

L'étude de l'origine et de la diffusion des noms de famille est une thématique fortement interdisciplinaire. Ce ne sont pas seulement les sciences historiques et démographiques qui peuvent y contribuer, mais aussi la génétique, qui permet d'étudier sur la longue durée la distribution des noms, ainsi que la physique statistique, qui s'est approchée de cette dernière matière dans la décennie passée.

L'objectif de mon exposé est la présentation des principaux résultats obtenus pour ces domaines dans les études génétiques, ainsi que des thèmes et des techniques mobilisées récemment dans ce champ par la physique statistique, en indiquant les informations les plus importantes qui peuvent être obtenues grâce à ces méthodes tout en illustrant les résultats avec des exemples concrets.

NOMS ET ETUDES DE BIOLOGIE HUMAINE.

L'intérêt pour les noms de famille dans les études de biologie humaine s'est affirmé au cours de la deuxième moitié du XIX^e siècle, mais l'étude systématique de ceux-ci a été mise en train seulement durant les années soixante du XX^e siècle. Ces travaux se concentrent sur deux thèmes principaux, celui des études de consanguinité et celui de la distribution en fréquence des noms de famille et de leurs possibles origines. L'idée d'étudier le degré de consanguinité au moyen des noms de famille remonte à George Darwin (fils de Charles) qui en 1875 calcula (pour l'Angleterre) la probabilité que lors d'un mariage les conjoints aient fortuitement le même nom de famille. Il confronta ses calculs avec les données empiriques, en concluant que la consanguinité due aux mariages parmi des cousins atteignait 4,5% dans l'aristocratie, avec des valeurs plus basses pour la bourgeoisie et plus généralement pour la population urbaine.¹

Ce genre d'étude fut repris seulement après 1965, lorsque Crow et Mange montrèrent que, pour la plupart des typologies de mariages entre parents, la probabilité que des individus apparentés aient le même nom est proportionnelle au degré de consanguinité. Par conséquent le pourcentage des mariages isonymiques constitue une mesure assez directe du niveau de consanguinité de la population². Plus généralement, on peut établir le degré de parenté entre deux populations différentes à partir des pourcentages de personnes ayant le même nom de famille³. Il faut pourtant corriger les résultats pour tenir compte du fait que le même nom peut avoir des origines différentes et n'être pas dû à un ancêtre commun, en particulier du fait que les phénomènes migratoires peuvent influencer les distributions des noms. Il vaut la peine de signaler que certaines études empiriques, parmi les plus importants dans ce contexte, ont été réalisées sur des populations ibériques et ibéro-américaines, grâce au fait que le système espagnol des noms de famille semble offrir la possibilité d'un contrôle immédiat et facile de l'isonymie des ancêtres⁴.

La seconde ligne d'enquête fut ouverte par les travaux de Galton et Watson (1874) sur la probabilité d'extinction des noms de famille⁵. En son temps, ce travail fut motivé par l'inquiétude née de la constatation que de nombreuses familles de l'aristocratie britannique risquaient de s'éteindre. Ces études suscitérent une ligne de recherches mathématiques sur les procès de ramification, les *branching processes*. Les premiers résultats aboutirent à la conclusion erronée que pour chaque nom, la probabilité de s'éteindre après un temps suffisamment long était égale à un ; ensuite, il fut précisé que cette probabilité n'était pas égale à 1 dans une population croissante, mais que toutefois, elle a une valeur non négligeable et proche de 0,5.

¹ G.H. DARWIN, « Marriages between first cousins in England and their effects ».

² J.F. CROW and A.P. MANGE, « Measurement of inbreeding from the frequency of marriages between persons of the same surname ».

³ G.W. LASKER, « A coefficient of relationship by isonymy: A Method for Estimating the genetic Relationship between Populations » ; G.W. LASKER, *Surnames and Genetic Structure*.

⁴ J. PINTO-CISTERNAS *et alii*, « Estimation of Inbreeding from Isonymy in Iberoamerican Populations » ; G.W. LASKER, « Datos sobre los apellidos hispanoamericanos en los estudios de biología humana » ; S.E. COLANTONIO *et alii*, « Use of Surname Models in Human Population Biology: A Review of Recent Developments ».

⁵ F. GALTON and H.W. WATSON, « On the Probability of the Extinction of Families ».

Il est intéressant d'appliquer les principes statistiques qui sont à la base des études d'isonymie pour faire la comparaison entre le nombre de mariages qui ont eu lieu entre les familles d'une population particulière et la prédiction théorique obtenue par le calcul de la probabilité aléatoire. Si par exemple nous prenons les mariages endogamiques des Xuetes de Majorque (entre 1565 et 1750), qui ont été étudiés par E. Porqueres,⁶ et considérons seulement les onze familles (dans un total de 17) qui sont mentionnées le plus souvent (les chiffres trop petites ne pouvant être sérieusement analysées du point de vue statistique) nous pouvons construire le tableau qui suit, tableau I.

Tableau I - LES INTERMARIAGES DES FAMILLES XUETES DE MAJORQUE

	Fuster	Forteza	Aguilò	Cortés	Pinya	Pomar	Bonnin	Martí	Mirò	Picò	Valls
Fuster	22	13	10	7	13	10	5	2	4	7	10
Forteza	11	19	15	17	5	7	2	10	7	1	3
Aguilò	15	18	12	15	7	6	8	11	4	2	5
Cortés	4	13	13	18	2	6	2	14	0	0	6
Pinya	17	7	11	0	6	13	10	2	3	10	7
Pomar	12	12	6	2	11	10	8	7	4	4	4
Bonnin	14	5	7	4	9	11	4	0	8	8	5
Martí	5	16	6	20	0	5	2	10	0	0	0
Mirò	7	3	8	1	5	6	10	0	4	9	4
Picò	8	2	4	0	10	6	4	0	6	3	4
Valls	4	7	3	4	7	5	0	1	4	3	1

Le calcul des probabilités nous permet d'obtenir la valeur plus probable pour la distribution des intermariages du même groupe de personnes : nos résultats sont présentés dans le tableau II.

Tableau II - LES INTERMARIAGES DES XUETES : VALEUR STATISTIQUEMENT PLUS PROBABLE

	Fuster	Forteza	Aguilò	Cortés	Pinya	Pomar	Bonnin	Martí	Mirò	Picò	Valls
Fuster	15	15	13	12	10	11	7	7	6	6	7
Forteza	14	13	11	11	9	9	6	7	5	5	6
Aguilò	14	14	12	11	10	10	7	7	5	6	6
Cortés	11	11	9	9	8	8	5	6	4	4	5
Pinya	12	11	10	9	8	8	5	6	4	5	5
Pomar	11	10	9	8	7	7	5	5	4	4	5
Bonnin	11	10	9	8	7	7	5	5	4	4	5
Martí	9	9	8	7	6	7	4	5	3	4	4
Mirò	8	8	7	6	5	6	4	4	3	3	4
Picò	7	6	5	5	4	5	3	3	2	3	3
Valls	5	5	4	4	4	4	3	3	2	2	2

La comparaison entre les deux tableaux montre qu'il y a de toute évidence des phénomènes qui ne peuvent être expliqués comme simples fluctuations statistiques et qui nécessitent donc une explication anthropologique : il faut au moins mentionner l'excès de mariages isonymiques des Fuster, des Forteza, des Cortés et des Martí, l'excès de mariages du type Pinya-Picò, du type Picò-Mirò et à l'intérieur du groupe Forteza-Cortés-Martí ; et par contre la faiblesse des mariages des

⁶ E. PORQUERES-I-GENE, *Lourde alliance. Mariage et identité chez les descendants des juifs convertis de Majorque (1435-1750)*

individus appartenant au groupe Forteza-Cortés-Martí avec les membres de la plupart des autres familles.

NOMS ET GENETIQUE

La génétique ne s'intéressa guère aux noms de famille jusqu'aux années soixante-dix, lorsque Yasuda, Cavalli-Sforza et leurs collaborateurs reconnurent que la transmission des noms obéit aux mêmes lois biologiques et statistiques que la transmission des allèles neutres (la large partie du patrimoine génétique qui ne conditionne pas le phénotype, et qui donc n'est pas assujettie à la pression évolutive)⁷. En particulier les noms de famille européens se transmettent exactement comme le chromosome Y, en passant de père en fils dans la ligne masculine. Comme pour les gènes, leur distribution se développe dans le temps sous l'effet de deux phénomènes concomitants : la mutation et la migration. Par conséquent on appliqua aux noms de famille les mêmes conclusions et les mêmes déjà appliquées pour les allèles neutres, et en particulier la théorie mathématique formulée en 1967 par Karlin et McGregor. Celle-ci tenait explicitement compte de la possibilité qu'avec le temps le patrimoine génétique soit sujet à des mutations⁸. La version simplifiée de la théorie applicable dans le cas de populations très nombreuses (distribution de Fisher) devint rapidement le moyen ordinaire de paramétrer les distributions des noms⁹.

Un grand intérêt fut aussi soulevé par l'observation de Fox et Lasker (1983), qui relevèrent sur une base purement empirique que la fréquence des noms était décrite avec une remarquable précision par la courbe de Pareto, originellement introduite pour décrire la distribution des revenus, mais plus tard réutilisée dans l'analyse quantitative de nombreux phénomènes sociaux¹⁰. Un des aspects plus intéressants des analyses sur la distribution des noms de famille est la possibilité de calculer, sur la base des données empiriques relatives à la distribution actuelle, une estimation statistique pour la fraction de noms qui se sont modifiés ou ajoutés à une hypothétique population initiale. Ainsi, on peut estimer l'importance des phénomènes migratoires si le nombre des mutations peut être considéré comme relativement négligeable.

L'étude des phénomènes migratoires a fait l'objet de beaucoup d'attention de la part des génétistes, qui ont étudié les dynamiques migratoires en Sardaigne (1984)¹¹ aussi bien que pour le reste de l'Italie (1987)¹² et la France (1992)¹³. Le groupe dirigé par J. Barraï, en exploitant à partir de 1995 les possibilités offertes par les versions digitalisées des listes téléphoniques, a étendu l'analyse à tous les principaux Pays européens, jusqu'à couvrir environ 9% des populations. Des estimations intéressantes sur les migrations à l'échelle continentale ont ainsi pu être établies, ainsi que sur les effets de séparation et d'isolement dus à la distance spatiale et aux barrières géographiques et linguistiques¹⁴.

Une autre ligne de recherche a été engendrée par l'observation de l'analogie entre le nom de famille et le chromosome Y, et concerne la question de l'origine (monophylétique ou polyphylétique) des noms de famille. Les études que Sykes et Irven (2000) ont conduit sur le DNA d'un groupe de

⁷ N. YASUDA *et alii*, « The Evolution of Surnames: An Analysis of Their Distribution and Extinction ».

⁸ S. KARLIN and J. MC GREGOR, « The number of mutant forms maintained in a population ».

⁹ R.A. FISHER *et alii*, « The Relation Between the Number of Species and the Number of Individuals in a random sample of an Animal Population ».

¹⁰ W.R. FOX and G.W. LASKER, « The Distribution of Surname Frequencies ».

¹¹ G. ZEI *et alii*, « Surnames as Neutral Alleles: Observations in Sardinia » ; G. ZEI *et alii*, « Surnames in Sardinia I. Fit of frequency distributions for neutral alleles and genetic population structure » ; G. ZEI *et alii*, « Surnames in Sardinia III. The spatial distribution of surnames for testing neutrality of genes » ; E. WIJMAN *et alii*, « Surnames in Sardinia II. Computation of migration matrices from surname distributions in different periods ».

¹² A. PIAZZA *et alii*, « Migration rates of human populations from surname distributions » ;

A. PIAZZA *et alii*, « A genetic history of Italy ».

¹³ P. DARLU et J. RUFFIE, « L'immigration dans les départements français étudiée par la méthode des patronymes » ; P. DARLU et J. RUFFIE, « Relationships between consanguinity and migration rate from surname distributions and isonymy in France » ; P. DARLU *et alii*, « Quelques statistiques sur la distribution des patronymes en France » ; A. DEGIOANNI *et alii*, « Patronymes italiens et migration italienne en France ».

¹⁴ C. SCAPOLI *et alii*, « Correlations between Isonymy Parameters » ; C. SCAPOLI *et alii*, « Surnames in Western Europe: A comparison of the subcontinental populations through isonymy ».

porteurs mâles du nom Sykes, pour lequel les sources documentaires fouillées par les historiens semblaient indiquer une origine multiple, a montré que les données génétiques sont cohérentes avec l'hypothèse que, compte tenu des inévitables cas de non-paternité, tous les Sykes descendent d'un seul ancêtre mâle commun¹⁵. Parmi d'autres études intéressantes citons celles relatives au nom hébreu Cohen, pour lequel l'hypothèse biblique d'une descendance patrilinéaire directe à partir d'Aaron frère de Moïse résulte génétiquement plausible. Il en va de même des études sur l'ascendance séparée des irlandais ayant un nom de famille d'origine gaélique et ceux ayant un nom anglo-saxon ; évoquons aussi, tout récemment (2007), les recherches sur l'influence génétique des Vikings dans le nord-ouest de l'Angleterre.

NOMS ET PHYSIQUE STATISTIQUE

Pour leur part, mathématiciens et physiciens ont attirés notre attention sur l'existence de structures universelles en phénomènes très différents, cherchant à expliquer l'origine dynamique de celles-ci par la construction de modèles mathématiques, en l'occurrence des modèles d'évolution des populations.

L'analyse de la consanguinité a reçu une importante contribution avec la publication des travaux de Derrida, Manrubia et Zanette (1999), dans lesquels ont été comparés d'une part un modèle théorique et d'autre part les valeurs empiriques de la répétition des ancêtres dans l'arbre généalogique du roi d'Angleterre d'Edouard III¹⁶. Un autre exemple très clair du phénomène de la répétition des ancêtres est donné par Don Carlos d'Habsbourg, petit-fils de l'Empereur Charles Quint. À la quatrième génération ses ancêtres étaient seulement quatre (en lieu d'huit), à la sixième génération ils étaient douze (en principe 32) et à la dixième ils étaient 83, contre 512 selon le modèle théorique. Malheureusement, malgré l'énorme diffusion de matériels généalogiques via Internet et la création de bases de données contenant des centaines de milliers de noms et leurs connexions familiales, peu d'études ont eu lieu pour parvenir à une validation définitive de ces modèles.

Pour ce qui concerne la distribution des noms, le résultat le plus avancé est dû au groupe coréen de Baek, Tiet et Kim, qui a formulé en 2007 un modèle très général de dynamique des populations¹⁷. Leur modèle englobe les modèles antérieurs. En effet, il prend en compte les probabilités de naissance, de décès et de migration tout en établissant des lois statistiques pour la croissance du nombre des noms de famille par rapport à la croissance globale de la population. Il ne néglige pas pour autant la probabilité qu'un nom apparaisse dans la population en fonction du nombre des individus porteurs de ce nom¹⁸. Dans tous les cas, on a montré que la distribution tend vers une courbe de Pareto, mais le paramètre caractéristique de la courbe dépend des modalités avec lesquelles les nouveaux noms sont ajoutés. Elle prend des valeurs significativement différentes dans les cas d'absence de mutation et dans celui d'absence de migration, avec des valeurs intermédiaires dans les situations mixtes. L'analyse des données empiriques, en particulier pour des Pays tels que la Corée et la Chine, où les mutations des noms ont été négligeables, montre un remarquable accord entre le modèle et la réalité historique.

Ces résultats ont été tout récemment reproduits dans la thèse d'A. De Luca, qui a utilisé le groupe de *renormalisation*, un outil tout propre de la physique théorique, ce qui a permis de reconduire la dynamique des noms de famille au contexte plus général des phénomènes de criticité auto-organisée.¹⁹ Ces phénomènes semblent caractériser beaucoup de systèmes complexes, soit physiques que biologiques, économiques et sociaux, dans lesquels ne soit pas présente une échelle caractéristique et les conditions d'auto similarité soient vérifiées.

¹⁵ B. SYKES and C. IRVEN, « Surnames and the Y Chromosome » ; M. A. JOBLING, « In the name of the father: surnames and genetics ».

¹⁶ B. DERRIDA *et alii*, « Statistical Properties of Genealogical Trees » ; B. DERRIDA *et alii*, « Distribution of repetitions of ancestors in genealogical trees ».

¹⁷ D.H. ZANETTE and S.C. MANRUBIA, « Vertical transmission of culture and distribution of family names » ; W.J. REED et B.D. HUGHES, « On the distribution of family names ».

¹⁸ S.K. BAECK *et alii*, « Family name distributions: Master equation approach ».

¹⁹ A. DE LUCA et P. ROSSI, « Renormalization group evaluation of exponents in family-name distributions »

BIBLIOGRAPHIE

- BAECK, Seung Ki, TIET, Hoang Anh Tuan et KIM, Beom Jun, « Family name distributions: Master equation approach », *Physical Review E*, 76, 2007, 046113 pp. 1-7
- COLANTONIO, Sonia E., LASKER, Gabriel W., KAPLAN, Bernice A. et FUSTER, Vicente, « Use of Surname Models in Human Population Biology: A Review of Recent Developments », *Human Biology*, 75, 2003, pp. 785-807.
- CROW, James F. et MANGE, Arthur P., « Measurement of inbreeding from the frequency of marriages between persons of the same surname », *Eugenics Quarterly*, 12, 1965, pp.199-203.
- DARLU, Pierre, DEGIOANNI, Anna et RUFFIE, Jacques, « Quelques statistiques sur la distribution des patronymes en France », *Population*, 52 (3), 1997, pp. 607-634.
- DARLU, Pierre et RUFFIE, Jacques, « L'immigration dans les départements français étudiée par la méthode des patronymes », *Population*, 47 (3), 1992, pp. 719-734.
- DARLU, Pierre et RUFFIÉ, Jacques, « Relationships between consanguinity and migration rate from surname distributions and isonymy in France », *Annals of Human Biology*, 19, 1992, pp.133-137.
- DARWIN, George H., « Marriages between first cousins in England and their effects », *Journal of the Statistical Society*, 38, 1875, pp.153-184.
- DE LUCA Andrea et ROSSI, Paolo, « Renormalization group evaluation of exponents in family-name distributions », *Physica A*, 2009, (à paraître)
- DEGIOANNI Anna, LISA, Antonella, ZEI, Gianna et DARLU, Pierre, « Patronymes italiens et migration italienne en France », *Population*, 51 (6), 1996, pp. 1153-1180.
- DERRIDA, Bernard, MANRUBIA, Susanna C. et ZANETTE, Damián H., « Statistical Properties of Genealogical Tree », *Physical Review Letters*, 82, 1999, pp. 1987-1990.
- DERRIDA, Bernard, MANRUBIA, Susanna C. et ZANETTE, Damián H., « Distribution of repetitions of ancestors in genealogical trees », *Physica A*, 281 (2000) pp. 1-16.
- Fischer, Ronald A., CORBET A. Steven et WILLIAMS, Carrington B., « The Relation Between the Number of Species and the Number of Individuals in a random sample of an Animal Population », *Journal of Animal Ecology*, 12, 1943, pp. 42-58.
- FOX, Wendy R. et LASKER Gabriel W., « The Distribution of Surname Frequencies », *International Statistical Review*, 51, 1983, pp. 81-87.
- GALTON, Francis et WATSON, Henry W., « On the Probability of the Extinction of Families », *Journal of the Anthropological Institute of Great Britain and Ireland*, 4, 1874, pp.138-144.
- JOBLING, Mark A., « In the name of the father: surnames and genetics », *Trends in Genetics*, 17, 2001, pp. 353-357.
- KARLIN, Samuel et MC GREGOR, James, « The number of mutant forms maintained in a population », *Proceedings of the 5th Berkeley Symposium on Mathematics, Statistics and Probability*, 4, University of California Press, 1967, pp. 415-438.
- LASKER, Gabriel W., « A coefficient of relationship by isonymy: A Method for Estimating the genetic Relationship between Populations », *Human Biology*, 49, 1977, pp. 489-493.
- LASKER Gabriel W., *Surnames and Genetic Structure*, Cambridge University Press, 1985.
- LASKER, Gabriel W., « Datos sobre los apellidos hispanoamericanos en los estudios de biología humana », *Anales de Antropología (Mexico)*, 28, 1993, pp.107-128.
- PIAZZA, Alberto, CAPPELLO, Nazario, OLIVETTI, Eleonora et RENDINE, Sabina, « A genetic history of Italy », *Annals of Human Genetics*, 52, 1988, pp. 203-213.
- PIAZZA, Alberto, RENDINE, Sabina, ZEI, Gianna, MORONI, Antonio et CAVALLI-SFORZA, Luigi Luca, « Migration rates of human populations from surname distributions », *Nature*, 329, 1987, pp. 714-716 .
- PINTO-CISTERNAS, Juan, PINEDA, L. et BARRAI, Italo, « Estimation of Inbreeding from Isonymy in Iberoamerican Populations: an Extension of the Method of Crow and Mange », *American Journal of Human Genetics*, 37, 1985, pp. 373-385.

- PORQUERES I GENE, Enric, *Lourde alliance. Mariage et identité chez les descendants des juifs convertis de Majorque (1435-1750)*, Paris, Editions Kimé, 1995.
- REED, William J. et HUGHES, Barry D., « On the distribution of family names », *Physica A*, 319, 2003, pp. 579-590
- SCAPOLI, Chiara, MAMOLINI, Elisabetta, CARRIERI, Alberto, RODRÍGUEZ-LARRALDE, Alvaro et BARRAI, Italo, « Surnames in Western Europe: A comparison of the subcontinental populations through isonymy », *Theoretical Population Biology*, 71, 2007, pp. 37-48.
- SCAPOLI, Chiara, RODRÍGUEZ-LARRALDE, Alvaro, BERETTA, M., NESTI, C., LUCCHETTI, A. et BARRAI, Italo, « Correlations between Isonymy Parameters », *International Journal of Anthropology*, 12, 1997, pp. 17-37.
- B. SYKES Bryan et IRVEN, Catherine, « Surnames and the Y Chromosome », *American Journal of Human Genetics*, 66, 2000, pp.1417-1419.
- WIJMAN, Ellen M., ZEI, Gianna, MORONI, Antonio et CAVALLI-SFORZA, Luigi Luca, « Surnames in Sardinia II. Computation of migration matrices from surname distributions in different periods », *Annals of Human Genetics*, 48, 1984, pp. 65-78.
- YASUDA, N., CAVALLI-SFORZA, Luigi Luca, SKOLNICK, Mark H. et MORONI, Antonio, « The Evolution of Surnames: An Analysis of Their Distribution and Extinction », *Theoretical Population Biology*, 5, 1974, pp. 123-142 .
- ZANETTE, Damián H. et MANRUBIA, Susanna C., « Vertical transmission of culture and distribution of family names », *Physica A*, 295, 2001, pp. 1-8
- ZEI, Gianna, GUGLIELMINO, C.Rosalba, SIRI, Enzo, MORONI, Antonio et CAVALLI-SFORZA Luigi Luca, « Surnames as Neutral Alleles: Observations in Sardinia », *Human Biology*, 55, 1983, pp. 357-365.
- ZEI, Gianna, GUGLIELMINO MATESSI, Rosalba., SIRI, Enzo, MORONI, Antonio et CAVALLI-SFORZA Luigi L., « Surnames in Sardinia I. Fit of frequency distributions for neutral alleles and genetic population structure », *Annals of Human Genetics*, 47, 1983, pp. 329-352.
- ZEI, Gianna, PIAZZA, Alberto, MORONI, Antonio et CAVALLI-SFORZA, Luigi Luca, « Surnames in Sardinia III. The spatial distribution of surnames for testing neutrality of genes », *Annals of Human Genetics*, 50, 1986, pp. 169-180.