

Chroniques génomiques

Sous le capot des sosies

Bertrand Jordan

Vrais jumeaux, faux jumeaux et sosies

On sait depuis longtemps que les traits d'un visage humain sont en grande partie génétiquement déterminés [1], comme en témoigne la très forte ressemblance entre jumeaux monozygotes même lorsqu'ils ont été séparés à la naissance. Les études menées depuis une vingtaine d'années ont permis d'identifier plusieurs centaines de locus impliqués dans l'architecture du visage [2, 3] (→), et aussi de montrer que certaines dissemblances phénotypiques entre jumeaux monozygotes sont liées à des différences au niveau épigénétique [4]. Mais il arrive (très rarement) que deux individus qui ne sont ni jumeaux ni même apparentés présentent une ressemblance extrêmement forte (Figure 1). Un artiste et photographe

(→) Voir la Chronique
génomique de
B. Jordan, m/s n° 3,
mars 2021, page 300



Biologiste, généticien et immunologiste, Président d'Aprògène (Association pour la promotion de la Génomique), 13007 Marseille, France.
brjordan@orange.fr

Québécois, François Brunelle, mène depuis 1999 un projet intitulé « Je ne suis pas un sosie » pour lequel il recherche de tels sosies non apparentés. Il en a trouvé (sur la base du bouche-à-oreille et du volontariat) plus de 250 dans une trentaine de villes de part le monde, et a réalisé sur ce thème un documentaire et plusieurs expositions¹.

Ce travail a attiré l'attention de plusieurs scientifiques et en particulier de l'éditeur de la revue *Epigenetics*, Manel Esteller, qui travaille à Barcelone sur le rôle des modifications épigénétiques chez l'homme et en particulier chez les jumeaux. Avec plusieurs laboratoires de cette ville, Esteller a lancé une étude portant sur le génome, l'épigénome et le microbiome d'une série de couples de sosies, dont les résultats publiés à l'été 2022 [5] font l'objet de cette chronique.

Définition d'un jeu de sosies

Les chercheurs sont partis d'un jeu de 32 couples de sosies ayant donné leur accord pour l'étude et pour la fourniture d'un prélèvement de salive. Les analyses ont été effectuées sur les photos en noir et blanc des participants, et par *microarray* sur un large jeu de variants dans l'ADN (plus de quatre millions), sur le patron de méthylation de cet ADN pour près d'un million de sites CpG, et sur le microbiome présent dans l'échantillon salivaire. Ils ont d'abord cherché à définir de manière objective la ressemblance des sosies, et ont employé pour cela les trois systèmes de reconnaissance faciale les plus performants : *Custom-Net*, un logiciel commercial de reconnaissance faciale développé par une

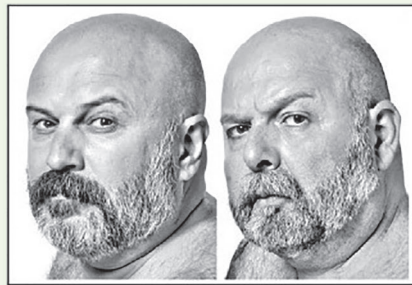


Figure 1. Deux couples de sosies (look-alike) de la collection de François Brunelle (extrait partiel et modifié de la figure 1 de [5]).

¹ <http://www.francoisbrunelle.com/webn/e-project.html>

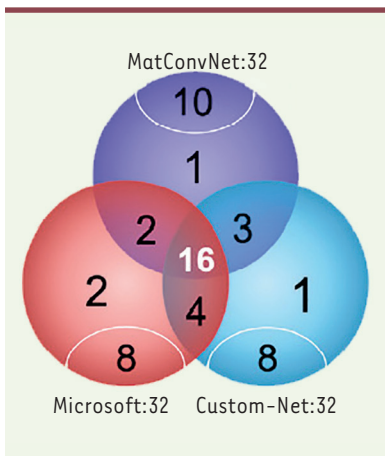


Figure 2. Diagramme de Venn montrant l'identification des couples de sosies par les trois systèmes de reconnaissance utilisés. Seize couples sont reconnus par les trois systèmes et sont considérés comme des « super-sosies ». Les couples non reconnus par un algorithme donné sont indiqués dans la partie du cercle correspondant isolée par une ligne blanche. (extrait partiel et remanié de la figure 1 de [5]).

Ressemblances génétiques entre les super-sosies

Ces 16 super-sosies, identifiés objectivement grâce à l'analyse de leurs portraits par trois systèmes différents, peuvent maintenant être examinés au niveau du génome, de l'épigénome et du microbiome grâce aux prélèvements biologiques autorisés par les sujets, qui ont également répondu à un questionnaire détaillé sur leurs caractéristiques biométriques et leur style de vie (voir plus loin). L'examen des profils de SNP (*single nucleotide polymorphism*) grâce à un système de recherche de parenté [7]⁵ a permis d'éliminer toute lien de parenté même lointain entre les membres de chaque couple. L'analyse globale des quatre millions de SNP montre de très nombreuses différences, conformément à ce que l'on observe pour deux personnes prises au hasard, mais la majorité des couples de sosies (9 sur 16) partagent les mêmes allèles pour une série de près de vingt mille SNP touchant 3730 gènes : les deux allèles sont identiques au sein d'un couple pour chacun de ces SNP, éventuellement différents mais également identiques au sein d'un autre couple. On peut donc imaginer que les gènes auxquels appartiennent ces SNP influencent la forme du visage et soient collectivement responsables de la très forte ressemblance observée.

entreprise locale², *MatConvNet*, un système de réseaux de neurones mis au point à l'université d'Oxford [6], et *Microsoft (Microsoft Oxford Face Project API³)*, un système de reconnaissance faciale généraliste. L'analyse des photos par ces trois algorithmes donne des résultats assez – mais pas totalement – cohérents (Figure 2) : 16 couples (sur 32) sont identifiés par les trois systèmes, et 25 par au moins deux.

La suite des analyses a, logiquement, privilégié les 16 « super-sosies ». La comparaison des scores de similitude avec ceux d'un jeu de 50 jumeaux monozygotes⁴ et de 50 individus pris au hasard est montrée sur la Figure 3. On voit que dans chaque cas la similitude des sosies est proche de celle des vrais jumeaux, et que l'algorithme de *Microsoft* semble être le plus performant.

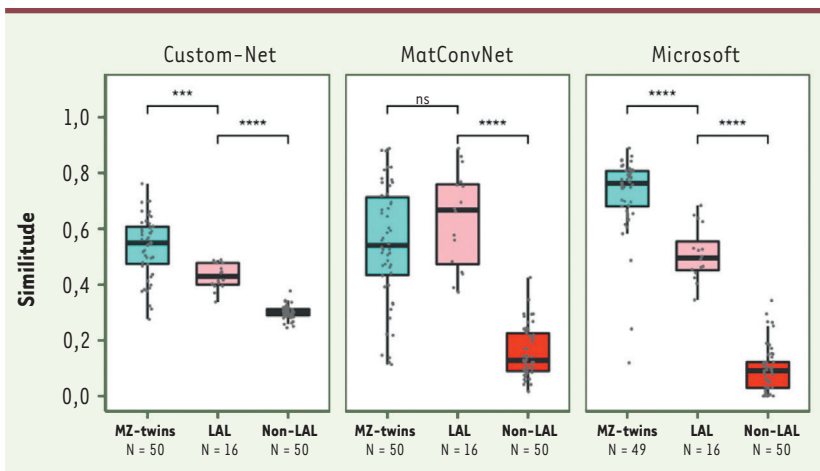


Figure 3. Performance des trois systèmes de classification pour une série de 50 vrais jumeaux (MZ-twins), de 16 super-sosies (LAL, look-alike), et de 50 individus pris au hasard (Non-LAL). Le score de similitude va de 0 (aucune ressemblance) à 1 (identité) (extrait partiel et modifié de la figure 1 de [5]).

La consultation des bases de données fonctionnelles comme *Gene Ontology*⁶ et du site *Facebase*⁷ spécialisé pour les « gènes du visage » montre qu'environ un quart des gènes ainsi désignés sont déjà connus pour leur implication dans divers aspects de la morphologie faciale. Il est donc clair que ces individus très ressemblants mais non apparentés partagent un nombre important d'allèles de gènes connus pour leur implication dans la morphologie faciale. Parmi les autres gènes se trouvent sans doute des déterminants encore inconnus de cette morphologie. Au niveau épigénétique (méthylation de l'ADN aux sites CpG) et du microbiome oral, les ressemblances entre sosies sont peu significatives, du moins dans les limites de l'étude pratiquée.

² www.hertasecurity.com

³ <https://azure.microsoft.com/es-es/services/cognitive-services/face/>

⁴ <https://cvrl.nd.edu/projects/data/#>

⁵ Ce système repose sur la recherche de longues zones d'ADN portant le même assortiment d'allèles pour tous leurs SNP, qui indiquent un ancêtre commun proche.

⁶ <https://geneontology.org/>

⁷ <https://www.facebase.org/>

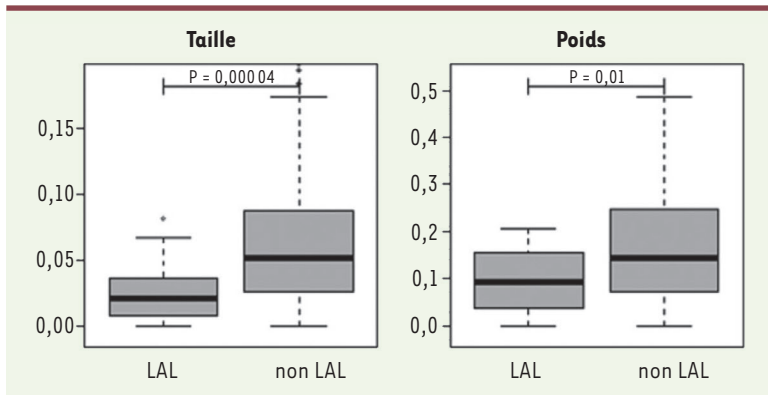


Figure 4. Différences phénotypiques (distances euclidiennes dans une représentation spatiale des données) observées entre sosies (LAL) et non sosies (non LAL) pour la taille et le poids. Les sosies sont un peu plus proches que les non-sosies (extrait partiel et remanié de la figure 4 de [5]).

Sosies et comportement

La dernière partie de l'article se Joshi *et al.* [5] est consacrée à l'étude du comportement (au sens large) des sosies, afin d'examiner si les ressemblances observées du point de vue physique et génétique se retrouvent à ce niveau. À cet effet, les participants à l'étude ont répondu à un questionnaire très détaillé (68 items, depuis le poids jusqu'au tabagisme en passant par la présence ou non d'animaux de compagnie [5]). L'analyse montre une certaine ressemblance entre les sosies (Figure 4), mais elle est beaucoup moins frappante que celle que l'on observe généralement [8] entre couples de jumeaux monozygotes.

Ce résultat est assez logique : nos couples de sosies ont été sélectionnés sur la base de leur ressemblance, on a vu que cela implique des similitudes au niveau des gènes impliqués dans la forme du visage, mais tout le reste du génome peut être totalement différent, tout comme l'histoire personnelle de individus. On est donc très loin du cas des jumeaux monozygotes qui portent les mêmes allèles sur tous leurs gènes.

Un résultat « évident » mais utile

La recherche de sosies menée depuis plus de vingt ans par François Brunelle a ainsi sélectionné des couples chez lesquels l'assortiment d'allèles impliqués dans le visage est le même (dans chaque couple) mais où tout le reste du génome est différent. C'est un résultat assez logique même s'il est plutôt contre-intuitif : on a tendance à oublier que notre Terre porte plus de huit milliards d'individus, et qu'avec un tel effectif des coïncidences statistiquement très peu probables peuvent avoir lieu. C'est aussi un résultat utile puisque chacun des gènes identifiés par leur ressemblance chez les couples de sosies devient un gène-candidat pour le déterminisme de la forme du visage. Il serait intéressant de pratiquer une telle étude sur un échan-

tillon plus large comme les 500 000 personnes de la *UK Biobank*⁸, pour lesquelles de très nombreuses informations phénotypiques sont déjà archivées ainsi que des séquences génomiques complètes : on pourrait ainsi évaluer la fréquence des sosies, mieux caractériser les gènes responsables de leurs ressemblances, et ainsi avancer dans la compréhension des déterminants génétiques de notre apparence. ♦

SUMMARY

What “look-alike” individuals tell us

Rare unrelated individuals show striking resemblance, almost as high as monozygotic twins. Extensive genetic analysis of a set of such individuals shows that every couple shares a given allele at a large set of loci, enriched for “face genes”. The similarity is limited to this set of loci, and the rest of the genome is quite diverse in these “look alike” individuals. ♦

LIENS D'INTÉRÊT

L'auteur déclare n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

RÉFÉRENCES

- Weinberg SM, Parsons TE, Marazita ML, Maher BS. Heritability of Face Shape in Twins: A Preliminary Study using 3D Stereophotogrammetry and Geometric Morphometrics. *Dentistry* 2013 ; 1 : 14.
- White JD, Indencleef K, Naqvi S, *et al.* Insights into the genetic architecture of the human face. *Nat Genet.* 2021 ; 53 : 45-53.
- Jordan B. Les gènes du visage. *Med Sci (Paris)* 2021 ; 37 : 300-3.
- Fraga MF, Ballestar E, Paz MF, *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci USA* 2005 ; 102 : 10604-9.
- Joshi RS, Rigau M, García-Prieto CA, *et al.* Look-alike humans identified by facial recognition algorithms show genetic similarities. *Cell Rep* 2022 ; 40 : 111257.
- Vedaldi A, Lenc K. MatConvNet : Convolutional Neural Networks for MATLAB. *Proceedings of the 23rd ACM International Conference on Multimedia (MM '15) 2015*, 689-92. <https://doi.org/10.1145/2733373.2807412>
- Manichaikul A, Mychaleckyj JC, Rich, SS, *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010 ; 26 : 2867-73.
- Hagenbeek FA, Hirzinger JS, Breunig, S, *et al.* Maximizing the value of twin studies in health and behaviour. *Nat Hum Behav* 2023 ; 7 : 849-60.

TIRÉS À PART

B. Jordan

⁸ <https://www.ukbiobank.ac.uk/>