dimensions » qui peuvent faire l'objet d'analyse. Que cela tienne à la forme de l'arbre (ex. distribution des longueurs des branches) ou à la population même (nombre d'ancêtres dans l'arbre, au total et distincts) et de son homogénéité (complétude générationnelle et totale). Que l'on s'intéresse à la structure interne résultant de l'existence d'implexes (globalement avec le CRDG - coefficient de réduction de la diversité généalogique -, en détail par la distribution des familles « implexées » ou des poids ancestraux). On pourrait aussi considérer les aspects géographiques portant sur les déplacements, et les concentrations spatiales, dimensions pour lesquelles nous n'avons pas encore d'outils spécifiques et que nous éviterons pour l'instant.

Le texte est donc organisé autour de deux axes, les conventions et les analyses.

# I - Les conventions

## Les sujets

Les sujets pour lesquels les arbres sont construits (connus aussi comme de cujus) doivent être pratiquement du même âge car si l'on veut faire des comparaisons par génération il convient qu'elles se situent temporellement dans le même environnement. Il est parfois impossible de maintenir cette condition sur toute l'étendue de l'arbre car certaines branches peuvent « pousser plus vite » que d'autres; la condition initiale reste une façon de minimiser ces risques de décalage.

## Identification des personnes

On admet qu'une personne est bien identifiée si son nom et son prénom sont connus. Toute identification manquante ou partielle revient à dire que l'on a affaire à un inconnu, donc cette personne ne peut être prise en compte comme un ancêtre. La quasi-totalité des cas se rencontrent parmi les têtes de lignée, c'est-à-dire les personnes sans parents connus.

Certaines pratiques « nommant » par exemple les parents de X. Untel comme Monsieur Untel et Madame Untel sont à proscrire. Toute tête de lignée pourrait se voir attribuer de tels parents ce qui n'apporterait rien à la connaissance des ancêtres.

L'utilisation de termes généraux comme « ancêtre_A Untel » qui est utile quand on connait l'existence de « frères/sœurs » mais pas celle de leurs parents pourrait être tolérée si au moins deux de ces frères/sœurs font partie du même arbre ; cela permettrait de marquer l'existence d'un implexe.

Il est cependant des cas où une identification vague ou générale pourrait être acceptée. Un cas précis serait en Amérique du Nord un conjoint « autochtone » non identifié mais portant une mention comme « indienne abénaquis » ou « princesse indienne ». Même si l'identification ne respecte pas la règle de base, le maintien dans l'arbre de cette personne fournit une information généalogique fondamentale.

Une tête de lignée pourrait être qualifié de « parent inconnu », c'est-à-dire qu'au moment de la naissance ou du mariage d'un enfant ce parent est déclaré « inconnu » par l'autre. Cette information peut être d'intérêt pour un généalogiste voulant poursuivre des recherches mais n'ajoute pas d'information dans un arbre. Il serait donc à éliminer.

## Ascendance biologique ou légale

La question touche la façon de traiter les cas d'adoption reconnus. Est-ce que les parents (et leurs ascendances) à retenir dans la construction des arbres doivent être les parents biologiques ou les parents adoptifs? Est-ce que l'aspect hérédité chromosomique est à favoriser plutôt que l'hérédité culturelle?

Il est bien entendu que les changements sociaux récents aident les adoptés à retrouver leurs parents naturels. Mais la situation dans le passé n'est pas aussi favorable à ces « retrouvailles », l'absence d'information sur les parents biologiques fermant la porte à toute découverte. Il y a eu aussi certainement des cas d'adoption non documentés.

Il ne semble pas y avoir quelque consensus que ce soit sur le sujet mais il convient que les conventions d'avant comparaison spécifient cet aspect qui peut affecter profondément la structure d'un arbre.

## L'étendue d'un arbre.

Un arbre construit avec les informations disponibles et dans le seul but de compiler le plus d'ancêtres possible peut comporter des branches très longues mais rares. Dans l'état actuel des ressources, on

ne peut pas imaginer, par exemple, d'aller systématiquement au-delà de la fin du 16ême siècle, moment auquel les archives d'état civil d'Europe occidentale ont commencé à s'organiser. Il faut donc se mettre d'accord sur l'horizon à adopter pour les deux arbres.

Un horizon fixe établit le nombre maximum de générations à conserver. Au-delà de cet horizon, toute information est oubliée, en-deçà, toute information est conservée.

Dans certaines circonstances, des conditions particulières permettraient d'utiliser un horizon variable défini par un événement déterminé. Au Québec par exemple, l'événement serait l'arrivée d'un immigrant suivie de la formation d'une famille. Chaque branche dont à la tête serait un « fondateur » peut donc être de longueur variable, dépendant de la date d'arrivée. Mais cette définition est un peu réductrice car elle ne prend pas en compte le mode d'arrivée. Le plus grand nombre de cas vient certes de l'immigration reconnue de France, la mieux documentée, mais il y avait d'autres arrivées comme les habitants de la Nouvelle Angleterre « enlevés » lors d'expéditions et intégrés par la suite dans la société québécoise, comme les acadiens expulsés durant le Grand Dérangement après la conquête anglaise, comme ceux qui ont « exploré » le Golfe (St-Pierre et Miquelon, Terre Neuve, l'Île du Prince Édouard) avant de se fixer au Québec, comme et surtout les autochtones dont plusieurs ont épousé des colons. Chacune de ces sources posent des problèmes en particulier quel serait l'ancêtre que l'on reconnaitrait comme « fondateur » ou au moins formant la tête de lignée « limite ».

## Les « arbres »

Le principe de construction d'un arbre d'ascendants est basé sur la notion symétrique que chaque ancêtre a deux parents et qu'un couple a un enfant. Cette image d'un arbre *binaire* permet de construire un arbre *théorique* avec des cases et des liens. Dans une présentation verticale, le sujet est à la base et les cases d'une même génération sont sur une horizontale; le nombre de cases suit une loi exponentielle, doublant à chaque génération. Si le sujet est la génération 1, le nombre de cases dans la génération G est égal à $2^{G-1}$ (2 à la puissance [G-1], ou 2 multiplié par lui-même [G-1] fois). Pour des fins de repérage, les cases peuvent être numérotées selon, par exemple, le principe avancé par Sosa : 1 pour le sujet, puis 2 et 3 pour les parents, 4 à 7 pour les grands-parents … toujours en continu et de gauche à droite.

Quand on remplit cette structure par les identifiants des parents, on obtient un *arbre entier*, même si certaines cases restent vides.

Dans la réalité, on peut observer que certains ancêtres sont répétés et avec eux leurs propres ancêtres. La raison en est qu'un couple d'ancêtres a eu deux ou plusieurs enfants qui sont des ancêtres du sujet; ce sont des situations d'implexe. Si l'on veut rendre compte de ces situations, la symétrie binaire sur laquelle l'arbre théorique est basé est brisée : un enfant a toujours deux parents mais un couple de parents peut avoir plusieurs enfants. Pour faire face à cette situation et pour conserver l'aspect binaire de l'arbre, certaines conventions ont été adoptées : parmi les enfants d'un tel couple, seul un (par exemple, celui avec le plus petit Sosa) va avoir une ascendance détaillée; les cases des parents des autres sont remplacées par des renvois aux parents conservés, ce sont des *marqueurs d'implexes*. Comme les répétitions sont ainsi éliminées, nous avons là un *arbre élagué*.

La distinction entre entier et élagué est importante car elle sous-tend une différence dans la disponibilité de l'information utile aux comparaisons. Par exemple, l'arbre élagué livre facilement toute l'information relative aux implexes et à leurs conséquences, mais ne donne pas le détail nécessaire pour identifier toutes les têtes de lignée. Il faudra glisser plusieurs fois de l'un à l'autre et parfois dépendre des deux pour en faire une synthèse.

# II - Les analyses

## Conventions adoptées pour comparer les arbres Fra et Reg.

Fra et Reg sont deux personnes nées au Québec avec moins de 5 ans d'écart à la fin des années 1940.

Nous avons adopté un horizon limite avec la 13ème génération; toute information au-delà de cette génération est ignorée. Tenant compte de cette condition globale, chaque branche est bornée par l'arrivée au Québec d'un ancêtre. Cette notion d'arrivée au Québec est elle-même nuancée pour tenir compte de situations particulières comme a/ les ancêtres Acadiens pour lesquels c'est l'arrivée en Acadie qui est prise en compte b/ ceux venus de la Nouvelle Angleterre pour lesquels on peut remonter jusqu'à leur arrivée en Amérique c/ les autochtones sans limite autre que l'horizon de la 13ème génération.

Les personnes présentes dans l'arbre doivent être identifiées par nom et prénom. Toute information insuffisante disqualifie la personne. Seules sont acceptées les personnes dont l'absence d'information est « justifiée » comme un enfant né d'un père déclaré inconnu. N'ayant pas eu à faire face à un cas d'adoption connue, aucune décision n'a été prise à ce sujet.

Tout couple incomplet est éliminé.

Compte tenu de ces restrictions, les têtes de lignée sont toujours des couples. Cette réserve amène certaines entorses à la définition des limites de branches car il y a plusieurs cas de parents venus avec des enfants mais dont le conjoint n'a pas immigré (en général à la suite d'un veuvage); ce parent est alors retenu aussi.

## Arbres entiers.

### Taille

La longueur des deux arbres est la même à cause de l'horizon limite fixé à 13 générations. Remarquons en passant que l'utilisation d'un tel horizon coupe les liaisons avec les origines françaises par exemple alors qu'elles peuvent contenir des informations sur l'existence de familles implexées, couples de parents dont au moins deux enfants sont parmi les ancêtres du sujet. Cette dimension particulière, qui déborde du cadre de cet article, n'est que rarement prise en compte.

Cette même impression se retrouve dans un autre indicateur de taille, celui de la profondeur ancestrale moyenne[1]. Cette mesure donne en fait la longueur moyenne de toutes les lignées de l'arbre et s'exprime en nombre de générations. Fra = 10.42 et Reg = 10.60. Ces deux arbres venant du même milieu font pratiquement le plein des générations au moins jusqu'à la neuvième, presque à la dixième; les différences existant par la suite sont presque gommées par cette valeur initiale très forte. Dans une telle situation, il faudrait utiliser des outils capables de mettre en valeur ces différences.

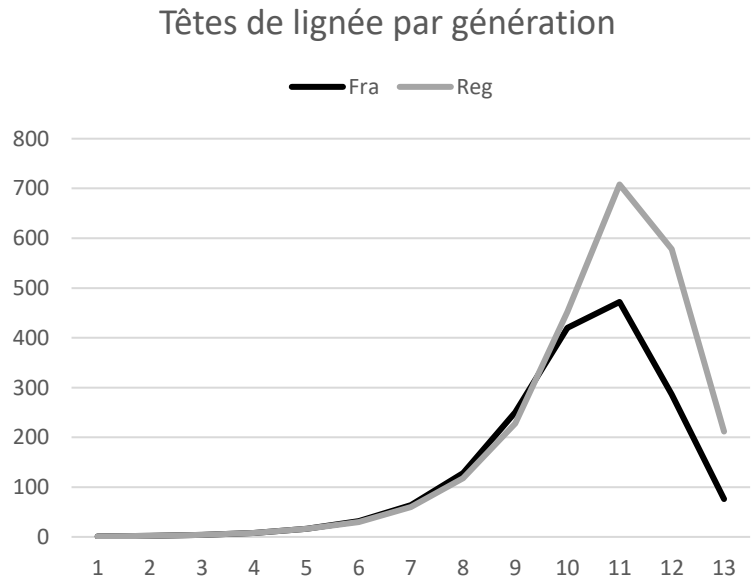
## Forme des arbres

La place occupée par un arbre entier dans son espace théorique donne une idée de son volume global. Le nombre de cases remplies dans les arbres entiers est ici assez différent : Fra = 1759, Reg = 2417 ce qui peut se traduire par des coefficients de complétude globale (rapport au nombre total de cases dans un arbre théorique de 13 génération, soit 8191) de Fra = 0.215 et Reg = 0.295. La différence entre totaux semble bien plus marquée que celle entre coefficients mais ce n'est qu'une question d'échelle puisque les deux valeurs ont été divisées par le même nombre. L'arbre entier de Reg est plus dévloppé que celui de FRA.

Pour révéler la forme des arbres (par ex. élancés vs trapus), nous allons utiliser l'enveloppe de toutes les lignées c'est-à-dire le distribution du nombre de têtes de lignées par génération.

***Têtes de lignée par génération***

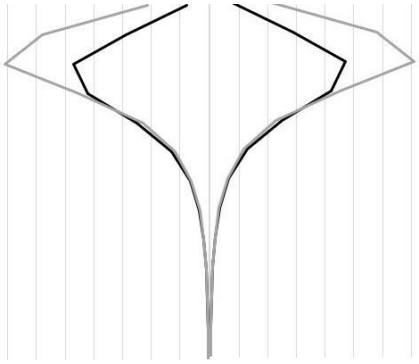
| Génération | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fra | 1 | 2 | 4 | 8 | 16 | 32 | 64 | 128 | 250 | 420 | 472 | 286 | 76 |
| Reg | 1 | 2 | 4 | 8 | 16 | 30 | 60 | 118 | 228 | 452 | 708 | 578 | 212 |




[1] C'est la somme des colonnes dans le tableau plus bas de « complétude générationnelle » qui est l'équivalent de l' « Indice de complétude » utilisé dans « Origines et contributions génétiques des fondatrices et des fondateurs de la population québécoise », Hélène Vézina, Marc Tremblay, Bertrand Desjardins et Louis Houde, Cahiers québécois de démographie Vol. 34, No. 2 automne 2005

On peut voir clairement sur le graphique que les différences notables entre ces deux arbres sont très claires après la génération 10 même si les premières têtes de lignée apparaissent dès la sixième.

Une petite manipulation de ce graphique (rotation plus miroir) donne une image qui pourrait ressembler plus à un arbre. Cette transformation purement visuelle peut être pour certains beaucoup plus « parlante ».



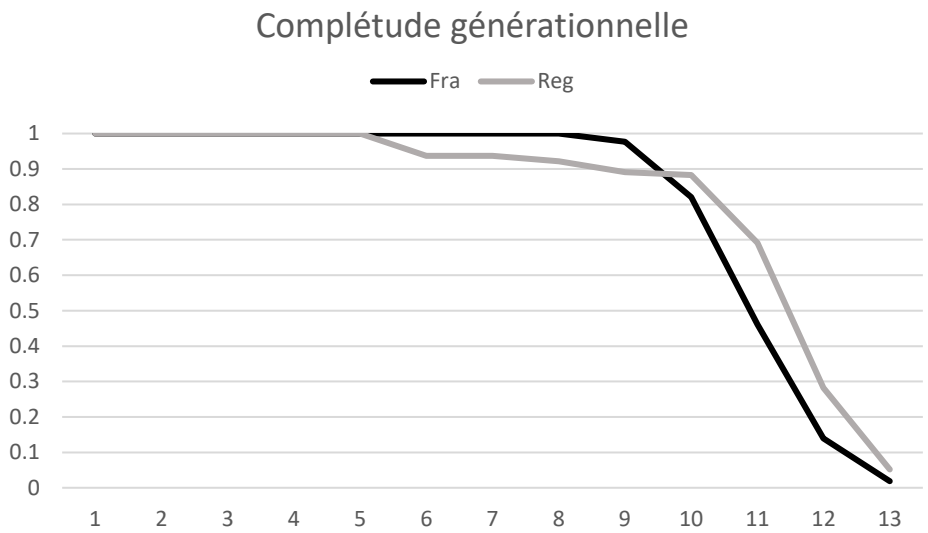
**Complétude et cases vides par génération**

La complétude générationnelle se calcule comme la globale pour chaque génération; c'est le rapport des cases remplies par rapport aux cases théoriques.

***Complétude générationnelle***

| Génération | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fra | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.98 | 0.82 | 0.46 | 0.14 | 0.02 |
| Reg | 1 | 1 | 1 | 1 | 1 | 0.94 | 0.94 | 0.92 | 0.89 | 0.88 | 0.69 | 0.28 | 0.05 |

n.b. la somme des valeurs d'une rangée donne la profondeur ancestrale moyenne



Les têtes de lignée qui sont apparues dès la sixième génération chez Reg grèvent la performance jusqu'à la dixième à partir de laquelle Reg prend le dessus.

Mais ces cases vides n'indiquent pas le nombre de nouvelle têtes de lignée dans une génération car elles contiennent aussi les cases vides correspondant aux têtes apparues précédemment. Si elles ne peuvent pas être obtenues par un décompte direct dans l'arbre, on peut les trouver en tenant compte des apparitions précédentes.

Le petit tableau qui suit montre comment on peut passer de nombres de cases occupées à celui des nouvelles cases vides.

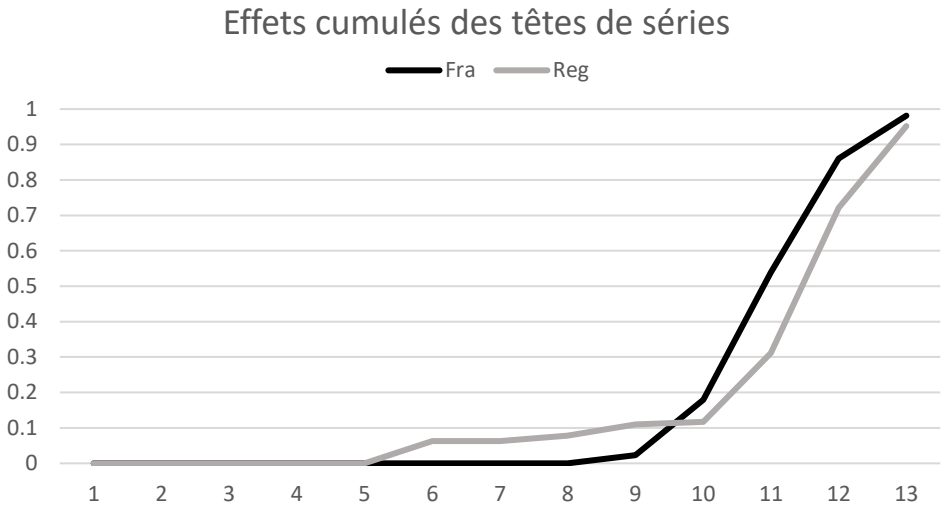
| | | | | | | |
|---|---|---|---|---|---|---|
| Génération | 9 | 10 | 11 | 12 | 13 | |
| Cases théoriques | 256 | 512 | 1024 | 2048 | 4096 | |
| Cases occupées | 250 | 420 | 472 | 286 | 76 | |
| Cases vides | 6 | 92 | 552 | 1762 | 4020 | |
| | | | | | | |
| | 6 | 12 | 24 | 48 | 96 | 6 = première valeur observée |
| | | 80 | 160 | 320 | 640 | 80 = 92 – 12 |
| | | | 368 | 736 | 1472 | 368 = 552 – 24 -160 |
| | | | | 658 | 1316 | 658 = 1762 – 48 – 320 – 736 |
| | | | | | 496 | 496 = 4020 – 96 – 640 – 1472 – 1316 |
| | | | | | | |
| Vides 'nouvelles' | 6 | 80 | 368 | 658 | 496 | |

On peut alors calculer l'impact des nouvelles cases vides par génération (rapport à la population théorique) et en faire le cumul.

***Effets des têtes de lignée***

| Génération | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Théorique | 1 | 2 | 4 | 8 | 16 | 32 | 64 | 128 | 256 | 512 | 1024 | 2048 | 4096 |
| **Fra** | | | | | | | | | | | | | |
| cases vides | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 92 | 552 | 1762 | 4020 |
| nouvelles | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 80 | 368 | 658 | 496 |
| contributions | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.02 | 0.16 | 0.36 | 0.32 | 0.12 |
| cumul | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.02 | 0.18 | 0.54 | 0.86 | 0.98 |
| **Reg** | | | | | | | | | | | | | |
| cases vides | 0 | 0 | 0 | 0 | 0 | 2 | 4 | 10 | 28 | 60 | 316 | 1470 | 3884 |
| nouvelles | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 8 | 4 | 199 | 840 | 944 |
| contributions | 0 | 0 | 0 | 0 | 0 | 0.06 | 0 | 0.02 | 0.03 | 0.01 | 0.19 | 0.41 | 0.23 |
| cumul | 0 | 0 | 0 | 0 | 0 | 0.06 | 0.06 | 0.08 | 0.11 | 0.12 | 0.31 | 0.72 | 0.95 |

Les deux courbes du cumul présentent alors un aspect intéressant. C'est en fait l'image miroir des courbes de complétude générationnelle, la « somme » des deux étant toujours égale à l'unité.




Vu cette complémentarité, une seule est nécessaire puisqu'on peut toujours déduire l'une de l'autre.

## Arbres élagués

### Taille

La longueur des arbres reste constante puisque nous avons fixé un horizon limite.

Cependant, leur « population » est plus petite car toutes les répétitions ont été supprimées : Fra = 1219 et Reg = 1078. Il y a eu inversion de l'ordre des grandeurs. Le nombre des cases occupées de Reg était de 1.37 celle de Fra (Fra = 1759, Reg = 2417), mais un plus grand nombre de cases sont occupées par moins d'ancêtres distincts et leur rapport est tombé à 0.88.

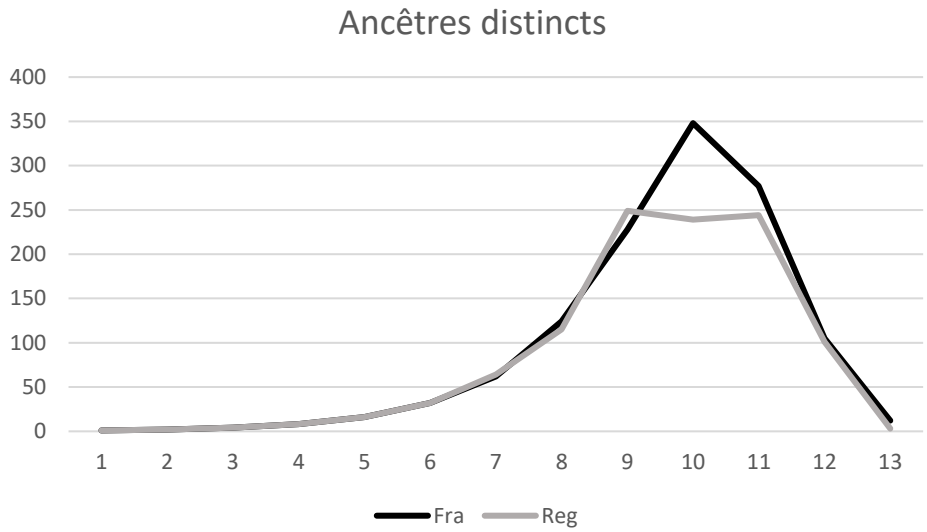
Cette inversion montre bien que l'utilisation d'un seul type de mesure ne rend pas bien compte de la « performance » d'un arbre.

### Forme des arbres

Nous utilisons la même technique de présentation que pour les cases occupées

***Ancêtres distincts par génération***

| Génération | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fra | 1 | 2 | 4 | 8 | 16 | 32 | 62 | 124 | 228 | 348 | 277 | 105 | 12 |
| Reg | 1 | 2 | 4 | 8 | 16 | 32 | 64 | 115 | 249 | 239 | 244 | 101 | 3 |




Si on compare ce graphique avec celui correspondant des cases occupées, on peut voir que les courbes sont affectées surtout à partir de la génération 9, celle de Reg étant notablement amputée dans ces dernières générations.

**Degré de répétitions**

Quand le nombre des cases occupées varient indépendamment du nombre des ancêtres distincts, on peut s'attendre à ce que le degré de répétition (le nombre moyen de fois qu'un ancêtre distinct se retrouve dans l'arbre entier) révèle des comportements différents

| | Cases occupées | Ancêtres distincts | Degré de répétition |
|---|---|---|---|
| Fra | 1759 | 1219 | 1.22 |
| Reg | 2417 | 1078 | 2.24 |

Ainsi il est presque deux fois plus élevé pour Reg que pour Fra.

**Source des changements entre cases occupées et ancêtres distincts**

Le degré de répétition est un indicateur des effets des implexes contenus dans un arbre entier. Plus les implexes sont fréquents, plus ils sont complexes, plus il y aura dans l'arbre entier des répétitions et dans l'arbre élagué des marqueurs d'implexes (ces renvois des parents répétés vers les parents conservés).

Le nombre de marqueurs dans Reg est unbquart de plus que celui dans Fra. Rien que cela indique que leur exclusion de l'arbre élagué aura plus d'impact global.

***Marqueurs d'implexe par génération***

| | 7 | 8 | 9 | 10 | 11 | 12 | 13 | Total |
|---|---|---|---|---|---|---|---|---|
| Fra | 2 | 0 | 14 | 28 | 109 | 47 | 6 | 206 |
| Reg | 1 | 11 | 33 | 99 | 76 | 21 | 12 | 252 |

La distribution par génération pour Fra montre une plus forte concentration dans la onzième (près de la moitié) avec des valeurs « progressives » de chaque côté ce qui ne ferait que rabaisser la pointe sans modifier la forme générale. Ce n'est pas le cas pour Reg où de fortes valeurs existent en dehors de la 11ème ce qui entraine des distorsions de l'arbre entier.

Il existe une mesure globale de l'effet des marqueurs d'implexe qui tient compte de la distribution de ces marqueurs. Il s'agit du « coefficient de réduction de la diversité généalogique »[2] (CDRG) qui est la somme des contributions de chaque génération (rapports du nombre de marqueurs à la population théorique de la génération). Ce coefficient donne la proportion des cases d'un arbre qui seraient des ancêtres occultés par un implexe (les répétitions et leurs ascendances connues ou non).

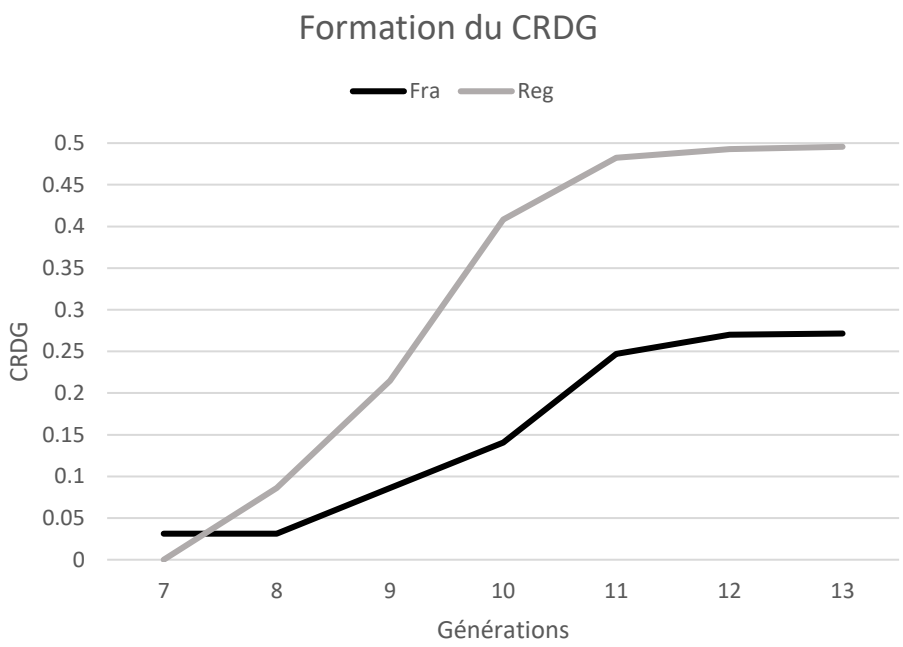
***Calcul du CRDG***

| Génération | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|
| Pop. théorique | 64 | 128 | 256 | 512 | 1024 | 2048 | 4096 |
| **FRA** | | | | | | | |
| Marqueurs d'impl. | 2 | 0 | 14 | 28 | 109 | 47 | 6 |
| Contributions | 0.031 | 0 | 0.055 | 0.055 | 0.106 | 0.023 | 0.001 |
| Cumul | 0.031 | 0.031 | 0.086 | 0.141 | 0.247 | 0.270 | 0.271 |
| **REG** | | | | | | | |
| Marqueurs d'impl. | 0 | 11 | 33 | 99 | 76 | 21 | 12 |
| Contributions | 0 | 0.086 | 0.129 | 0.193 | 0.074 | 0.01 | 0.003 |
| Cumul | 0 | 0.086 | 0.215 | 0.408 | 0.482 | 0.493 | 0.496 |

Le CRDG de Reg est 1.8 fois celui de Fra, ce qui montre que l'impact réel des marqueurs d'implexes est beaucoup plus marqué avec cette mesure que le simple rapport du nombre de marqueurs pour ces deux cas. La raison provient de la distribution des marqueurs par génération et de son résultat; il ne faut pas oublier qu'un marqueur a deux fois plus d'impact si on le rapproche d'une génération du sujet. Nous avons signalé dans le cadre de l'étude comparative des distributions des têtes de lignée et ancêtres distincts que la suppression des répétitions altère plus la courbe ancêtres distincts de Reg car la distribution des marqueurs d'implexe est moins « régulière ». Les CRDG montrent bien que c'est dans les générations 9, 10 et 11 que la différence s'établie.

Rappelons que le CRDG indique à quel point les implexes présents dans un arbre réduisent l'éventail imaginable des ancêtres d'origine. Le CDRG de Reg est extrêmement élevé, un des plus hauts que nous ayons rencontrés, alors que celui de Fra est beaucoup plus faible, parmi les plus bas. Est-ce que cette concentration sur moins d'ancêtres distincts aurait créé un pool génétique plus « critique », ce n'est pas à un généalogiste d'y répondre mais il peut certainement poser la question.

[2] 'Parenté, consanguinité, implexe et diversité généalogique' dans « Macro-Généalogie », Jacques Paris, Éditions Universitaires Européennes, 2021

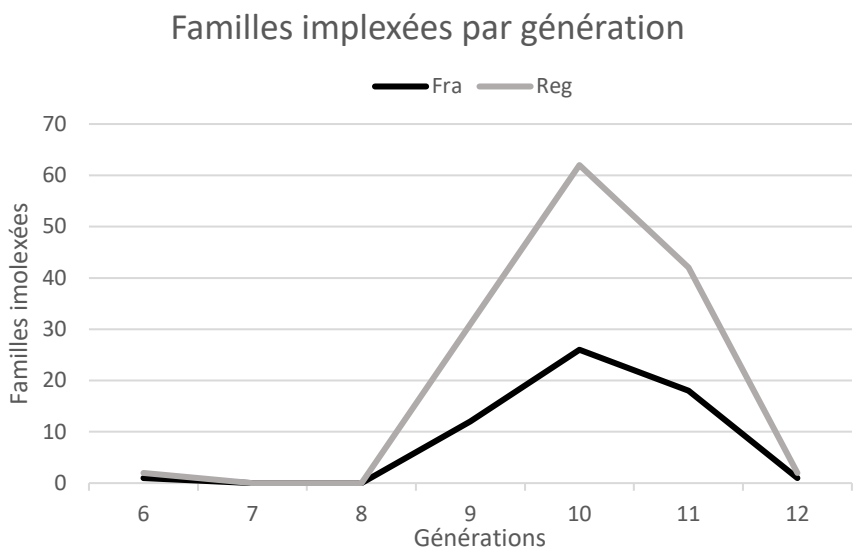


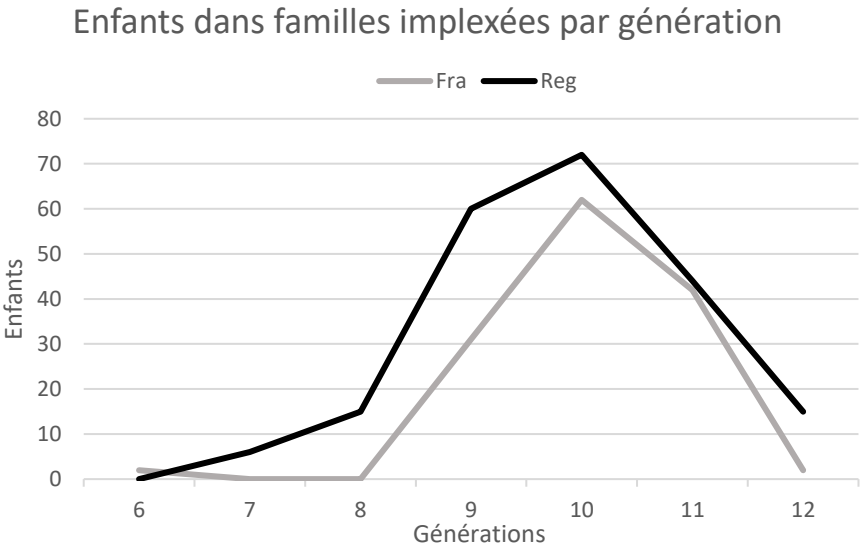

**Familles implexées**

Nous appelons familles implexées celles formées avec deux parents et au moins deux enfants présents dans l'arbre comme ancêtres du sujet. Ces familles pouvant contenir plus de deux enfants, il faut donc tenir compte du nombre de familles et de celui des d'enfants.

| Génération | 6 | 7 | 8 | 9 | 10 | 11 | 12 | Total |
|---|---|---|---|---|---|---|---|---|
| ***Familles implexées*** | | | | | | | | |
| Fra | 1 | 0 | 0 | 12 | 26 | 18 | 1 | 58 |
| Reg | 0 | 3 | 7 | 23 | 25 | 19 | 7 | 84 |
| ***Enfants dans familles implexées*** | | | | | | | | |
| Fra | 2 | 0 | 0 | 31 | 62 | 42 | 2 | 139 |
| Reg | 0 | 6 | 15 | 60 | 72 | 44 | 15 | 212 |

Une mesure globale permet de se donner une image de ces familles : le nombre d'enfants moyen par famille : Fra = 2.40 et Reg = 2.52. Encore ici, la différence sur une valeur d'ensemble ne renseigne pas sur certains détails importants. Par exemple, si les distributions des familles montrent de plus grandes valeurs dans les générations 8 et 9 pour Reg, celle des enfants de Reg est nettement décalée vers le haut sur toute son étendue; c'est une moyenne d'enfants par famille supérieure à la moyenne dans les générations 9 et 10 qui ont tiré vers le haut la courbe à cet endroit. Il est probable que ces écarts à la moyenne révèlent quelque aspect de la construction des familles implexées; il leur faut attendre une certaine ancienneté pour pouvoir regrouper le plus d'implexes possibles.




Nous n'avons pas voulu aborder ici un point qui, s'il peut ajouter une certaine information, est plus difficile à traiter systématiquement et à exploiter; il s'agit des familles implexées complexes[3]. Ces familles sont formées à la suite de remariages d'un des parents d'une famille implexée, remariages dont au moins un enfant est présent dans l'arbre. Un regroupement de plus de trois parents n'est pas à exclure alors, mais en dehors de la description de telles structures, il est difficile d'intégrer ces cas dans une vue d'ensemble.

Ces structures complexes sont la source de différences possibles entre parents de telles familles en ce qui concerne le nombre d'enfants qui peuvent leur être attribué; par exemple un père a deux enfants d'une femme et un d'une autre; on lui attribue 3 descendants alors que la première femme en reçoit deux et la deuxième un. Ces inégalités obligent à devoir traiter les coules comme deux personnes distinctes. C'est ce que l'on tente de faire avec les poids ancestraux.

[3] Sur cet aspect particulier, voir 'Familles implexées' dans « Macro-Généalogie », Éditions Européennes Universitaires, 2022

**Poids ancestraux**

Le poids ancestral d'une personne est le nombre de cheminements distincts qui existent entre le sujet et cette personne[4]. Initialement, chaque personne dans l'arbre a le poids de 1 mais à cause des implexes existant plusieurs cheminements deviennent possibles et les poids s'accumulent. Mise en valeur quand on considère un arbre réduit comme un réseau[5], cette mesure peut s'obtenir aussi d'une recherche de parenté entre sujet et ancêtre offerte dans certains logiciels (nombre de façons dont la parenté existe).

Toutes les personnes faisant partie d'un même arbre ont un poids ancestral mais dans la perspective de cet article nous allons nous limiter aux seules têtes de séries.

| **Fra** | Poids ancestral | | | | | |
|---|---|---|---|---|---|---|
| Génération | 1 | 3 | 5 | 7 | 9 | total |
| 9 | 2 | 1 | | | | 3 |
| 10 | 84 | 2 | 1 | | | 87 |
| 11 | 259 | 7 | 2 | 1 | 1 | 270 |
| 12 | 362 | 9 | 1 | | | 372 |
| 13 | 204 | 2 | | | | 206 |
| total | 911 | 21 | 4 | 1 | 1 | 938 |
| Poids | 911 | 63 | 20 | 7 | 9 | 1010 |

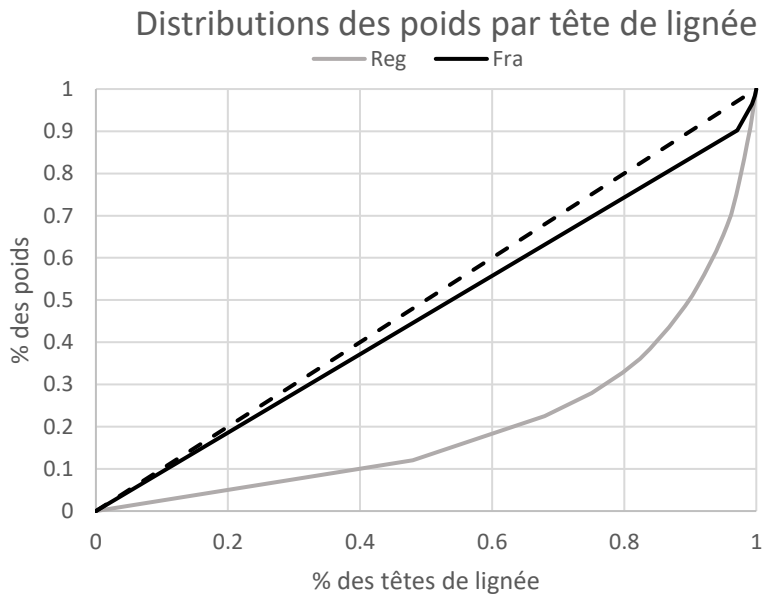
| **Reg** | Poids ancestral | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Génération | 1 | 2 | 3 | 4 à 9 | 10 à 29 | 30 | 32 | 39 | Total |
| 8 | 2 | | | | | | | | 2 |
| 9 | 10 | 2 | | | | | | | 12 |
| 10 | 16 | 6 | | 21 | | | | | 43 |
| 11 | 59 | 43 | 14 | 32 | 21 | | | 2 | 171 |
| 12 | 144 | 58 | 15 | 127 | 16 | 2 | 1 | 4 | 367 |
| 13 | 86 | 21 | 18 | 229 | 15 | 1 | 2 | | 372 |
| Total | 317 | 130 | 47 | 409 | 52 | 3 | 3 | 6 | 967 |
| Poids combinés | 317 | 260 | 141 | 593 | 838 | 90 | 96 | 234 | 2569 |

Réduites aux seules têtes de lignée les arbres de Fra et Reg ne sont pas différents (938 et 967) mais la différence dans les poids (1010 et 2569) peut expliquer celle observée dans les CRDG et celle dans leurs distributions sont énormes (poids maximum 9 et 39).

4 Il faut voir une certaine parenté de cette mesure avec la notion de « contribution génétique d'un fondateur à un groupe de sujets » utilisée dans l'article cité dans la note 1, à la différence près que le poids ancestral n'est valable que par rapport au sujet de l'arbre et ne peut pas être étendue à un groupe de sujets.

5 Voir en particulier le chapitre 'Ascendance comme un réseau' dans « Macro-Généalogie », Jacques Paris, Éditions Universitaires Européennes, 2021. Un gabarit pour calcul sur Excel est inclus dans l'annexe B.

Pour mieux saisir l'impact des valeurs extrêmes, nous allons nous servir de la mesure de distorsion[6] classique dans ces cas; plus la courbe est proche de la diagonale (en pointillé) moins forte est l'influence des valeurs extrêmes. Dans notre cas, il n'est pas besoin de calculer l'indice, la différence est beaucoup trop visible.




**Différences ou similitudes mais surtout convergence**

Nous avons vu comment mettre en évidence des ressemblances et des différences entre deux arbres, mais il existe une dimension vers laquelle deux arbres peuvent converger, celle de la parenté entre les deux sujets. Rappelons d'abord qu'il ne peut y avoir de parenté que s'il y a eu des implexes. Le CREG qui mesure l'impact des implexes sur la diversité des ancêtres, plus spécialement les têtes de lignée, est aussi un indicateur de la force des liens de parenté à l'intérieur d'un arbre. Nous allons utiliser cet outil pour rechercher la parenté pouvant exister entre les deux sujets.

Supposons que chaque arbre est dans un fichier généalogique distinct qui ne contienne que cet arbre. Le nombre de personnes dans le fichier est donc celui de l'arbre. Faisons une fusion de ces arbres, c'est-à-dire créons un nouveau fichier avec un des arbres et ajoutons-y le fichier du second. Durant cette opération, le logiciel de généalogie enregistre ce qui apparait être des doublons, des paires de personnes jugées semblables; nous pouvons accepter que ces doublons soient fusionnés. Il ne restera alors dans le fichier que des ancêtres distincts. Dans notre exemple, si Fra=1219 et Reg=1078, nous nous attendrions à voir que leur fusion donne 2297 personnes mais il n'y en a que 2102. Les 195 personnes manquantes correspondent à des nouveaux implexes créés durant la fusion et dont les répétions ont été évitées; ce sont les effets des doublons repérés et acceptés.

[6] Voir en particulier « Mesure de distorsion » annexe A dans « Macro-Généalogie ». Jacques Paris, Éditions Universitaires Européennes, 2021.

Pour pouvoir utiliser pleinement le CRDG, nous allons « donner » un enfant au couple virtuel formé des deux sujets. Ceci nous permettra de calculer le CRDG de cet « enfant » et de le comparer à ceux des parents. L'arbre unique que nous venons de construire contient tous les ancêtres du côté paternel avec leur structure d'implexes et de même du côté maternel. Le CRDG de l'enfant pourrait donc se déduire de ceux des parents comme la somme des deux CRDG divisé par 2 puisque l'enfant est une génération plus éloigné des ancêtres que les parents. La différence entre ce CRDG et celui obtenu directement avec l'arbre de l'enfant est une mesure de la consanguinité chez l'enfant[7], et le double de cette valeur (pour tenir compte d'un saut d'une génération) mesure la parenté globale existant entre les parents.

Parenté globale entre parents = 2 x (CRDG[enfant] – {CRDG[père] + CRDG[mère]} / 2)

Quel est le CRDG de l'enfant?

***Calcul du CRDG***

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Génération | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| Marqueurs d'impl. | 2 | 11 | 49 | 155 | 189 | 106 | 14 |
| Contributions | 0.016 | 0.043 | 0.096 | 0.151 | 0.092 | 0.026 | 0.002 |
| Cumul | 0.016 | 0.059 | 0.154 | 0.306 | 0.398 | 0.424 | 0.426 |

Remarquons que le premier marqueur d'implexe chez l'enfant se rencontre dans la génération 8, qui est la génération 7 pour les parents.

Parenté globale entre parents = 2 x (0.426 – {0.271 + 0.496} / 2) = 0.084

La parenté globale entre Fra et Reg n'est donc pas particulièrement prononcée mais elle existe bien. En comparant la somme des marqueurs des « parents » avec le nombre chez l'enfant tout en ajustant les générations, nous pouvons faire un bilan de l'impact de la fusion des deux généalogies.

***Impact de la fusion***

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Génération | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| Différences | 0 | 0 | 2 | 28 | 4 | 38 | -4 |

Même si des différences sont bien visibles comme 28 et 38, elles sont cantonnées dans des générations anciennes (10 et 12) ce qui explique que la parenté globale n'est pas plus élevée. D'un autre côté, la valeur négative montre que toutes les fusions peuvent modifier jusqu'à un certain point les nouvelles structures avec des effets inattendus.

Si une recherche de parenté entre Fra et Reg trouve 573 façons dont ces personnes sont apparentées, le lien le plus « proche » n'est pas avant la génération 8 pour Fra, 9 pour Reg. Le grand éloignement de ces liens explique que malgré leurs nombres, leurs contributions à cette parenté reste faible.

[7] La consanguinité est une mesure qui concerne l'individu même. La parenté met en jeu deux personnes; si deux personnes se marient et ont un enfant, celui-ci pourrait recevoir d'elles une certaine consanguinité si elles sont apparentées.

# Conclusions

Si nous avons peut-être donné un poids important en commençant par détailler le genre de conventions à reconnaitre avant d'entreprendre une comparaison entre deux arbres, c'est que nous jugeons important d'éviter toute source de différences qui pourrait influencer les mesures et les interprétations qui pourraient en être tirées.

Nous nous sommes limités à présenter des mesures simples et que nous pensons efficaces. Nous avons montré que la constitution même des arbres ouvraient des portes différentes mais complémentaires : un arbre entier permet en particulier de parler de leur forme et de comparer leur complétude ; un arbre élagué fait valoir entre autres la diversité généalogique des ancêtres du sujet et ouvre la porte à de nouvelles dimensions comme les questions de parenté et de consanguinité.

Chacun des arbres a ainsi dévoilé dans ses différences une histoire généalogique propre mettant en particulier en évidence le rôle fondamental des implexes. Issus pourtant d'un milieu globalement identique, le Québec, ils font supposer que des circonstances régionales différentes pourraient en être la cause; se pourrait-il alors que de nouvelles approches permettraient un jour de le préciser?