Veille citoyenne d'information sur les OGM et les semences

When algorithms decide on the genetic modification of living organisms

Par Eric MEUNIER

Publié le 05/06/2025, modifié le 09/03/2026

For many years now, multinationals have been collecting an increasing amount of genetic, proteic sequences and epigenetic informations. They are reducing living organisms to data compiled in digital databases. Using "*artificial intelligence*" algorithms, they claim to have the tools to determine which genetic modifications will produce a given new characteristic. In a society where genetic modification techniques and patents are intimately linked, these algorithms will above all accelerate the claim to own living organisms.



In the field of genetic technologies, the digitisation of living beings and the computer processing of the data derived from them are growing rapidly. Using sequences of genomes present in biodiversity, the owners of "*artificial intelligence*" algorithms now claim to be able to predict the genetic modifications required for new characteristics.

A still partial reading of a large number of genomes

In 2023, an article reported how researchers at the Wellcome Sanger Institute (UK) had fed algorithms factories to make them predict "*the best fix for a given genetic flaw*"[i]. The researchers established 3,064 DNA sequences of variable length to be inserted into a genome. After implementing a method for inserting each of these sequences, they sequenced the modified genome obtained to see whether the genetic modification had been carried out correctly. Using their results, they fed an algorithm used in '*machine learning*' so that it could "*detect patterns that determine insertion success, such as length and the type of DNA repair involved*".

Others go further in their ambitions for such algorithms. In 2024, Science published an article[ii] reporting on the work of US researchers to develop an algorithmic model of "*artificial intelligence*" (AI). According to the article, this algorithm can "*interpret and generate large-scale genomic sequences*". Called Evo, it is capable of analysing "*millions of microbial genomes*" and then "*generate realistic genome-length sequences*". Without any modesty, the authors explain that the Evo algorithm "*has developed a comprehensive understanding of life's complex genetic code, from individual DNA bases to entire genomes*". According to the researchers, Evo is therefore capable of predicting how small changes in DNA could modify the characteristics of organisms, just as it can "*generate realistic genome-length sequences, and design new biological systems*", if a biological system is defined solely by its genetic sequences.

According to a report by the German association Save Our Seeds, published in January 2025[iii], this dynamic of feeding algorithms with genetic sequence information is shared by several companies. Among the given examples is the algorithmic model developed by the company Inari, called FloraBERT. According to the report, the algorithm was fed with the regulatory sequences of 93 plant species and 25 different maize varieties. The expected use of this algorithm is that it will be able to predict how genetic modifications to regulatory sequences in the maize genome may alter the characteristics of this plant. A partnership between Google and Instadeep has also resulted in an algorithm, AgroNT (for Agronomic Nucleotide Transformer), finalised at the end of 2023. Fed with genetic sequences from 48 plant species, the algorithm is said to have been used to simulate more than 10 million mutations in cassava, with a prediction of changes in characteristics.

Combining genetic modifications and algorithms

The use of algorithms by researchers aims to establish which genetic modifications in which sequences would be the most effective in obtaining a given characteristic. This combination is the result of the work described above, namely the analysis by these algorithms of a growing number of genomes. But, as the Save Our Seeds report[iv] points out, such work is still in its infancy, having been carried out after 2022.

To date, the genetic modifications envisaged with the use of algorithms mainly concern so-called regulatory sequences. These sequences regulate the expression of other genetic sequences according to internal or external signals received. Genetically modifying these regulatory sequences can be aimed at turning off the expression of a sequence rather than removing it. Other work using algorithms aims not to switch off the expression of a genetic sequence, but to modulate its level of expression. The iCREPCP algorithm from Huazhong University (China) is being used to identify promising sequences and suggest the most '*promising*' genetic modifications[v]. Other sequences are also targeted: those of the small RNAs used in Crispr complexes.

"*Artificial intelligence*" algorithms are also being used to predict genetic modifications that could potentially alter protein structures. Work on genetic modifications to increase tomato resistance to the fungus Phytophthora infestans has been identified. This research was carried out using the

AlphaFold algorithm, which led to the targeting of two building blocks of a protein to be modified in the hope of increasing resistance. Other researchers have used the same algorithm to increase the viscosity of potatoes. For maize, the aim is to modify the architecture of the plants so that they can be grown more densely in the fields. In the case of wheat, proteins are targeted to make wheat more easily '*workable*' in the food industry...

Algorithms to replace laboratory assistants

Rather than remain confined to the role of work tools, some people are working to ensure that AI algorithms become laboratory assistants. At the end of 2024[vi], a press article reported how Google DeepMind and BioNTech were working on developing automated systems to help choose the scientific experiments to be carried out, from their protocols to the collection and analysis of results. For Karim Beguir, CEO of BioNTech, such assistants would be "*productivity accelerator*" for researchers.

As we have already seen, assistant projects are being developed to analyse genetic sequences or predict the structure of proteins. Others could also be used to "*design, plan, and execute complex chemistry experiments*". Anthropic's Claude 3.5 could even be used further upstream, this time to generate ideas for experiments. As the author of the 2024 article[vii] points out, this use has yet to be tested against the actual usefulness of such experimental proposals.

Multinationals in the front row

Bayer, BASF, Syngenta, Corteva... these multinational seed companies are already using their algorithms to genetically modify plants. Save Our Seeds[viii] points out that these companies have been collecting vast quantities of genetic and protein data for several years, precisely to feed their algorithms.

Corteva, for example, has developed its own algorithm using Google's BigBird software. To obtain predictions of genetic modifications, in this case mutations, to be carried out in regulatory sequences, Corteva has provided its algorithm with sequence information from 14 plant species, including barley, rice, wheat, maize, canola (a rapeseed) and soya. Corteva is also working with Tropic Bioscience in the hope of being able to carry out genetic modifications to resist disease.

For its part, like BASF, Bayer is working with Evogene and its algorithm to define the genetic modifications needed to obtain disease resistance in various plants. Save Our Seeds adds that it has also invested in two companies that use algorithms and the Crispr/Cas genetic modification tool, Ukko and Amfora.

Finally, Syngenta has developed its own algorithm, AgroNT, which we have already mentioned, in conjunction with InstaDeep and Google. With this tool, the company is now seeking to establish genetic modifications to be carried out in maize and soya.

The algorithms used predict genetic modifications solely on the basis of knowledge of genetic sequences, without taking into account the cellular context within the organism, or the diversity of environments in which the organism is disseminated. This approach is reductive, to say the least, and the reliability of its results depends on the data it is proposed to process, which in no way constitutes all the data concerning living organisms. This data is going to be produced at an ever-increasing rate, and is taking precedence over data derived from direct observation of living organisms, which necessarily takes place over a slower but more collaborative period...

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

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

iii B. Vogel, "When chatbots breed new plant varieties", Save Our Seeds, January 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 January 2023.

vi Edd Gent, "DeepMind and BioNTech Bet AI Lab Assistants Will Accelerate Science", *SingularityHub*, 7 October 2024.

vii *Ibid.*

viii B. Vogel, "When chatbots breed new plant varieties", Save Our Seeds, January 2025.

Adresse de cet article : https://infogm.org/en/when-algorithms-decide-on-the-genetic-modification-of-living-organisms/