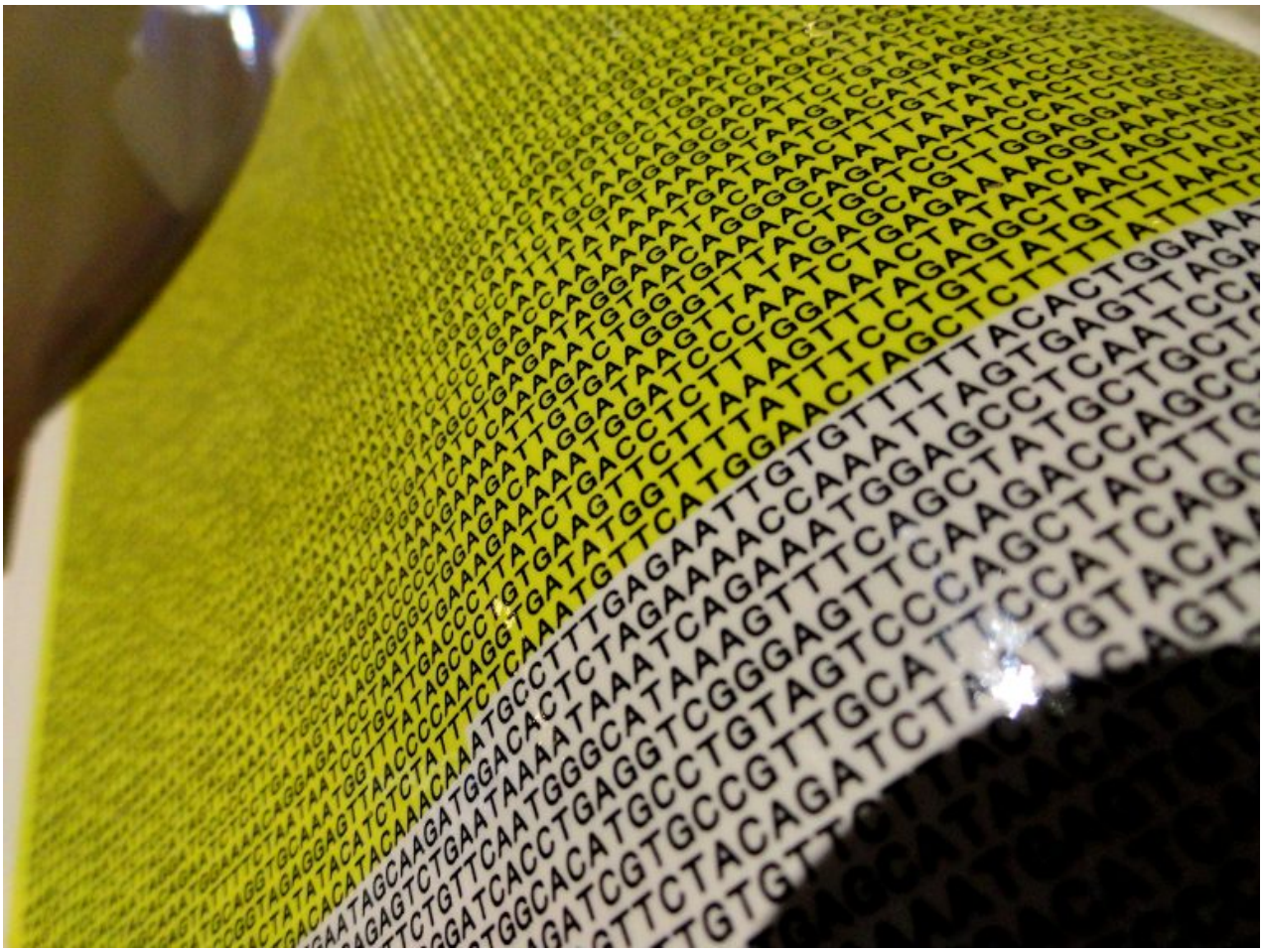
Veille citoyenne d'information sur les OGM et les semences

L’ « intelligence artificielle » au service de la numérisation des génomes

Par Eric MEUNIER

Publié le 08/04/2025, modifié le 25/03/2026

La numérisation du vivant fait l’objet de projets de plus en plus nombreux. Les données informatiques, générées et stockées dans des « *data centers* » (centre de données) de plus en plus grands, sont utilisées par les matrices de l' « *intelligence artificielle* ». Ces données sont de toutes natures : séquences génétiques, protéines, etc. Dans ces domaines nécessitant des ressources naturelles croissantes, les investissements se multiplient.



L’utilisation d’algorithmes informatiques, réunis au sein de véritables usines, afin d’exploiter des informations de plus en plus nombreuses est un des sujets de l’année 2025. Imparfaitement

nommées « *intelligence artificielle* », ces usines informatiques nécessitent des serveurs informatiques de plus en plus grands, entraînant la consommation de ressources en eau et énergie de plus en plus problématiques. Mais c'est surtout leur utilisation dans le domaine de la numérisation du vivant qui mérite d'être regardée de plus près pour celles et ceux qui suivent le dossier des OGM. A l'heure où les multinationales multiplient les manœuvres pour s'approprier le vivant, en quoi ces algorithmes deviennent-ils des outils d'accaparement ?

Une croissance exponentielle de données à générer et traiter

Depuis 2008, le Earth BioGenome Project vise à séquencer le génome des 1,84 million d'espèces eucaryotes (organisme à cellules possédant un noyau) connues, sur les 12 à 15 millions estimées[i]. Le coût ? Minimum cinq milliards de dollars… Quand on sait qu'en décembre 2024, le génome de 3 039 espèces avait été séquencé, l'objectif affiché d'un délai de 10 années fixé au départ paraît être une gageure. Mais l'ambition de séquençage est là et la quantité de données attendues faramineuse. Dans un rapport publié en janvier 2025[ii], l'organisation Save our seeds donne plusieurs chiffres parlant. Les séquences de 13,8 millions de protéines issues de 342 espèces végétales différentes sont enregistrées dans une base de données informatique créée par des chercheurs universitaires et nommée PlantMWpIDB. Une autre base de données publique, PlantExp, contient les séquences de 131 400 transcriptomes, c'est-à-dire l'ensemble des molécules recensées d'ARN issues d'un génome. Une autre base, Pmhub, contient de son côté la description de 188 837 métabolites végétaux. Ces chiffres sont donnés à titre d'exemples. Avoir une idée de la quantité totale d'informations générées et stockées est impossible. Pour le projet ABS Biotrade, le constat est que « *le séquençage des informations génétiques est devenus une pratique tellement courant dans le champ des sciences biologiques que cela a conduit à une énorme quantité d'informations produites et stockées tous les jours* »[iii].

Sur le plan technique, la génération de ces données a été notamment permise par l'abaissement du coût des méthodes de séquençage et leur rapidité croissante. Comme le détaille Save our Seeds dans son rapport, le premier séquençage d'un génome complet, celui d'*Arabidopsis Thaliana*, publié en 2000, mobilisa 10 années de travail pour un coût de 100 millions de dollars. 25 années plus tard, en 2025, le séquençage du génome d'une plante prend une semaine pour un coût inférieur à 1 000 dollars. A ces méthodes de séquençage s'ajoutent les méthodes dites « *omiques* », qui étudient tout les types de molécules présentes dans une cellule : protéine et ARN en sont deux exemples. L'objectif de plusieurs entreprises est aujourd'hui de dresser une carte des molécules produites et présentes dans une cellule, une sorte de « *Google Map des plantes* », comme le nomme Save our Seeds. En 2021, une conférence sur le premier atlas cellulaire végétal a réuni 500 représentants de gouvernements, instituts de recherche et entreprises, comme Bayer, Google ou encore Syngenta.

Des capacités techniques énergivores et aquavores

Pour traiter cette quantité de données croissantes et pour remplir les tâches demandées, les algorithmes ont besoin de matériel, d'énergie et d'eau. Les chiffres prennent ici une dimension que l'on a du mal à appréhender. Si on prend l'exemple de l'entreprise d'Elon Musk, xAI, qui prétend « *accélérer les découvertes scientifiques* », son serveur Colossus, installé à Memphis (États-Unis), a des besoins qui ont été rarement atteints. Ces serveurs fonctionnent avec 100 000 processeurs fournis par l'entreprise Nvidia. Ces 100 000 processeurs du serveur Colossus nécessitent l'équivalent de 150 mégawatts[iv] / heure pour fonctionner, soit l'équivalent des besoins de 100 000 foyers pour la même durée. Pour obtenir une telle quantité d'énergie, le serveur de xAI fonctionne aujourd'hui avec des turbines utilisant du méthane. Pour les années à venir, les acteurs du domaine travaillent à des mini-centrales atomiques. En France, l'entreprise EDF a lancé en 2023

une filiale, nommée Nuward, pour concevoir et construire de tels réacteurs[v].

L'énergie n'est pas le seul problème à résoudre. Ces serveurs, comme toute installation informatique, doivent être refroidis continuellement. Pour cela, les besoins en eau sont également faramineux. Si on regarde à nouveau le serveur de xAI, il nécessite plus de 3,5 millions de litres d'eau par jour[vi]. Rapporté à un nombre d'habitant en France par exemple, cela correspond à la consommation quotidienne moyenne de plus de 22 300 personnes[vii]… Et à l'automne 2024, xAI a annoncé vouloir doubler la taille de son serveur Colossus.

Les GAFAM mobilisés

La quantité des données « *omiques* » ou de séquences devraient croître de manière exponentielle. Par exemple, le Earth BioGenome Project mobilise plus de 60 projets dans le monde entier, sur terre comme en mer, visant à prélever des échantillons d'organismes et à séquencer le maximum, sinon toutes les molécules composant ces organismes. Des données qui vont finir enregistrées sous forme informatique dans des serveurs. On parle depuis quelques années d'informations de séquences numérisées (*digital sequence information*, DSI) pour toutes celles qui concernent les séquences de molécules (ADN, ARN, protéines…). Ce statut « *numérisé* » est à la base des tentatives en cours de la part des multinationales d'échapper à des textes internationaux, comme celui de la Convention sur la diversité biologique (CDB) visant à protéger la biodiversité. Il ouvre également la porte à de possibles brevets pour des inventions utilisant des DSI comme matière première mais qui, obtenus, seront étendus par leurs détenteurs à la matière physique composant la biodiversité.

Plusieurs cas illustrent que les algorithmes privés sont de plus en plus utilisés pour se nourrir des ces données numériques, notamment pour imaginer des modifications génétiques que les chercheurs de laboratoire des multinationales souhaiteraient mettre en place, comme nous le verrons dans un prochain article. Les entreprises du monde informatique sont depuis plusieurs années mobilisées au croisement des algorithmes et des biotechnologies. Un rapport[viii], publié en septembre 2024 par le Centre Africain pour la Biodiversité, ETC Group et Third World Network, détaille par exemple que Google, Microsoft, Amazon et Nvidia ont investi et développé des projets touchant au vivant. Retenons ici que Google a par exemple créé une entreprise avec Gigko Bioworks pour créer de nouvelles protéines en utilisant ses algorithmes de Google DeepMind. De son côté, Jeff Bezos, fondateur d'Amazon, a investi 100 millions de dollars *via* le Bezos Earth Fund pour à son tour créer de nouvelles protéines en utilisant les algorithmes.

Des investissements en croissance

Le développement des capacités informatiques pour faire tourner ces algorithmes fait l'objet, ces dernières années, de recherches croissantes de fonds. Dernier exemple en date, le 11 février 2025, l'Union européenne a lancé une initiative pour rassembler 200 milliards d'euros à investir dans le domaine. Selon la banque JP Morgan, plus de 1 000 milliards de dollars pourraient être investis entre 2024 et 2027[ix]. Au croisement des domaines informatiques et biotechnologiques, les investissements sont également substantiels. Alors que les champs concernés vont des molécules utilisées en médecine aux plantes génétiquement modifiées utilisées en agriculture, en passant par l'agroalimentaire, les agrocarburants… la valeur du marché des biotechnologies était estimé, en 2023, à 1 500 milliards de dollars[x].

Alors que la Commission européenne propose depuis 2023 de déréglementer de nombreux OGM, l'émergence de cette technologie (mal)nommée « *intelligence artificielle* » apparaît bien comme un élément de changement complet du paradigme technique existant. Le paysage même des

multinationales impliquées dans la chaîne agroalimentaire pourrait continuer d'être bouleversé dans les années à venir. Quant aux citoyens, leur place dans ce choix technologique est pour l'instant celle de simple spectateur, quand il n'est pas lui-même fournisseur de données personnelles.

i Eric Meunier, « Le séquençage du génome de 1,8 millions d'espèces est en cours », *Inf'OGM*, 3 décembre 2024.

ii B. Vogel, « When chatbots breed new plant varieties », Save our seeds, janvier 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 novembre 2024.

v Kevin Champeau, « Pour accélerer son mini réacteur nucléaire, EDF lance une nouvelle filiale », *Révolution Énergétique*, 10 avril 2023.

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

vii Misitia Ravaloson, « Consommation moyenne eau : estimation par personne et usage », *Selectra*, 9 août 2024.

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

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

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

Adresse de cet article : https://infogm.org/l-intelligence-artificielle-au-service-de-la-numerisation-des-genomes/