

Distances génétiques et distances linguistiques en Europe

Les différences génétiques entre populations d'une même espèce peuvent être évaluées en termes de « distances génétiques ». Ces distances génétiques sont en relation avec les distances géographiques qui séparent les populations. A l'échelle d'un continent, l'Europe par exemple, intervient en outre un facteur historique. Des populations venues d'ailleurs, dotées de marqueurs et de langages spécifiques, vont occuper un espace déterminé, dont la comparaison avec les espaces voisins doit permettre de reconstituer l'histoire des migrations. Cependant les immigrés ne restent pas isolés, et des métissages modifient aussi bien les caractéristiques génétiques que les mœurs et le langage. Dans ces conditions, les distances génétiques sont-elles liées autant, ou davantage, aux différences de langage qu'à l'éloignement géographique ?

Un travail considérable a été accompli par Stokal [1, 2] en colligeant les données fournies par l'étude de 26 systèmes génétiques (antigènes de groupes sanguins, enzymes, protéines — il n'y a pas encore assez de données sur l'ADN pour de telles comparaisons) provenant de 3 369 localités européennes. On ne peut, dans une telle étude, éviter tous les biais : les données émanent d'auteurs différents, on ne dispose pas de tous les systèmes pour chaque localité... Pour les minimiser, on évalue séparément les résultats de chaque système avant de les intégrer dans un ensemble. On obtient ainsi une carte des distances génétiques, que l'on va corrélérer aux distances géographiques, puis aux distances linguistiques.

Les douze familles de langages utilisés en Europe se répartissent en cinq phyla : (a) Indo-européen, le plus important, qui comprend les rameaux : Albanais, Balte, Celte, Germanique, Grec, Roman (Romanche), Slave ; (b) Finno-Ougrien : Finnois et Ougrien (Hon-

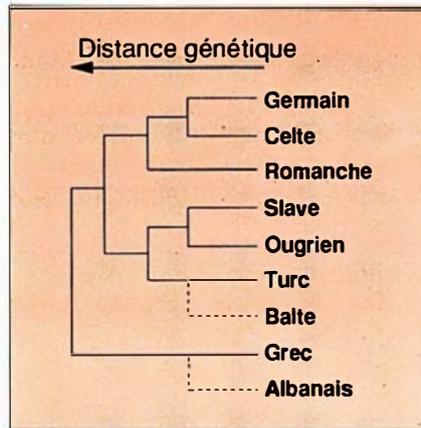


Figure 1. **Distances génétiques moyennes entre les familles de langages.** Balte et Albanais ont été représentés en pointillé en raison de l'incertitude due au trop petit nombre de mesures les concernant. Finnois, Basque et Sémitique, extérieurs à cet ensemble, n'ont pas été représentés. (Modifié d'après [2].)

grois) ; (c) Altaïque : Turc ; (d) Sémitique : Maltais ; (e) un isolat, le Basque.

On peut construire une matrice de distance linguistique en affectant la comparaison entre deux familles de phyla différents d'un coefficient plus élevé qu'à l'intérieur d'un même phylum (figure 1).

Une difficulté propre à ce type de recherche est que le nombre de données n'est pas le même pour toutes les familles. En particulier, le petit nombre de systèmes utilisables pour les familles sémitique, balte et albanaise impose d'être très prudent dans les conclusions qui les concernent. Deux types de résultats peuvent être tirés de ce travail.

(1) Si l'on compare les distances génétiques entre les neuf familles de langage pour lesquelles on dispose de données suffisantes, on voit, d'une part, un groupe Germano-Celte rejoint ensuite par le Romanche, d'autre part un groupe Slavo-Ougrien rejoint par le Turc ; le Grec ne rejoint que plus tard, et Finnois

et Basque restent extérieurs à l'ensemble. Si l'on prend en compte les trois autres familles, l'Albanais est proche du Grec, le Balte du Turc, le Sémitique reste en dehors. L'ensemble est schématisé sur la figure 1.

(2) Les distances génétiques montrent une corrélation significative avec les distances géographiques mais non avec le langage. Si on groupait les familles par leurs affinités linguistiques, les plus proches seraient, parmi les langages indo-européens, Balte et Slave, et parmi les autres, Finnois et Ougrien. Or on trouve des affinités génétiques plus grandes entre, par exemple, Turcs et Ougriens, ou Slaves et Ougriens, qu'entre Celtes et Grecs, ou Finnois et Ougriens, qui appartiennent au même phylum. Il faut dire que dans ce dernier cas il existe un biais, du fait que des Lapons parlant Finnois sont inclus dans l'échantillon.

La conclusion qui se dégage de ce travail est que dans l'ensemble la distance génétique est essentiellement régie par la géographie, même si des barrières linguistiques ont été maintenues entre peuples voisins. On en veut pour preuve les observations que les distances génétiques les plus faibles sont en général obtenues pour des populations voisines, comme Celtes et Germains, Grecs et Albanais. Quelques exceptions doivent cependant être mentionnées : les plus marquantes sont celles des Finnois (peut-être, comme nous l'avons dit, du fait de l'inclusion des lapons) et des Basques, longtemps demeurés une enclave isolée [3].

J.-C. D.

1. Sokal RR. Genetic, geographic and linguistic distances in Europe. *Proc Natl Acad Sci USA* 1988 ; 85 : 1722-6.

2. Harding RM, Sokal RR. Classification of the European language families by genetic distances. *Proc Natl Acad Sci USA* 1988 ; 85 : 9370-2.

3. Allières J. Les Basques. Paris : Presses Univ France, 1986.