

« Alimentation : la révolution des omiques »

Synthèse

Hervé **This**, Eric **Dugat-Bony**, Justine **Bertrand-Michel**,
Hélène **Chiapello**, Douglas **Rutledge**

Les méthodes dites "-omiques" s'imposent largement en biologie, depuis quelques décennies : génomique, transcriptomique, protéomique, métabolomique lorsqu'elles s'intéressent à des organismes isolés ou leurs déclinaisons « méta » (métagénomique, métranscriptomique, méta protéomique, métabolomique) lorsqu'elles sont employées dans le cadre d'études de communautés d'organismes (ex : communautés microbiennes). Elles visent à détecter, identifier, quantifier simultanément et à haut débit différentes molécules qualifiées de briques de la vie : ADN, ARN, protéines et métabolites, respectivement.

Les méthodes « omiques » se sont développées grâce au développement rapide des techniques de production de données haut-débit dans le secteur des biotechnologies, d'une part, et des méthodes bio-informatiques et statistiques d'interprétation des résultats, d'autre part.

Dans ce « grand débat » de l'Académie d'agriculture de France, nous avons abordé comment cet essor récent des méthodes omiques a redéfini en profondeur notre manière d'aborder les questions scientifiques, notamment dans le domaine de l'alimentation.

D'un point de vue analytique, les méthodes « omiques » reposent sur deux types de technologies principaux : les techniques de séquençage pour les analyses génomiques et transcriptomiques, plus avancées technologiquement, et les techniques de chimie analytique pour les analyses protéomiques, lipidomique et métabolomiques, d'autre part. Le premier type de méthodes donne accès à l'information portée par les acides nucléiques (génome complet ou gènes exprimés dans une condition donnée), tandis que le second permet de caractériser les protéines, les petites molécules et les métabolites qui reflètent l'état fonctionnel d'un système biologique. Pour le traitement des données, l'essor de la bio-informatique a été déterminant, dans les deux cas, mais, également, les progrès de la chimiométrie ont largement contribué à valoriser des résultats qui n'étaient pas pleinement valorisés jusqu'alors. Et, évidemment, les progrès constants des méthodes numériques, avec désormais une place grandissante prise par les technologies fondées sur l'intelligence artificielle, accompagnent cette révolution des « omiques » pour l'analyses et l'interprétation des résultats.

Une évolution technique rapide qui conduit à des données massives et d'une grande diversité

Un des premiers domaines à se développer a été la génomique, c'est-à-dire le séquençage de l'ADN de tous types d'organismes. En 1977, il fallait plusieurs mois, grâce à une toute nouvelle technique, pour décrypter les onze gènes d'un petit virus de bactérie. Aujourd'hui, le séquençage du génome humain se fait en quelques heures, et pour un coût inférieur à 100 euros. Pas étonnant, dans ces conditions, que la génomique ait rapidement gagné tous les domaines : végétal (séquençage d'une plante modèle de la famille des brassicacées, *Arabidopsis thaliana* en 2000, ou encore du blé en

2018), animal (génomé bovin séquencé en 2018) et microbien (séquençage de la levure de boulanger *Saccharomyces cerevisiae* en 1996).

La métagénomique, connexe, étudie les métagénomes, c'est-à-dire le matériel génétique récupéré directement dans des échantillons environnementaux, sans nécessité d'isoler et cultiver les bactéries au préalable. Elle a été une évolution très importante en sciences de la vie et a contribué à l'émergence d'un nouveau champ de recherche : l'écologie microbienne. Il s'agit ainsi d'étudier des communautés microbiennes, leur fonctionnement, leur évolution dans un grand nombre d'écosystèmes, notamment alimentaires.

Pour l'étude des métabolites, c'est donc la métabolomique, qui nécessite d'autres outils d'analyse, notamment de la spectrométrie de résonance magnétique nucléaire et des spectrométries de masse variées, notamment en couplage de chromatographies en phase gazeuse ou liquide.

Et comme pour la génomique, la métabolomique se particularise selon les classes de composés : protéomique pour les protéines ; lipidomique, pour explorer l'ensemble des composés lipidiques d'un organisme, y compris les modifications apportées par un organisme à cet ensemble particulier de lipides, produit par un organisme ou un système ; sécrétomique pour l'exploration de l'ensemble des composés organiques et inorganiques sécrétés par des cellules, des tissus, des organes et des organismes biologiques ; glycomique pour l'étude complète du glycome, c'est-à-dire des glucides ; pharmacomicrobiomique pour l'exploration de l'effet des variations du microbiome humain sur les médicaments et *vice versa* ; toxicogénomique, pour la collecte, l'interprétation et le stockage d'informations sur l'activité des gènes et des protéines dans une cellule ou un tissu particulier d'un organisme en réponse à des substances toxiques.

Toutes ces méthodes omiques produisent des jeux de données massifs avec des techniques de plus en plus facilement accessibles, souvent grâce à des partenariats avec soit des plateformes ou infrastructures de recherches, soit avec des entreprises internationales du secteur des biotechnologies. Mais elles posent aussi un certain nombre de défis en termes d'analyse, d'intégration et d'interprétation.

Des questions d'analyse

Pour des analyses « haut-débit », Justine Bertrand-Michel, de l'Inserm à Toulouse, indique que c'est l'automatisation instrumentale, couplée au développement des approches numérique, qui a été le facteur clé du développement des méthodes omiques. On l'a oublié, mais, pour la découverte des premières hormones, il a fallu collecter des milliers de litres d'urine de juments gravides, pour ne récupérer finalement que quelques microgrammes d'hormone. Puis avec l'amélioration des méthodes de chromatographie, d'abord sur couches minces, puis sur colonne, en phases gazeuse ou liquide, les chimistes sont parvenus à détecter des quantités aussi faibles que la dizaine de femtogramme (millionième de milliardième de gramme). Et on doit aussi se souvenir que, il y a seulement 50 ans, l'interprétation des spectres, et notamment la détermination des quantités des composés séparés se faisant par découpage des spectres produits par l'analyse, et pesée du papier sur une balance.

Progressivement, les méthodes chromatographiques ont été épaulées par des spectromètres de masse, qui identifiaient les composés séparés, mais, plus récemment, les méthodes de spectrométries de masse sont devenues si efficaces (trappe orbitale, résonance cyclotronique ionique, etc.) qu'elles peuvent se passer de séparation préalable. Pour les haut débit (centaines à milliers d'échantillon, la *flow-injection* couplée à la spectrométrie de masse se fait en moins d'une minute par échantillon. La chromatographie en phase liquide ultra haute performance couplée à la spectrométrie de masse est plus lente, mais avec une grande qualité des résultats. La spectrométrie de résonance magnétique nucléaire, elle, peut être très rapide (quelques minutes par échantillon) ; elle est peu sensible, mais très reproductible.

En tout cas, les spectres produits par ces appareils comportent des centaines de milliers de points,

de sorte que des milieux complexes peuvent être analysés : fluides biologiques, par exemple. Ce seront autant de résultats qu'il faudra ensuite traiter, soit pour des dosages, soit pour l'étude de mécanismes du vivant.

Pour les explorations du génome ou du transcriptome, les techniques de séquençage des acides nucléiques ont beaucoup évolué dans les années 2000 avec l'essor des techniques dites de seconde génération. Les coûts de séquençage de l'ADN ((meta)génomique) ou de l'ARN ((meta)transcriptomique) ont alors beaucoup baissé, et des instruments très performants se sont imposés dans le domaine, avec des séquences assez courtes (quelques centaines de paires de bases) et taux d'erreurs très faibles ($< 1\%$). Depuis quelques années, les techniques dites de troisième génération permettent de séquencer des longs fragments d'acides nucléiques (plusieurs dizaines de kilobases) et permettent ainsi de décrypter des génomes complexes beaucoup plus rapidement. Ces techniques sont performantes et robustes, mais les taux d'erreurs sont un peu supérieurs (5 à 10 %) et nécessitent des stratégies de correction adaptées.

Et des questions de traitements des résultats d'analyse

Produire rapidement de nombreux résultats est évidemment un atout... si l'on peut analyser les résultats. Les méthodes omiques produisent des données massives, hétérogènes, issues de techniques différentes. Il faut les stocker, les standardiser, les analyser, les comparer, les visualiser, les intégrer et leur donner du sens, cela pose un grand nombre de défis en sciences du numérique, bio-informatique, ingénierie des connaissances et analyse statistique de données.

Il reste de nombreuses questions scientifiques pour en extraire des connaissances toujours plus complètes ; en voici quelques unes : comment prendre en compte les structures tridimensionnelles des molécules, notamment dans les interactions hôte-microbiome observées ? comment passer de corrélations à des relations causales ? comment intégrer et représenter la dynamique spatio-temporelle de ces communautés microbiennes ? comment garantir la robustesse et la reproductibilité des résultats ?

En 2011, quand sont apparues les toutes premières plateformes de spectrométrie de masse par imagerie, où un laser balaie une coupe de tissu pixel par pixel, avec un spectre de masse acquis pour chaque position, les laboratoires s'arrachaient les exceptionnels bio-informaticiens capables d'utiliser les résultats d'analyse, en vue d'établir les cartes de répartition spatiale des molécules.

Progressivement, grâce aux collaborations scientifiques, des algorithmes ont été élaborés et partagés entre les différentes plateformes de génomique ou de métabolomique. Reste que les méthodes statistiques doivent encore évoluer.

Lors de la séance organisée, Douglas Rutledge a ainsi comparé les techniques d'analyses en composantes principales avec l'analyse en composantes indépendantes, et la technique ComDim. Classiquement les analyses en composantes principales avaient été introduites (par Pearson, en 1901) pour explorer des grandes quantités de nombres, par exemple quand des « individus » étaient caractérisés par plusieurs grandeurs (taille, poids, couleur des yeux, indice de masse corporelle, âge, etc.) : ces individus sont alors représentables par des points dans espaces ayant autant de dimensions que l'on a de paramètres par individu ; et l'analyse en composante principale cherche des groupements dans ces espaces. L'analyse en composantes indépendantes vise à extraire des signaux sources inconnus (et leur proportions) à partir de signaux observés. Les techniques ComDim (analyse en composantes communes et poids spécifiques), elles, font l'analyse simultanée de plusieurs tableaux de nombres avec différentes variables décrivant les mêmes objets. Par exemple, elles permettent d'estimer la proportion d'information métabolique commune extraites des données spectrales de différentes plateformes de spectrométrie de masse et de résonance magnétique nucléaire.

Et de merveilleux résultats

Que fait-on de tous ces jeux de données omiques issus d'analyse haut débit ? Lors de la séance du 24 juin, Eric Dugat-Bony (Inrae, UMR Sayfood) a évoqué un récent travail collaboratif français concernant les fromages d'appellation d'origine protégée (programme MetaPDOcheese) : il s'est agi de cartographier la diversité microbienne des laits et des fromages AOP français et d'explorer les déterminants liés au procédé de fabrication, au terroir et aux pratiques des producteurs associés à la composition des communautés microbiennes. Une campagne d'échantillonnage sans précédent a été organisée : 2800 échantillons de laits et fromages AOP ont été récupérés et analysés par une méthode de métagénomique ciblée. De plus, 100 variables associées à ces échantillons ont été collectés auprès des producteurs couvrant les facteurs imposés par l'AOP (topographie, espèce laitière, temps d'affinage, etc.), les pratiques d'élevage (alimentation, hygiène, etc.), et des facteurs conditionnels (saison, type de production, méthode de salage, etc.). Les résultats ont été à la hauteur des efforts, avec une caractérisation inédite des profils microbiens de l'ensemble des échantillons, la caractérisation des espèces communes, des espèces particulières, la mise en relation de la communauté microbienne des laits et des fromages. Les facteurs les plus significatifs associés à la richesse microbienne des fromages sont la famille technologique, le label AOP et les facteurs dépendants de l'AOP (région française, durée d'affinage) dans toutes les familles de fromages, en cœur et sur la croûte, la topographie, les pratiques de soin de la croûte, l'utilisation de planches en bois, la méthode de salage dans certaines familles de fromages.

Les limites

Aujourd'hui les méthodes métabolomiques sont plus longues et complexes à mettre en oeuvre que les méthodes génomiques, notamment parce que les échantillons couvrent plusieurs ordres de grandeur de concentrations, et aussi plusieurs ordres de grandeur de tailles, avec des extractions qui s'imposent encore pour de nombreuses techniques d'analyse. Un des problèmes importants des analyses métabolomiques est celui de la robustesse des résultats, qui peuvent varier beaucoup selon le type d'instrument d'analyse utilisé.

Cela étant, avec quelques décennies de recul nous pouvons nous interroger sur la place exacte que peuvent avoir ces démarches omiques, notamment pour les études scientifiques ou technologiques de l'alimentation. Elles se sont imposées dans le secteur médicale, les séquençage génomique étant déjà à la base des médecines personnalisées, mais en viendra-t-on aussi à de l'alimentation personnalisée ? Du point de vue technologique, elles commencent à donner des résultats importants, comme, il y a quelques mois, avec la mise en œuvre de techniques métagénomiques pour explorer l'origine des fins chocolats : on a découvert que ce goût spécifique recherché découle moins de la torréfaction et de la variété de cacao que du type de fermentation, dans des récipients en bois, avec des souches microbiennes qui ont été caractérisées. En tout cas, les méthodes omiques sont maintenant omniprésentes pour des travaux scientifiques : la biologie moderne en fait un usage quasi permanent.

Ce qui n'a pas été formalisé, c'est la nature des champs où ces techniques sont utiles. On comprend bien qu'elles pourraient nous aider à envisager non pas seulement les lipides ou les protéines, mais d'autres classes importantes de composés organiques, tels les chlorophylles, ou les caroténoïdes, ou les mélanoidines, dans des chlorophyllomiques, caroténoïdomiques ou mélanoidiniques. Mais pourraient-elles, par exemple, aller jusqu'à nous aider à savoir si les sauces mayonnaises sont ou non des émulsions de Ramsden, avec des granules du jaune d'oeuf abaissant l'énergie de surface ? Ou contribueraient-elles à des études des « échanges » entre de tissus végétaux ou animaux et un environnement liquide ? Alors que plus de cinq millions d'articles « scientifiques ou technologiques » ont été consacrés à l'étude du thé, que pas un n'examine les mécanismes par lesquels des composés des feuilles de *Camelia sinensis* passent en solution aqueuse, des techniques omiques pourraient-elle contribuer à nous éclairer sur ces mécanismes.

On peut rêver à des grandes solutions, de grands programmes, de grands travaux, mais on n'oubliera pas que, à ce jour, les extractions précèdent souvent les analyses d'échantillons, souvent biologiques. Or il a été observé que les meilleures méthodes d'extraction des métabolites (notamment des sucres) des racines de carottes ne permettaient pas de voir la quantité réelle de ces composés dans les racines de carotte (*Daucus carotta* L.). La technique d'analyse par spectrométrie de résonance magnétique nucléaire *in situ* quantitative a pallié cette insuffisance, mais elle reste à généraliser, et, plus généralement, il se préoccuper des échantillons que l'on analyse avant d'interpréter des résultats de mesure. En améliorant l'analyse, d'une part, et le traitement des résultats de mesure, d'autre part, il n'y a pas à douter que les méthodes omiques ne sont qu'au début de leur essor.

Références

- Irlinger, F., Mariadassou, M., Dugat-Bony, E., Rué, O., Neuveglise, C., Renault, P., Rifa, E., Theil, S., Loux, V., Cruaud, C., Gavory, F., Barbe, V., Lasbleiz, R., Gaucheron, F., Spelle, C., Delbès, C., 2024. A comprehensive, large-scale analysis of "terroir" cheese and milk microbiota reveals profiles strongly shaped by both geographical and human factors. ISME Communications 4, ycae095. <https://doi.org/10.1093/ismeco/ycae095>
- Esteves da Silva C, Vitoria Leimann F, Ribeiro Sanches YY, de Carvalho Rodrigues V, Tfayli A, Rutledge DN, Março PH. Food freshness and composition evaluated by Colorimetry, TPA, and spectroscopy through ICA-based ComDim: A case study of a peanut-based protein-enriched food, Food Control, DOI: 10.1016/j.foodcont.2024.110978
- Bouveresse Jouan-Rimbaud D, Rutledge DN. 2024. A synthetic review of some recent extensions of ComDim, Chemometrics, 38(5), e3454.
- De Figueiredo M, Giannoukos S, Wütrich C, Zenobi R, Rutledge DN. A tutorial on the analysis of multifactorial designs from one or more data sources using AComDim, Journal of Chemometrics, DOI: 10.1002/cem.3384
- This H, Weberskirch L, Plassais P, Luna A, His A, Skoglund S. La RMN du liquide voit le coeur des légumes et des viandes... puisque ce sont des gels. L'Actualité chimique. 2010 (1), 337, pp. 10-13.
- Bauchard E, This H. 2015. Investigating the performance of *in situ* quantitative nuclear magnetic resonance analysis and applying the method to determine the distribution of saccharides in various parts of carrot roots (*Daucus carota* L.), *Talanta*, 131, 335-341, DOI: 10.1016/j.talanta.2014.07.097.
- This H, Febvay L, Bria M. 2021. L'intégration des signaux (incluant des résonances) dans les spectres unidimensionnels obtenus par spectroscopie de résonance magnétique quantitative *in situ*, *Cahiers Techniques de l'INRAE*, 104(4), https://www6.inrae.fr/cahier_des_techniques/Les-Cahiers-parus/Les-N-reguliers/2021/Cahier-N-104/Art4-ct104-2021.
- Chiapello H, Rodolphe FF. Introduction to bioinformatics, Functional Plant Genomics, Science Publishers, 2007, 978-1-57808-506-4.
- Dechêne-Tempier M, Bougeard S, Loux V, Chiapello H, Libante V, et al.. Pan-immune system, mobilome and resistome in *Streptococcus suis*. Microbial Genomics, 2025, 11 (9), pp.001521