infOGM

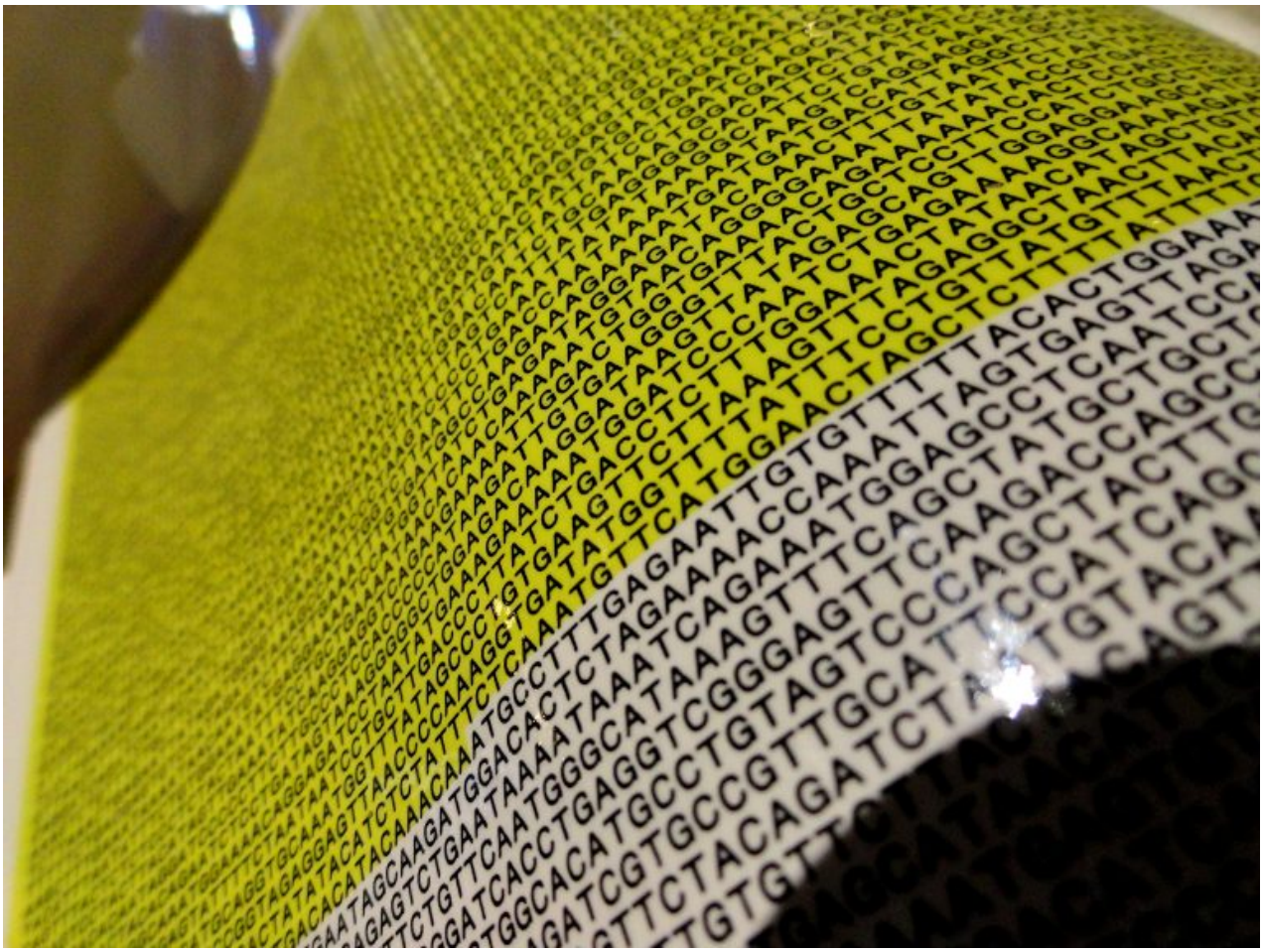
Veille citoyenne d'information sur les OGM et les semences

“Artificial intelligence” to digitize genomes

Par Eric MEUNIER

Publié le 18/04/2025, modifié le 09/03/2026

The digitization of living organisms is the subject of a growing number of projects. Computer data, generated and stored in ever-larger “*data centers*”, are used by “*artificial intelligence*” matrices. These data are of all kinds: genetic sequences, proteins, etc. In these fields, which require increasing natural resources, investments are multiplying.



The use of computer algorithms, brought together in real factories, to exploit ever-increasing amounts of information is one of the topics of the year 2025. Inaccurately named “*artificial intelligence*”, these computer factories require ever-larger computer servers, which in turn consume increasingly problematic water and energy resources. But it is above all their use in the digitization of living organisms that deserves a closer look for those following the GMO dossier. At a time when

multinationals are stepping up their efforts to appropriate living organisms, in what way are these algorithms becoming tools of monopolization?

Exponential growth in the amount of data to be generated and processed

Since 2008, the Earth BioGenome Project aims at sequencing the genomes of the 1.84 million known eukaryotic species (organisms with cells possessing a nucleus), out of the estimated 12 to 15 million[i]. The cost? At least five billion dollars... When you consider that, by December 2024, the genomes of 3,039 species had been sequenced, the initial target of 10 years seems a tall order. But the ambition for sequencing is there, and the amount of data expected is staggering. In a report published in January 2025[ii], the organization Save our seeds gives several telling figures. The sequences of 13.8 million proteins from 342 different plant species are stored in a computer database created by university researchers and named PlantMWpIDB. Another public database, PlantExp, contains the sequences of 131,400 transcriptomes, i.e. all the RNA molecules identified in a genome. Another database, Pmhub, contains descriptions of 188,837 plant metabolites. These figures are given as examples. Having an idea of the total amount of information generated and stored is impossible. For the ABS Biotrade project, the observation is that "*the sequencing of genetic information became such a common practice in all fields of biological sciences that it led to an enormous amount of information being generated and stored every day*"[iii].

From a technical point of view, the generation of this data has been made possible by the lower cost and increasing speed of sequencing methods. As Save our Seeds details in its report, the first sequencing of a complete genome, that of Arabidopsis Thaliana, published in 2000, required 10 years' work at a cost of $100 million. 25 years later, in 2025, sequencing the genome of a plant will take just one week, at a cost of less than $1,000. In addition to these sequencing methods are the so-called "*omics*", which study all types of molecules present in a cell: protein and RNA are two examples. The aim of several companies today is to draw up a map of the molecules produced and present in a cell, a sort of "*Google Map of plants*", as Save our Seeds calls it. In 2021, a conference on the first plant cell atlas brought together 500 representatives from governments, research institutes and companies such as Bayer, Google and Syngenta.

Energy- and water-hungry technical capabilities

To process this growing quantity of data and to fulfil the tasks required, algorithms need hardware, energy and water. The numbers here take on a dimension that is hard to fathom. If we take the example of Elon Musk's company, xAI, which claims to "*accelerate scientific discovery*", its Colossus server, based in Memphis (USA), has requirements that have rarely been met. These servers run on 100,000 processors supplied by Nvidia. These 100,000 processors on the Colossus server require the equivalent of 150 megawatts[iv] /hour to operate, equivalent to the needs of 100,000 homes for the same duration. To obtain this amount of energy, xAI's server is currently powered by methane-fuelled turbines. For the years to come, players in the field are working on mini atomic power plants. In France, EDF has launched a subsidiary, Nuward, to design and build such reactors[v] in 2023.

Energy is not the only problem to be solved. These servers, like all IT installations, need to be cooled continuously. To do this, the water requirements are also staggering. If we look again at the xAI server, it requires over 3.5 million liters of water per day[vi]. Relative to the population of France, for example, this corresponds to the average daily consumption of over 22,300 people[vii]... And in the autumn of 2024, xAI announced plans to double the size of its Colossus server.

GAFAM mobilized

The amount of "*omics*" or sequence data is set to grow exponentially. The Earth BioGenome Project is mobilizing more than 60 projects worldwide, on land and at sea, to take samples of organisms and sequence as many, if not all, of the molecules making up these organisms as possible. The data will then be stored in computerized form on servers. For some years now, the term "*digital sequence information*" (DSI) has been used for all data concerning molecular sequences (DNA, RNA, proteins, etc.). This "digitized" status is at the root of current attempts by multinationals to evade international texts such as the Convention on Biological Diversity (CBD), which aims to protect biodiversity. It also opens the door to the possibility of patents for inventions using ISDs as raw material, but which, once obtained, will be extended by their holders to the physical matter making up biodiversity.

Several cases illustrate that private algorithms are increasingly being used to feed off this digital data, in particular to imagine genetic modifications that multinationals' laboratory researchers would like to implement, as we'll see in a forthcoming article. For several years now, IT companies have been mobilized at the crossroads of algorithms and biotechnologies. A report[viii] published in September 2024 by the African Center for Biodiversity, ETC Group and Third World Network, for example, details that Google, Microsoft, Amazon and Nvidia have invested in and developed projects concerning the Living. For example, Google has set up a company with Gigko Bioworks to create new proteins using its Google DeepMind algorithms. Amazon founder Jeff Bezos, meanwhile, has invested $100 million *via* the Bezos Earth Fund to create new proteins using his algorithms.

Growing investments

In recent years, the development of the computing capacity to run these algorithms has been the focus of a growing search for funding. As the latest example, on February 11, 2025, the European Union launched an initiative to raise 200 billion euros for investment in this field. According to JP Morgan, over $1,000 billion could be invested between 2024 and 2027[ix]. At the crossroads of informatic and biotech, investments are also substantial. While the fields concerned range from molecules used in medicine to genetically modified plants used in agriculture, agri-food, agrofuels... the value of the biotech market was estimated, in 2023, at 1,500 billion dollars[x].

At a time when the European Commission has been proposing to deregulate many GMOs since 2023, the emergence of this (mis)named "*artificial intelligence*" technology appears to be a complete shift in the existing technical paradigm. The very landscape of the multinationals involved in the agri-food chain could continue to be turned upside down in the years to come. As for citizens, their place in this technological choice is, for the moment, that of mere spectator, when they are not themselves suppliers of personal data.

i Eric Meunier, « The genome of 1.8 million species is being sequenced », *Inf'OGM*, 11 December 2024.

ii B. Vogel, « When chatbots breed new plant varieties », Save our seeds, January 2025.

iii ABS Capacity Development Initiative, BioInnovation Africa and ABioSA, « Digital Sequence Information on Genetic Resources ».

iv Magna, « Elon Musk's xAI Receives Approval For 150MW, Allowing 100,000 GPUs To Operate Simultaneously », *AI Secret*, 27 November 2024.

v Kevin Champeau, « Pour accélerer son mini réacteur nucléaire, EDF lance une nouvelle filiale », *Révolution Énergétique*, 10 April 2023 (in French) .

vi Magna, « Elon Musk's xAI Receives Approval For 150MW, Allowing 100,000 GPUs To Operate Simultaneously », *AI Secret*, 27 November 2024.

vii Misitia Ravaloson, « Consommation moyenne eau : estimation par personne et usage », *Selectra*, 9 August 2024 (in french).

viii African Centre for Biodiversity (ACB), Third World Network (TWN) and ETC Group, «'Black Box' Biotechnology – Integration of artificial intelligence with synthetic biology », September 2024.

ix Karen Ward *et al.* , « AI investing: More broadening than bubble », J.P. Morgan, 19 November 2024.

x Monica Johnson, « The Outlook for Global Biotechnology: AI Expected to Drive Market Growth », *International Banker*, 6 December 2024.

Adresse de cet article : https://infogm.org/en/artificial-intelligence-to-digitize-genomes/