

Etude des parentés génétiques dans les populations humaines anciennes : estimation de la fiabilité et de l'efficacité des méthodes d'analyse

L'étude des parentés génétiques permet à l'anthropologie d'identifier la place du sujet au sein des différentes structures dans lesquelles il évolue : l'individu est membre d'une famille biologique, d'un groupe social et d'une population. L'application des méthodes probabilistes classiques (établies pour répondre à des problématiques de médecine légale, comme la méthode des Likelihood Ratios (LR) ou « Rapports de vraisemblance ») aux données STR issues du matériel archéologique a permis la découverte de nombreux liens de parenté, qui ensemble constituent des généalogies parfois complexes. Notre pratique prolongée de ces méthodes nous a cependant amenés à identifier certaines limites de l'interprétation des données STR, en particulier dans les cas de parentés complexes, distantes ou consanguines, ou dans des populations isolées, méconnues ou disparues. Ce travail de thèse s'attache en premier lieu à quantifier la fiabilité et l'efficacité de la méthode des LR dans quatre situations : une population moderne avec une grande diversité allélique, une population moderne avec une faible diversité allélique, une population ancienne de grande taille et une population ancienne de petite taille.

Les publications récentes font usage des marqueurs plus nombreux issus des nouvelles technologies de séquençage (NGS) pour mettre en place de nouvelles stratégies de détection des parentés, basées en particulier sur l'analyse des segments chromosomiques partagés par ascendance entre les individus (segments IBD). Ces méthodes ont rendu possible l'estimation plus fiable de probabilités de parenté dans le matériel ancien. Elles sont néanmoins inadaptées à certaines situations caractéristiques de la génétique des parentés archéologiques : elles ne sont pas conçues pour fonctionner avec une seule paire isolée d'individus et reposent, comme les méthodes classiques, sur l'estimation de la diversité allélique dans la population. Nous proposons donc une quantification de la fiabilité et de l'efficacité de la méthode des segments partagés à partir de données NGS, en s'attachant à déterminer la qualité des résultats dans les différentes situations qui correspondent à des tailles de population plus ou moins importantes et à une hétérogénéité plus ou moins grande de l'échantillonnage.

Ce travail de thèse s'est également attaché à proposer une méthode adaptée aux caractéristiques particulières de l'ADN ancien (petit nombre d'individus, populations mal définies, erreurs techniques...). Une méthode consacrée à la détermination des parentés pour les paires isolées d'individus dont la variabilité allélique est inconnue est donc proposée et soumise aux mêmes mesures de fiabilité et d'efficacité que la méthode des LR et la méthode des segments IBD.

La fiabilité et l'efficacité de chaque méthode ont pu être estimées à partir d'ADN issu de 727 individus modernes, pour lesquels plus de 1000 relations de parenté étaient déclarées et pour lesquels différents types de marqueurs génétiques (STR et SNP) ont été étudiés. Ces méthodes ont ensuite été appliquées à 186 individus anciens issus de fouilles archéologiques, afin d'établir leurs généalogies et de proposer des solutions à l'étude des parentés complexes ou distantes dans le contexte archéogénétique.

Résumé vulgarisé pour le grand public

L'étude de la parenté génétique a toujours été d'un grand intérêt en médecine légale, qui y voit l'opportunité de confirmer des paternités, d'identifier les victimes de catastrophes ou de crimes, voire, dans certains cas particuliers, de participer à l'identification de leurs auteurs. Les méthodes de détermination génétique des parentés ont donc été conçues pour répondre à ces problématiques judiciaires avant d'être appliquées à l'étude de l'ADN ancien des individus provenant de sites archéologiques vieux de centaines ou de milliers d'années. Avec l'apparition de nouvelles méthodes, et alors que l'application des méthodes classiques de détermination des parentés a permis la découverte d'informations sur les généalogies anciennes, il devient nécessaire d'estimer la qualité des déterminations par des études empiriques. Ce travail de thèse utilise plus de 1000 liens biologiques identifiés chez des populations modernes et applique trois méthodes de détermination des parentés pour estimer leur fiabilité et leur efficacité et proposer des solutions spécifiques à l'étude de l'ADN ancien.