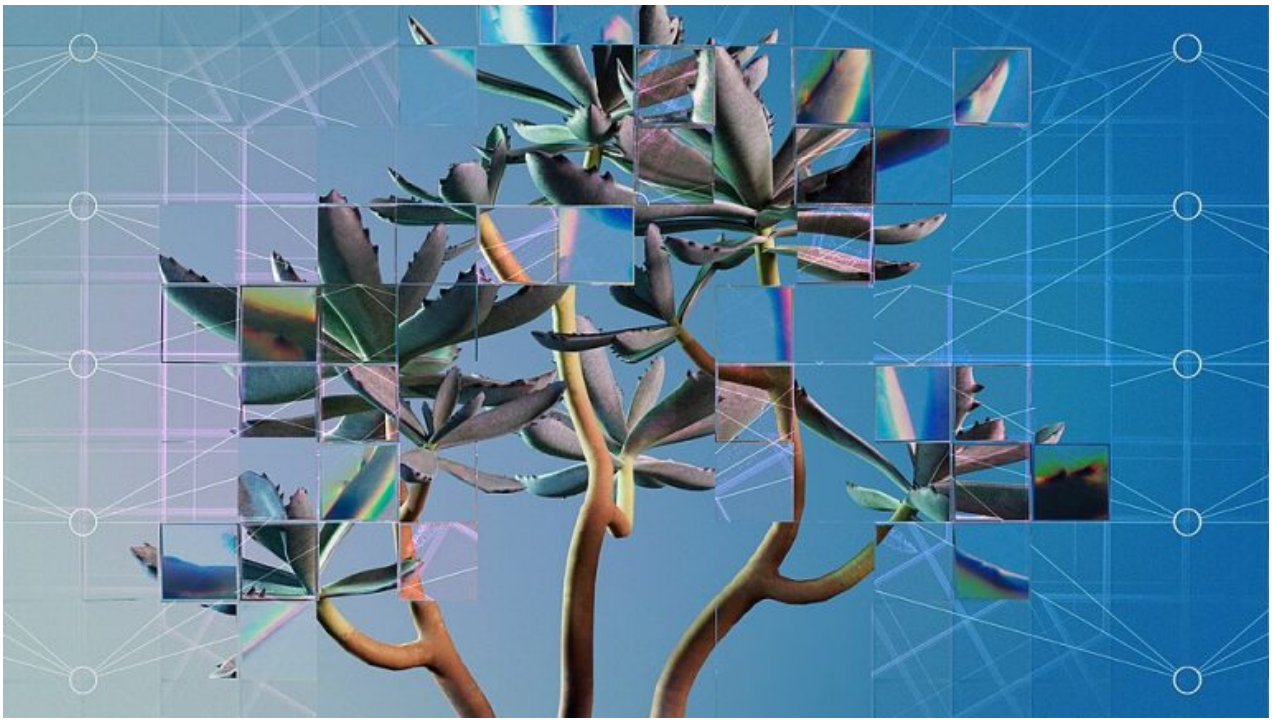
Veille citoyenne d'information sur les OGM et les semences

Quand les algorithmes décident des modifications génétiques du vivant

Par Eric MEUNIER

Publié le 29/05/2025, modifié le 17/03/2026

Depuis de nombreuses années, les multinationales collectent un nombre croissant d'informations de séquences génétiques et protéiques ou d'informations épigénétiques. Elles réduisent le vivant à quelques données compilées dans des bases de données numériques. Grace aux algorithmes de l'« *intelligence artificielle* », elles prétendent disposer d'outils déterminant quelles modifications génétiques permettraient d'obtenir une nouvelle caractéristique donnée. Dans une société où techniques de modification génétique et brevets sont intimement liés, ces algorithmes vont surtout accélérer les prétentions à s'approprier le vivant.



Dans le domaine des technologies génétiques, la numérisation des êtres vivants et le traitement informatique des données qui en sont issues sont en pleine croissance. En utilisant des séquences de génomes présents dans la biodiversité, les détenteurs d'algorithmes de l'« *intelligence artificielle* » clament maintenant pouvoir prédire les modifications génétiques nécessaires pour de nouvelles caractéristiques.

Une lecture encore partielle d'un grand nombre de génome

En 2023, un article rapportait comment des chercheurs du Wellcome Sanger Institute (Royaume-Uni) avaient « *nourri* » des usines d'algorithmes pour leur faire prédire les « *meilleures réparations face à un défaut génétique* »[i]. Les chercheurs ont établi 3 064 séquences d'ADN de longueur variable à insérer dans un génome. Après avoir mis en œuvre une méthode d'insertion de chacune de ces séquences, ils ont séquencé le génome modifié obtenu pour voir si la modification génétique avait été bien réalisée. Avec leurs résultats, ils ont « *alimenté* » un algorithme utilisé en « *apprentissage machine* » (*machine learning* en anglais) afin que celui-ci « *détecte les profils permettant de déterminer les succès d'insertions comme la longueur ou le type de réparation d'ADN mis en jeu* ».

D'autres vont plus loin dans les ambitions qu'ils font porter à de tels algorithmes. En 2024, *Science* publiait un article[ii] rendant compte du travail de chercheurs étasuniens pour mettre au point un modèle algorithmique d'« *intelligence artificielle* » (IA). Selon cette publication, cet algorithme peut « *interpréter et générer des séquences génomiques à grande échelle* ». Nommé Evo, il serait capable d'analyser « *des millions de génomes microbiens* », puis de « *générer efficacement de longues séquences (d'ADN)* ». Sans modestie aucune, les auteurs expliquent que l'algorithme Evo « *a développé une compréhension globale du code génétique complexe de la Vie, des nucléotides de base de l'ADN aux génomes entiers* ». Selon les chercheurs, Evo serait donc capable de prédire comment de « *petits changements d'ADN* » peuvent modifier les caractéristiques d'organismes, tout comme il serait capable de « *générer de longues séquences de la longueur d'un génome et élaborer de nouveau système biologique* », si tant est qu'un système biologique ne se définit que par ses séquences génétiques.

Selon un rapport de l'association allemande Save Our Seeds, publié en janvier 2025[iii], cette dynamique consistant à nourrir des algorithmes avec des informations de séquences génétiques est partagée par plusieurs entreprises. Parmi les exemples cités, on trouve celui du modèle algorithmique mis au point par l'entreprise Inari, nommé FloraBERT. Selon le rapport, l'algorithme a été alimenté avec les séquences régulatrices de 93 espèces végétales et 25 variétés de maïs différentes. L'utilisation attendue de cet algorithme est qu'il puisse prévoir comment des modifications génétiques dans des séquences régulatrices du génome du maïs peuvent modifier les caractéristiques de cette plante. Un partenariat entre Google et Instadeep a également débouché sur un algorithme finalisé fin 2023, AgroNT (pour *Agronomic Nucleotide Transformer*). Alimenté avec des séquences génétiques de 48 espèces végétales, l'algorithme aurait été utilisé pour simuler plus de 10 millions de mutations chez le manioc, en y associant une prédiction de modification de caractéristiques.

Combinaison des modifications génétiques et des algorithmes

L'utilisation des algorithmes par les chercheurs vise à établir quelles modifications génétiques dans quelles séquences seraient les plus efficaces pour obtenir telle ou telle caractéristique. Une combinaison qui résulte du travail décrit précédemment, à savoir l'analyse par ces algorithmes d'un nombre croissant de génomes. Mais, comme le souligne le rapport de Save Our Seeds[iv], de tels travaux sont encore récents, puisque apparus après 2022.

Les modifications génétiques envisagées avec l'utilisation des algorithmes concernent, à ce jour, principalement les séquences dites de régulation. Ces séquences régulent l'expression d'autres séquences génétiques selon les signaux intérieurs ou extérieurs reçus. Modifier génétiquement ces séquences de régulation peut viser à éteindre l'expression d'une séquence plutôt que d'enlever cette dernière. D'autres travaux utilisant des algorithmes visent non pas à éteindre

l'expression d'une séquence génétique, mais à moduler son niveau d'expression. L'algorithme iCREPCP de l'Université de Huazhong (Chine) est justement utilisé pour identifier les séquences promotrices et suggérer les modifications génétiques les plus « *prometteuses* »[v]. D'autres séquences sont également ciblées : celles des petits ARN qui sont utilisés dans les complexes Crispr.

Les algorithmes de l' « *intelligence artificielle* » sont également utilisés pour prédire les modifications génétiques qui opéreraient potentiellement une modification des structures de protéines. Des travaux qui portent sur les modifications génétiques à apporter pour augmenter la résistance de tomates au champignon *Phytophthora infestans* ont ainsi été recensés . Ces recherches ont été réalisées avec l'algorithme AlphaFold, qui a conduit à cibler deux briques d'une protéine à modifier pour espérer augmenter cette résistance. D'autres chercheurs ont utilisé le même algorithme pour augmenter la viscosité de pommes de terre. Pour le maïs, il s'agit de modifier l'architecture des plantes pour qu'elles puissent être cultivées plus densément dans les champs. Dans le cas du blé, des protéines sont ciblées pour rendre le blé plus facilement « *travaillable* » en agro-alimentaire...

Des algorithmes pour remplacer les assistants de laboratoire

Plutôt que de rester cantonnés au rôle d'outils de travail, certains travaillent à ce que les algorithmes d'IA deviennent des assistants de laboratoire. Fin 2024, un article de presse[vi] renseignait comment Google DeepMind ou encore BioNTech travaillent à développer des automates intervenant dans le choix des expériences scientifiques à mener, de leurs protocoles à la collecte et analyse des résultats. Pour Karim Beguir, CEO de BioNTech, de tels assistants seraient des « *accélérateur de productivité* » pour les chercheurs.

Des projets d'assistants voient donc le jour pour analyser des séquences génétiques ou prédire la structure de protéines, comme nous l'avons déjà vu. D'autres pourraient aussi être utilisés pour « *élaborer, planifier et exécuter des expériences chimiques complexes* ». L'outil Anthropic's Claude 3.5 pourrait même être utilisé encore plus en amont, cette fois pour générer des idées d'expériences. Une utilisation qui, comme le souligne l'auteur de l'article de 2024[vii], doit encore être confronté à l'utilité réelle de telles propositions d'expériences.

Les multinationales aux premières loges

Bayer, BASF, Syngenta, Corteva… ces multinationales semencières utilisent d'ores et déjà leurs algorithme pour modifier génétiquement des plantes. Save our Seeds[viii] rappelle pertinemment que ces entreprises collectent depuis plusieurs années de grandes quantités de données génétiques, protéiques… justement pour alimenter leurs algorithmes.

Corteva, par exemple, a développé son propre algorithme en utilisant un outil de Google, le logiciel BigBird. Pour obtenir des prédictions de modifications génétiques, des mutations en l'occurrence, à conduire dans des séquences régulatrices, Corteva a fourni à son algorithme les informations de séquences de 14 espèces végétales, dont l'orge, le riz, le blé, le maïs, le canola (un colza) ou encore le soja. Corteva travaille également avec l'entreprise Tropic Bioscience dans l'espoir de pouvoir conduire des modifications génétiques en vue de résistances à des maladies.

De son côté, à l'instar de BASF, Bayer travaille avec Evogene et son algorithme pour définir les modifications génétiques à opérer pour obtenir des résistances à des maladies chez diverses plantes. Save our Seeds ajoute qu'elle a également investi dans deux entreprises qui utilisent des algorithmes et l'outil de modification génétique Crispr/Cas, les entreprises Ukko et Amfora.

Enfin, Syngenta a développé son propre algorithme, que nous avons déjà mentionné, AgroNT, en lien avec InstaDeep et Google. Avec cet outil, l'entreprise cherche maintenant à établir des modifications génétiques à conduire dans le maïs et le soja.

Les algorithmes utilisés prédisent des modifications génétiques uniquement sur base de connaissances de séquences génétiques, sans tenir compte du contexte cellulaire au sein de l'organisme, ni de la diversité des environnements dans lesquels cet organisme serait disséminé. Une approche pour le moins réductrice et dont la fiabilité des résultats dépend des données qu'on lui propose de traiter, mais qui ne constituent en aucun cas toutes les données concernant le vivant. Des données dont la production va se faire à une vitesse croissante et qui, par ailleurs, prennent le pas sur celles issues de l'observation directe du vivant, qui se déroule sur un temps nécessairement plus lent mais plus collaboratif...

i « Prime Editing and Machine Learning Aid Researchers in Determining the Best Fix for Genetic Flaws », *Genetic Engineering & Biotechnology News*, 17 février 2023.

ii Eric Nguyen *et al.*, « Sequence modeling and design from molecular to genome scale with Evo », *Science*, Vol 386, Issue 6723, 15 novembre 2024.

iii B. Vogel, « When chatbots breed new plant varieties », Save Our Seeds, janvier 2025.

iv *Ibid.*

v Deng, K. *et al.*, « iCREPCP: A deep learning-based web server for identifying base-resolution *cis*-regulatory elements within plant core promoters », *Plant Communications*, Volume 4, Issue 1, 9 janvier 2023.

vi Edd Gent, « DeepMind and BioNTech Bet AI Lab Assistants Will Accelerate Science », *SingularityHub*, 7 octobre 2024.

vii *Ibid.*

viii B. Vogel, « When chatbots breed new plant varieties », Save Our Seeds, janvier 2025.

Adresse de cet article : https://infogm.org/quand-les-algorithmes-decident-des-modifications-genetiques-du-vivant/