

ANALYSE QUANTITATIVE DES RESEAUX GENEALOGIQUES ASCENDANTS

L'EXEMPLE DES LIGNEES FAMILIALES DE LA VALLEE DE LA VALSERINE

(JURA FRANÇAIS)

Guy Brunet (1), Stéphane Lallich (2) et Alain Bideau (3)

(1) Centre d'Etudes Démographiques, Université Lumière – Lyon 2

(2) Equipe de Recherche en Ingénierie des Connaissances, Université Lumière – Lyon 2

(3) Laboratoire de Recherche Historique Rhône-Alpes, Université Lumière – Lyon 2

Adresse de correspondance :

Guy BRUNET
Centre d'Etudes Démographiques
Institut des Sciences de l'Homme
14 avenue Berthelot
69363 Lyon cedex 07
Guy.Brunet@univ-lyon2.fr

La démographie historique a pendant longtemps ignoré la dimension généalogique, et les travaux visionnaires de Louis Henry (1956) et de Thomas Hollingsworth (1964) font figure d'exception. Il a fallu attendre l'arrivée de l'informatique dans les laboratoires de sciences humaines et sociales et la puissance de calcul des micro-ordinateurs pour qu'apparaissent des bases de données informatisées au sein desquelles chaque individu est rattaché à son réseau généalogique. Parmi les travaux pionniers de cette ère informatique on peut citer notamment les initiatives du Programme de Recherche en Démographie Historique de l'Université de Montréal (Desjardins, 1998). En France, un des principaux artisans de cette problématique a été Jacques Dupâquier, avec son étude sur 3000 familles (échantillon TRA) visant à observer la mobilité géographique et la mobilité sociale à l'échelle nationale de 1800 à la fin du XX^e siècle (Dupâquier et Kessler, 1992). Les *Annales de Démographie Historique* se sont à plusieurs reprises fait l'écho des problématiques et des acquis des recherches intégrant une dimension généalogique, notamment dans les volumes de 1984, 1998 et 2000 (voir par exemple Bideau et Poulain, 1984 ou Brunet et Bideau, 2002).

Dans une perspective différente, les généalogies ont souvent été utilisées, notamment en génétique des populations et en anthropologie biologique, pour le calcul des coefficients d'apparentement entre conjoints ou le coefficient de consanguinité d'une population (Boetsch, 1999), mais les exploitations strictement démographiques sont plus rares (Rallu, 1992). Dans des cas exceptionnels, les généalogies ont été utilisées pour percevoir l'histoire globale d'une population (Bouchard et De Braekeleer, 1991). Mais les réseaux généalogiques ont rarement été étudiés en eux-mêmes.

Nous souhaitons présenter ici de manière concrète quelques pistes d'exploitation rendues possibles par l'existence de fichiers démographiques informatisés issus d'une exploitation systématique des registres paroissiaux et des registres d'état civil. Les données utilisées concernent cinq communes de la vallée de la Valserine, dans le Jura français, étudiées de la fin du XVII^e siècle à nos jours (Bideau et Brunet, à paraître).

1 – A la recherche d'une « structure généalogique » de la population

Les biologistes et les généticiens de la population ont découvert avant les historiens – démographes l'intérêt des fichiers généalogiques. De nombreuses recherches portant sur la transmission de caractères héréditaires, sur l'origine de maladies génétiques ou sur la probabilité d'origine des gènes sont fondées sur l'exploitation de telles données. De ce fait, certains concepts de nature démographique ont été formulés par des généticiens. Tel est par exemple le cas du concept des « enfants utiles » : sont ainsi dénommés les enfants qui contribuent à la constitution de la génération suivante en se reproduisant (Heyer et Cazes, 1999 ; Boetsch et Prost, 2001). Par une démarche de généalogie descendante, il est ainsi possible de mesurer quelles lignées familiales se reproduisent sur place, avec quelle intensité au fil des générations, et quelle proportion de lignées disparaissent (Bideau et al., 1995). Chaque ascendant est identifié en fonction du nombre de générations ascendantes qui le séparent de l'individu étudié, sans tenir compte des dates précises de naissance de ces ascendants, des décalages chronologiques pouvant se produire entre les différentes lignées (Cazes, 1987).

Nous proposons ici le concept de « structure généalogique de la population », appellation provisoire qui mérite sans doute d'être discutée. L'interrogation est la suivante : au sein de la population présente dans la vallée à une date donnée, ou au sein de la population née dans la vallée durant une période donnée, peut-on savoir depuis combien de générations *les* ancêtres, ou *des* ancêtres, sont présents dans la vallée ? En d'autres termes, quelle est la proportion de

la population qui relève d'un enracinement déjà profond dans la vallée et quelle est la proportion qui est au contraire le fruit d'une implantation récente ? Peut-on par l'étude des arbres généalogiques ascendants percevoir si la population est homogène ou au contraire composée de groupes distincts quant à cet enracinement local ?

La littérature sur la question n'est pas abondante. D'un côté, on peut se référer au modèle, bien théorique en ce qui concerne les populations européennes, de l'isolat (Gomila, 1976). Dans une telle situation, tous les habitants de l'isolat posséderaient de nombreux ancêtres sur place de nombreuses générations auparavant. La situation semble se rapprocher de ce cas dans la vallée de Vallouise dans les Alpes briançonnaises (Boetsch et al., 2002). A contrario, dans la Dombes caractérisée par une forte mortalité et une forte mobilité géographique, peu d'habitants recensés en 1872 possédaient des ancêtres ayant vécu sur place deux ou trois générations auparavant (Brunet, 1979).

La population étudiée ici est celle cinq communes alignées dans la vallée de la Valserine : Lélex, Chézery, Forens, Champfromier et Montanges. Cette vallée est située entre les plateaux du Jura au nord, les montagnes du Haut-Bugey à l'ouest, les collines du Pays de Gex à l'est et le Bugey au sud. Partiellement isolée par ce relief, la vallée a tout de même été intégrée dans les mouvements migratoires, notamment au XIX^e siècle.

Ensemble les communes étudiées comptaient environ 3000 habitants vers 1700, 4500 vers 1800, 3000 de nouveau vers 1900 et 1400 actuellement. Le dépouillement a permis de compiler plus de 27000 actes de baptême ou naissance, 6100 actes de mariages et environ 21000 actes de décès enregistrés dans la vallée entre 1700 et 2000. Le couplage des informations mentionnées dans ces actes a abouti à la connaissance de 47037 individus différents. Chacun d'entre eux est relié, lorsque l'information est connue, à ses parents, ses enfants et son ou ses conjoints. Parmi ces 47037 individus, 28224 (60.0%) sont nés dans un des villages de la vallée, les autres étant connus uniquement par leur mariage, par leur décès ou par leur présence lors d'un de ces événements. L'ensemble des réseaux généalogiques peut ainsi être reconstruit, dans la mesure où les individus concernés sont enregistrés dans les registres paroissiaux ou d'état civil de la vallée.

Nous avons choisi, dans ce travail un peu expérimental, de faire porter l'observation sur quatre cohortes. Sont recherchées les ascendances de tous les individus nés dans une des cinq communes de la vallée de 1800 à 1819 (cohorte 1), de 1850 à 1869 (cohorte 2), de 1894 à 1913 (cohorte 3) et de 1950-à 1969 (cohorte 4). C'est ainsi sur l'observation de 6452 généalogies ascendantes, plus ou moins profondes et plus ou moins complètes, que reposent les analyses qui suivent. Nous pourrions également comparer les observations selon les quatre cohortes de naissance étudiées et déceler l'existence éventuelle de modifications dans cette « structure généalogique ».

Tableau 1 – Effectifs étudiés selon les cohortes

2 – Complétude et profondeur généalogique : des populations hétérogènes :

2.1 – Complétude des lignées ascendantes :

Par définition, tout individu possède 2 ascendants directs à la génération 1 (son père et sa mère), 4 ascendants à la génération 2 (ses grands-parents), et ainsi de suite avec un doublement du nombre d'ascendants à chaque génération (2, 4, 8, 16, 32, 64 ...)¹. On nomme « complétude » d'une généalogie le rapport entre le nombre théorique de rangs d'ascendants à

¹ Cette régularité peut bien entendu être rompue en cas de naissance d'enfant de père inconnu, mais ce phénomène est marginal dans la région étudiée.

une génération donnée et le nombre de rangs effectivement occupés par des individus identifiés. Par exemple, si parmi les 8 arrières-grands-parents d'un individu, 6 seulement sont connus, la complétude est de 75% à cette génération. L'étude de la complétude des lignées ascendantes des natifs de la vallée fournit donc une première approche globale de la présence des ascendants dans la vallée.

Tableau 2 – Complétude des généalogies : nombre moyen d'ancêtres connus selon la génération, pour les 4 cohortes.

Graphique 1 – Nombre moyen d'ancêtres connus selon la génération.

Logiquement, c'est pour la quatrième et dernière cohorte, naissances survenues entre 1950 et 1969, que la profondeur généalogique connue est la plus importante. Dans les meilleurs des cas, quelques ascendants de la 12^e génération sont identifiés, ce qui correspond à des ancêtres nés environ trois siècles et demi auparavant, c'est à dire dans la première moitié du XVII^e siècle. Le nombre d'ascendants connus est cependant nettement inférieur au nombre théorique d'ascendants, à chaque génération. En moyenne, dès la seconde génération (grands parents) seulement 3.2 ascendants sur 4 sont connus (80%), et à la troisième génération 4.9 arrières-grands-parents sur 8 sont connus (61.25%). Le nombre absolu d'ascendants connus augmente jusqu'à un maximum de 31.9 individus à la 7^e génération. Cependant, la complétude n'est alors que de 24.9%. Ceci témoigne du fait que, globalement, une grande majorité des ascendants des individus nés dans la vallée au milieu du XX^e siècle sont en fait extérieurs à la vallée étudiée dès les premières générations ascendantes.

Pour la cohorte précédente, naissances enregistrées entre 1894 et 1913, notre connaissance de l'ascendance s'arrête par définition à une profondeur moindre. La complétude est légèrement inférieure au 2^e et 3^e générations ascendantes, signe d'un apport de population extérieure plus important dans la seconde moitié du XIX^e siècle. Par contre de la 3^e à la 5^e génération ascendante la complétude est légèrement supérieure (15.1 ascendants connus sur 32, soit 47.2% au lieu de 11.6 soit 36.3.2% pour la cohorte 4), témoignant d'une moindre mobilité de la population durant les décennies précédentes. Par contre, à partir de la 6^e génération ascendante (150 à 200 ans plus tôt selon les lignées) l'information généalogique est moindre du fait essentiellement des lacunes documentaires du XVII^e siècle.

Pour la seconde cohorte de naissance (1850-1869) on observe le même phénomène : une meilleure complétude pour les premières générations ascendantes. Le nombre maximum d'ascendants est connu à la 4^e génération avec 8.8 ancêtres, soit un complétude de 55% contre respectivement 56.9% et 41.35% pour les cohortes précédentes de naissance. Par contre la profondeur généalogique connue se réduit encore pour les mêmes raisons que précédemment. Enfin, pour la première cohorte de naissances (1800-1819) le nombre maximum d'ascendants connus se situe à la 3^e génération, avec 4.3 individus, soit une complétude de 53.8% seulement. Les lacunes documentaires sont responsables de ce fait, ainsi que de l'absence de connaissance des ancêtres des générations précédentes

L'analyse de la distribution des généalogies selon le nombre d'ascendants connus par génération tempère ce premier constat global en soulignant la diversité de la connaissance que nous avons des ascendants de chaque individu étudié.

Tableau 3 – Distribution des généalogies selon le nombre d'individus connus à la seconde génération²

² Peu de généalogies comportant un nombre impair d'ascendants, les valeurs ne sont pas présentées pour alléger le tableau.

Par exemple à la seconde génération, pour les 493 individus composant la 4^o cohorte étudiée, nous connaissons les 4 grands-parents pour 282 individus (57.2%), deux grands-parents pour 94 individus (19.1%) et aucun grand-parent pour 113 individus (22.9%). Nous pouvons donc affirmer qu'un quart des individus nés dans la vallée au milieu du XX^e siècle n'y possédaient aucun ascendant deux générations plus tôt, c'est à dire qu'il s'agit de lignées familiales d'implantation récente. A contrario, pour la moitié des lignées nous connaissons tous les grands-parents, ce qui signale un ancrage local fort. Pour les cohortes précédentes, la part de natifs dont aucun grand-parent ne nous est connu est nettement plus faible, de l'ordre de 11 à 15%, révélant un meilleur ancrage local des ascendants. A l'opposé, nous connaissons pour ces trois cohortes précédentes les 4 grands-parents dans environ les trois-quarts des cas. La dernière cohorte, qui correspond aux individus nés vers le milieu du XX^e siècle se distingue donc nettement des autres.

Tableau 4 – Distribution des généalogies selon le nombre d'individus connus à la quatrième génération ³

Graphique 2 – Nombre moyen d'ancêtres connus à la quatrième génération

Par définition, notre connaissance des généalogies ne peut qu'être moindre à la 4^o génération ascendante. Pour la 4^o cohorte de naissance, nous connaissons en moyenne 5.6 ascendants par généalogie sur 16 ancêtres, soit une complétude de 35%. Mais les cas sont très divers. Pour 28.2% des individus nous ne connaissons aucun ascendant à cette génération, et pour 23% nous connaissons entre 2 et 6 ascendants. C'est donc pour la moitié des individus que la complétude est inférieure à 50%. A l'autre extrémité de la distribution, nous avons une connaissance complète (16 ascendants connus sur 16) pour 13.6% des généalogies, soit tout de même près de 1 sur 7.

Pour les 2^o et 3^o cohortes, le nombre moyen d'ancêtres connus à cette 4^o génération est pratiquement identique, respectivement 8.8 et 9.1 ascendants. Par contre les distributions sont différentes. Pour la 3^o cohorte (individus nés à la charnière des XIX^e et XX^e siècles) les situations sont très contrastées : la totalité des ascendants est connue dans 29.4% des cas tandis que aucun ascendant n'est connu dans 20.5% des cas, les situations intermédiaires étant moins nombreuses. La complétude est égale ou supérieure à 50% dans deux tiers des généalogies (65.6%).

Pour la seconde cohorte, ces valeurs extrêmes sont moins fréquentes. La complétude est aussi égale ou supérieure à 50% dans les deux tiers des généalogies (66.5%). Mais il est plus fréquent que 12 ou 14 des ascendants soient connus que pour la cohorte précédente, c'est à dire une information un peu moins complète. Toutefois, ces différences entre ces deux cohortes correspondent probablement plus à l'état des sources qu'à des modifications de comportements démographiques.

2.2 – Profondeur généalogique: génération ascendante la plus élevée atteinte

Tableau 5 – Distribution des ascendances selon la profondeur générationnelle maximum atteinte, selon les cohortes.

Graphique 3 – Profondeur générationnelle maximum de la cohorte 4

³ La première cohorte n'est pas présentée car l'observation est souvent incomplète du fait des sources. Peu de généalogies présentant un nombre impair d'ascendants, les valeurs ne sont pas présentées pour alléger le tableau.

Si l'on examine notre connaissance de toutes les lignées ascendantes des individus composant la 4^e cohorte (naissances en 1950-1969) on observe une distribution nettement bi-modale. D'une part, pour un quart environ des individus, notre connaissance généalogique ne remonte pas au-delà de la 3^e génération. Il s'agit là de lignées familiales récemment arrivées dans la vallée, par immigration durant la première moitié du XX^e siècle, ceci alors que la vallée connaissait globalement une forte dépopulation. Les situations intermédiaires sont rares, et à la droite de la distribution on trouve un second groupe, plus important que le premier, avec environ les trois quarts des individus pour lesquels on retrouve au moins un ascendant à la 9^e génération ou au-delà. Ce second groupe se caractérise donc par une implantation locale continue fort ancienne (environ trois siècles), au moins par un ascendant⁴.

Les résultats obtenus pour les autres cohortes de naissance vont dans le même sens, avec bien entendu une profondeur généalogique maximum atteinte moindre du fait de la documentation qui ne remonte pas au-delà du XVII^e siècle. La proportion d'individus n'ayant qu'une faible implantation généalogique locale varie relativement peu entre 11.1% (2^e cohorte) et 14.8% (3^e cohorte). Dans toutes les cohortes on observe une distribution bi-modale, avec deux groupes distincts parmi les individus : ceux d'implantation récente, minoritaires, et ceux d'implantation ancienne, représentant entre les deux tiers et les trois quarts des effectifs.

Tableau 6 – Distribution des ascendances selon la profondeur généalogique maximum atteinte : deux groupes principaux.

3 – Déperdition de l'information généalogique entre les générations 2 et 4

Pour trois des quatre cohortes retenues, il est possible de mesurer la perte d'information généalogique entre la seconde et la quatrième génération ascendante. Cela nous renseigne sur l'origine géographique des ascendants récents et sur l'importance de l'apport extérieur.

On forme les distributions du nombre d'ancêtres à la génération 4 conditionnellement au nombre d'ancêtres à la génération 2, selon les cohortes. La première cohorte est exclue en raison de la très importante perte d'information à la 4^e génération (ancêtres nés avant les registres paroissiaux)

Tableau 7a – Déperdition de l'information généalogique entre les générations 2 et 4

Ce tableau permet de souligner la diversité des généalogies telles qu'elles sont connues par les sources locales. Parmi les 493 individus composant la 4^e cohorte (milieu du XX^e siècle), 282 ont en commun le fait que leurs 4 grands-parents sont identifiés dans la base de données. Mais, lorsque l'on regarde le nombre d'ancêtres connus à la 4^e génération, ces 282 individus se répartissent en 11 situations différentes sur 12 possibles. Aux extrémités de la distribution, les 16 ancêtres à la 4^e génération sont connus pour 67 individus, tandis que aucun des 16 ascendants de la 4^e génération n'est connu pour 16 individus.

Une hétérogénéité de même nature se rencontre pour les autres cohortes. Toutefois, pour les individus nés à la charnière XIX^e-XX^e siècles (cohorte 3) et pour ceux nés au milieu du XIX^e siècle (cohorte 2), les effectifs les plus nombreux se rencontrent dans les meilleurs degrés de connaissance. Par exemple, pour les 1481 individus nés en cohorte 2 et pour lesquels les 4

⁴ Nous avons vu ci-dessus que la complétude des lignées à cette profondeur généalogique est très faible. La majorité des ascendants de la 9^e génération et au-delà ne sont pas connus en raison des lacunes documentaires ou sont originaires d'autres communes.

grands-parents sont connus, on connaît de 12 à 16 ancêtres à la 4^o génération dans 895 cas (60.4%). Par comparaison, on peut affirmer que la déperdition d'information supérieure entre la 2^o et la 4^o génération ascendante pour les individus nés au milieu du XX^e siècle signale un apport plus important d'individus nés hors de la vallée parmi les ancêtres de ces individus. Ce point est tout à fait compatible avec ce que l'on sait par ailleurs des mouvements migratoires (Bideau et al, 1986).

3.1 - Construction d'un coefficient de déperdition

Pour synthétiser la déperdition d'ancêtres entre une génération p et une génération q , $q > p$, dans une base de données généalogiques, nous proposons un ensemble de coefficients simples à calculer et d'interprétation aisée qui sont fondés sur le calcul du nombre d'ancêtres perdus entre les générations p et q . Pour un individu donné, on calcule son nombre potentiel d'ancêtres à la génération q compte tenu de son nombre d'ancêtres à la génération p . La différence entre ce nombre potentiel et le nombre réel est la déperdition absolue individuelle. Par exemple, un individu qui n'a que 2 ancêtres dans la base à la génération 2 a un nombre potentiel de 8 ancêtres dans la base à la génération 4. S'il n'a effectivement que 6 ancêtres dans la base, sa déperdition absolue individuelle est de 2 ancêtres. Dans la mesure où les déperditions absolues s'ajoutent, on calcule par sommation la déperdition absolue à i et j fixés, puis celle à i fixé et enfin la déperdition absolue totale pour l'ensemble des individus considérés. Pour comparer des situations où le nombre d'ancêtres à la génération p n'est pas le même, on forme le coefficient de déperdition relative. Pour un ensemble d'individus, celui-ci est le rapport de la déperdition absolue observée pour l'ensemble de ces individus au nombre potentiel d'ancêtres qu'ils représentent. On adopte les notations ci-dessous :

- X : nombre d'ancêtres à la génération p , prend les valeurs $i = 0, 1, \dots, 2^p$
- Y : nombre d'ancêtres à la génération q , prend les valeurs $j = 0, 1, \dots, 2^q$
- n_{ij} : nombre d'individus ayant i ancêtres à la génération p et j ancêtres à la génération q . On note n_{i+} et n_{+j} les fréquences absolues marginales.
- On désigne par \bar{X} et \bar{Y} , les moyennes marginales respectives de X et Y , par \bar{Y}_i la moyenne de la distribution conditionnelle $Y/X = i$.

Déperdition absolue. Les déperditions absolues s'ajoutent :

- déperdition absolue individuelle ; un individu qui a i ancêtres à la génération p et j à la génération q a potentiellement $i2^{q-p}$ ancêtres à la génération q , alors qu'il n'en a réellement que j , soit une déperdition absolue individuelle égale à $i2^{q-p} - j$.
- déperdition absolue conditionnelle au fait d'avoir i ancêtres à la génération p et j à la génération q , notée D_{ij} : il y a n_{ij} individus concernés, soit

$$D_{ij} = n_{ij} (i2^{q-p} - j) \quad [1]$$

- déperdition absolue conditionnelle au fait d'avoir i ancêtres à la génération p , notée D_i ; elle s'obtient en faisant la somme pour j des déperditions D_{ij} à i fixé, soit :

$$D_i = \sum_{j=0}^{i2^{q-p}} D_{ij} = \sum_{j=0}^{i2^{q-p}} n_{ij} (i2^{q-p} - j) = in_{i+} 2^{q-p} - \sum_{j=0}^{i2^{q-p}} jn_{ij} = n_{i+} (i2^{q-p} - \bar{Y}_i) \quad [2]$$

- déperdition absolue globale, notée D ; elle se calcule comme la somme des déperditions absolues de tous les individus, soit :

$$D = \sum_{i=0}^{2^p} D_i = 2^{q-p} \sum_{i=0}^{2^p} in_{i+} - \sum_{i=0}^{2^p} n_{i+} \bar{Y}_i = n(\bar{X} 2^{q-p} - \bar{Y}) \quad [3]$$

Déperdition relative. Le nombre d'ancêtres à la génération p jouant sur le nombre potentiel d'ancêtres à la génération q et sur la déperdition, on évalue la déperdition en divisant la déperdition absolue d'une entité par le nombre potentiel d'ancêtres de cette entité, ce qui

permet de comparer la déperdition de plusieurs entités n'ayant pas le même nombre d'ancêtres à la génération p .

- déperdition relative individuelle et déperdition relative conditionnelle au fait d'avoir i ancêtres à la génération p et j à la génération q , cette dernière notée Δ_{ij} : on divise chacune des deux déperditions absolues par le nombre potentiel d'ancêtres correspondants, obtenant évidemment le même résultat, donné pour $i \geq j$ par :

$$\Delta_{ij} = \frac{D_{ij}}{n_{ij}i2^{q-p}} = 1 - \frac{j}{i2^{q-p}} \quad [4]$$

- déperdition relative conditionnelle au fait d'avoir i ancêtres à la génération p , notée Δ_i : la déperdition conditionnelle relative des individus qui ont i ancêtres à la génération p est définie par :

$$\Delta_i = \frac{n_{i+}(i2^{q-p} - \bar{Y}_i)}{n_{i+}i2^{q-p}} = 1 - \frac{\bar{Y}_i}{i2^{q-p}} \quad [5]$$

- déperdition relative globale, notée Δ : la déperdition relative globale s'écrit :

$$\Delta = 1 - \frac{\bar{Y}}{X2^{q-p}} \quad [6]$$

3.2 - Evaluation de la déperdition entre les générations 2 et 4 dans la vallée de la Valserine

Le calcul pour chaque cohorte des coefficient de déperdition relative conditionnellement au nombre d'ancêtres i de la génération 2, puis globalement, sont rassemblés dans le tableau ci-dessous.

Tableau 7b – Déperdition relative conditionnelle et globale entre les générations 2 et 4

Deux constatations principales peuvent être énoncées à partir de ces données. Tout d'abord, la déperdition de la cohorte 4 est la plus importante, puisque près de 40 % des ancêtres potentiels ne sont pas retrouvés, alors que celle de la cohorte 3 est la plus faible. Paradoxalement, pour la cohorte 4, la déperdition Δ_2 correspondant aux individus qui ont 2 ancêtres à la génération 2 est nettement plus faible que celle liée à ceux qui ont 4 ancêtres ; elle est même inférieure à la déperdition Δ_2 pour les cohortes 2 et 3. En revanche, la déperdition Δ_4 est très élevée, étant à la base de la plus forte déperdition globale de la cohorte 4.

4 – Caractérisation des ascendances les mieux connues

Il est possible de mieux connaître l'implantation généalogique des individus composant les quatre cohortes en prenant en compte quelques caractéristiques telles que le rang SOSA ou le sexe des ascendants les plus éloignés ayant vécu dans la vallée.

4.1 – Situation des ancêtres dans les arbres généalogiques : les rangs SOSA

L'espagnol Sosa a mis au point, au XVII^e siècle une numérotation des individus en fonction du rang qu'ils occupent dans l'ascendance d'un sujet. Cette numérotation a été reprise ultérieurement par Stradonitz et elle est très largement utilisée par les généalogistes, y compris par les éditeurs de logiciels pour la gestion informatisée des généalogies. Dans cette numérotation, le sujet dont on étudie l'ascendance porte le n°1. Son père porte le n°2 et sa mère le n°3, son grand-père paternel porte le n°4, sa grand-mère paternelle le n°5, son grand-

père maternel le n°6, sa grand-mère maternelle le n°7, et ainsi de suite. Il découle de cette logique que tous les individus de sexe masculin sont porteurs d'un numéro pair et ceux de sexe féminin d'un numéro impair. Le numéro des individus situés sur la lignée agnatique (le père, son père, le père de celui-ci et ainsi de suite) portent un numéro qui correspond au nombre d'ascendants à cette génération : par exemple le grand-père paternel porte le n°4 et il y a 4 ascendants à cette génération. Tout individu de sexe masculin est porteur d'un numéro égal au double de celui de son enfant (par exemple l'ancêtre n°14, de sexe masculin, est le père de l'individu 7, de sexe féminin). Tout individu de sexe féminin est porteur d'un numéro égal au double + 1 du numéro de son enfant (par exemple l'ancêtre n°15, de sexe féminin, est la mère de l'individu n°7). Du fait de cette numérotation simple et logique il est possible de retrouver rapidement le cheminement qui mène d'un ascendant au sujet étudié⁵.

Tableau 8 – Fréquence des n° SOSA connus selon la génération, Rangs pairs de 2 à 31 (de la 2° à la 4° génération ascendante).

Graphique 4 – Fréquence des n° SOSA connus selon la génération

C'est encore une fois pour la quatrième cohorte (naissances en 1950-1969) que l'observation possède la profondeur la plus importante. On n'observe pratiquement aucune différence entre les sexes, ceci étant dû au fait que lorsque le père d'un individu est connu sa mère l'est en général également, et réciproquement. Pour cette raison, et afin d'éviter d'alourdir le tableau, nous ne présentons les fréquences que pour les rangs SOSA pairs (sexe masculin).

Dans les généalogies ascendantes des individus appartenant à la cohorte 4, on observe que les rangs SOSA correspondant aux lignées paternelles sont plus fréquemment occupés que les rangs correspondant aux lignées maternelles. Ainsi à la seconde génération on connaît 72% des grands-parents paternels contre 61.9% des grands-parents maternels. A la troisième génération ascendante on connaît 57% des pères des grands-pères paternels (rang SOSA 8) contre 45% des pères des grands-pères maternels (rang SOSA 12).

Le même phénomène se retrouve aux générations suivantes et est perceptible sur le graphique : la lignée agnatique est toujours mieux connue que les autres, ce qui signale que les ascendants de cette lignée sont plus souvent originaires de la vallée que ceux des autres lignées. Encore une fois, ceci est compatible avec ce que l'on sait des comportements démographiques et sociaux, et notamment de la tradition virilocale selon laquelle un nouveau couple s'installe de manière privilégiée dans le lieu d'origine, ou à défaut de résidence, de l'homme.

Globalement le même phénomène est observé au sein des généalogies des individus composant les cohortes précédentes. En comparant les résultats obtenus pour les différentes cohortes on remarque que les fréquences sont en général plus élevées pour les individus nés dans les cohortes les plus anciennes. Par exemple, le grand-père paternel (rang SOSA 4) est connu pour 85% des individus nés au XIX° siècle (cohortes 1 et 2), pour 81.4% des individus nés à la charnière du XIX° et du XX° siècle (cohorte 3) et pour 72.4% *seulement* de ceux nés au milieu du XX° siècle (cohorte 4).

On peut affirmer que l'ascendance locale correspond plus fréquemment aux lignées masculines qu'aux lignées féminines. En d'autres termes, la sédentarité au fil de plusieurs générations est plus fréquente lorsque la lignée paternelle est déjà implantée dans la vallée que lorsqu'il s'agit de la lignée maternelle. On « *est du pays* » avant tout si les grands parents paternels sont nés dans la vallée. Ceci est compatible avec les résultats présentés ci-dessus et

⁵ Par exemple, l'ancêtre n° 453 est une femme mère de 226. 226 est un homme père de 113. 113 est une femme mère de 56. 56 est un homme père de 28. 28 est un homme père de 14. 14 est un homme père de 7. 7 est une femme mère de 3. 3 est une femme, mère de 1, le sujet de la généalogie.

signale que la mobilité géographique s'est accrue à partir de la fin du XIX^e siècle. La méthode utilisée ici permet de mesurer précisément l'impact de ces flux migratoires sur la constitution des générations nées au XX^e siècle

4.2 – Des chemins remarquables : le rôle du sexe

En règle générale, on connaît les deux parents de chaque individu. C'est ainsi que globalement la proportion d'hommes et de femmes est quasiment la même dans chaque liste d'ascendance, ce qui suggère d'étudier le rôle du sexe à partir des chemins remarquables.

Comme nous l'avons déjà vu, la lecture du tableau présentant les rangs SOSA 2 à 31 (générations 1 à 4) montre que les parents du père sont plus souvent de la vallée que les parents de la mère, mais que la différence s'estompe au fil des générations.

Chaque n^o SOSA définit une lignée depuis la racine, marquée par une succession d'ascendants. Pour comparer les différentes lignées qui mènent de la racine aux extrémités de l'arbre d'ascendance tronqué à la 5^e génération et mettre en évidence le poids du sexe, nous avons classé les rangs SOSA des cohortes 3 et 4, les plus fiables à ce niveau de génération, en fonction de leur fréquence d'apparition. Dans ce tableau, plus une lignée est fréquente, plus petit est son rang. On remarque que les lignées les plus masculines sont favorisées.

Tableau 9 – Rang des lignées de la 5^{ème} génération (SOSA 32 à 63) selon leur fréquence et nombre d'hommes dans la lignée.

Graphique 5 – Rangs des lignées de la 5^{ème} génération.

Graphique 6 : Rangs moyens des lignées de 5^e génération en fonction du nombre d'hommes dans la lignée pour les cohortes 3 et 4.

Conclusion

C'est bien à un travail expérimental, de « généalogie quantitative historique », que nous nous sommes livrés ici. Les concepts et les méthodes mériteront d'être précisés et éventuellement modifiés. Mais il apparaît clairement que les corpus généalogiques retracés à partir d'un dépouillement systématique des registres d'état civil constituent bien une source importante pour étudier l'histoire et la constitution d'une population donnée. Ils nous renseignent sur l'origine locale ou extérieure des ascendants des individus natifs d'une région donnée, ici la vallée de la Valserine.

Les observations sont à chaque fois compatibles avec ce que l'on peut mesurer par ailleurs, à partir des registres paroissiaux, des registres d'état civil et des listes nominatives de recensement de la population. Mais ces observations permettent d'aller au-delà et de préciser les fréquences et la chronologie avec laquelle des ascendants originaires de l'extérieur de la vallée s'intègrent dans celle-ci et participent à la reproduction de la population. Bien sûr, les schémas généalogiques brefs correspondant aux familles dont le séjour dans la vallée est de courte durée et n'ayant pas donné lieu à au moins une naissance au sein des quatre cohortes étudiées, nous échappent. Mais par définition celles-ci ne jouent qu'un rôle marginal dans la reproduction de la population locale.

Il est possible d'aller plus loin dans la caractérisation et l'analyse des corpus généalogiques. Par exemple, nous pourrions prendre en considération le lieu de naissance des ancêtres, afin de percevoir si une origine proche est plus favorable qu'une origine plus lointaine pour

l'intégration d'un nouvel arrivant dans la vallée. De même, nous pourrions tenter de distinguer s'il existe des comportements différents selon les villages de la vallée alors que dans cet article nous avons considéré que les cinq communes de la vallée constituaient un ensemble. Des travaux précédents (Bideau et al., 1986) montrent que le processus migratoire n'a pas revêtu la même ampleur, la même chronologie et la même orientation géographique selon les communes. Notamment la commune située dans la partie nord de la vallée (Lélex) échangeait des conjoints et des migrants en proportion élevée avec les communes situées sur les plateaux du Jura, tandis que la commune située dans le sud de la vallée, Montanges, faisait de même avec les communes du Haut-Bugey situé plus au sud.

Cette démarche généalogique est également présente dans un autre aspect de notre étude de la population de cette vallée. Elle est en effet le siège de la principale concentration connue de porteurs de la maladie héréditaire dite « Rendu-Osler » (Plauchu et al., 1992). Nous avons alors cherché, grâce à l'ensemble de réseaux généalogiques, à identifier un éventuel effet fondateur, c'est à dire un ancêtre unique commun à tous les porteurs actuels de la maladie. Cet effet fondateur n'a pu être démontré, l'ancêtre commun, s'il a existé, ayant probablement vécu bien avant l'ouverture des registres paroissiaux qui nous sont parvenus. Par contre, nous avons pu établir que tous les porteurs actuels de la maladie résidant dans la vallée possèdent des ancêtres ayant vécu dans la vallée dès la fin du XVII^e siècle (Bideau et al., 1992). Ils appartiennent donc au groupe défini ci-dessus caractérisé par une implantation locale ancienne, avec une présence généalogique continue sur place sur une douzaine de générations. Pour reprendre le concept de « structure généalogique de la population », nous pensons pouvoir affirmer à partir des réseaux généalogiques que la population de la vallée de la Valserine, loin de constituer un isolat, a connu des flux continus d'entrée et de sortie au cours des trois derniers siècles, même si l'ampleur de ces flux a varié durant cette période. Par contre, un groupe de lignées familiales, que nous baptisons « noyau stable », s'est maintenu sur place sur toute cette période historique.

Références bibliographiques

BIDEAU A. et BRUNET G. (à paraître en 2006), *Essai de démographie historique et de génétique des populations. Une population du Jura méridional du XVII^e siècle à nos jours*. 205 p. dactyl., accepté pour publication, INED.

BIDEAU A., BRUNET G et PLAUCHU H., 1986, « La dynamique des structures familiales à Chézery-Forens(Haut-Jura). *Annales de Démographie Historique*, p.133-154.

BIDEAU A., BRUNET G., PLAUCHU H., HEYER E. et ROBERT J.M., 1992, “An abnormal concentration of cases of Rendu-Osler disease in the valley of Valserine (French Jura). Genealogical and demographic study”. *Annals of Human Biology*, 19 (3), p. 233-247

BIDEAU A., BRUNET G., DESJARDINS B. et PROST M., 1995, « La reproduction de la population aux XVII^e, XVIII^e et XIX^e siècles. Exemples français et québécois ». *Annales de Démographie Historique*, p.137-148.

BIDEAU A. et POULAIN M., 1984, « De la généalogie à la démographie historique : généalogie ascendante et analyse démographique ». *Annales de Démographie Historique*, p.55-70.

BOETSCH G., 1999, *L'anthropologie démographique*. Paris, PUF, coll. Que-sais-je, 126p.

BOETSCH G. et PROST M., 2001, « Descendances différentielles, reproduction générationnelle et enfants utiles dans une population isolée de la montagne alpine ». *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, 13, p.39-59.

BOETSCH G., PROST M. et RABINO-MASSA E., 2002, « Evolution of consanguinity in a French Alpine valley, the Vallouise in the Briançon region (17th-19th centuries). *Human Biology*, 75, p.285-300.

BOUCHARD G. et DE BRAEKELEER M., 1991, *Histoire d'un génome*. Québec, Presses de l'Université du Québec, 607p.

BRUNET G., 1979, *Un bourg rural de la Dombes aux XVIII^e et XIX^e siècles. Démographie, société et mentalités*. Doctorat d'Histoire, Université Lyon 2, 2 vol. dactyl.

BRUNET G. et BIDEAU A., 2002, « Démographie historique et généalogie ». *Annales de Démographie Historique* (2), p.101-110.

CAZES M.H., 1987, « Chacun appartient à une génération, mais laquelle ? ». *Population*, 42 (4-5), p.671-684.

DESJARDINS B., 1998, « Le registre de population du Québec ancien ». *Annales de Démographie Historique* (1), p.215-226.

DUPAQUIER J. et KESSLER D., 1992, *La société française au XIX^e siècle*. Paris, Fayard, 529 p.

GOMILA J., 1976, « Définir la population ». In *L'étude des isolats : espoirs et limites*, sous la direction d'A. Jacquard. Paris, PUF-INED, p.5-36.

HENRY L., 1956, *Anciennes familles genevoises. Etude démographique, XVI^e-XX^e siècles*. Paris, P.U.F., cahier de l'I.N.E.D. n°26.

HEYER E., 1993, "Population structure and immigration : a study of the Valserine Valley (French Jura) from the 17th century to today", *Annals of Human Biology*, 20 (6), p.565-573

HEYER E. et CAZES M.H., 1999, « Les enfants utiles. Une mesure démographique pour la génétique des populations ». *Population*, 54 (4-5), p.677-691.

HOLLINGSWORTH Th., 1964, « The demography of the British peerage », *Population Studies*, XVIII, 2. Supplément.

PLAUCHU H., BRUNET G., BIDEAU A. et ROBERT J.M., 1992, « La maladie de Rendu-Osler. Une cause d'epistaxis et d'angiomes réellement fréquente en France ». *Le Concours Médical*, 114 (29), p.2469-2476

RALLU J.L., 1992, « Note sur l'utilisation des généalogies ». In *Modèles de la démographie historique*, sous la direction de A.Blum, N.Bonneuil et D.Blanchet. Paris, INED, congrès et colloques n°11, p.45-53.