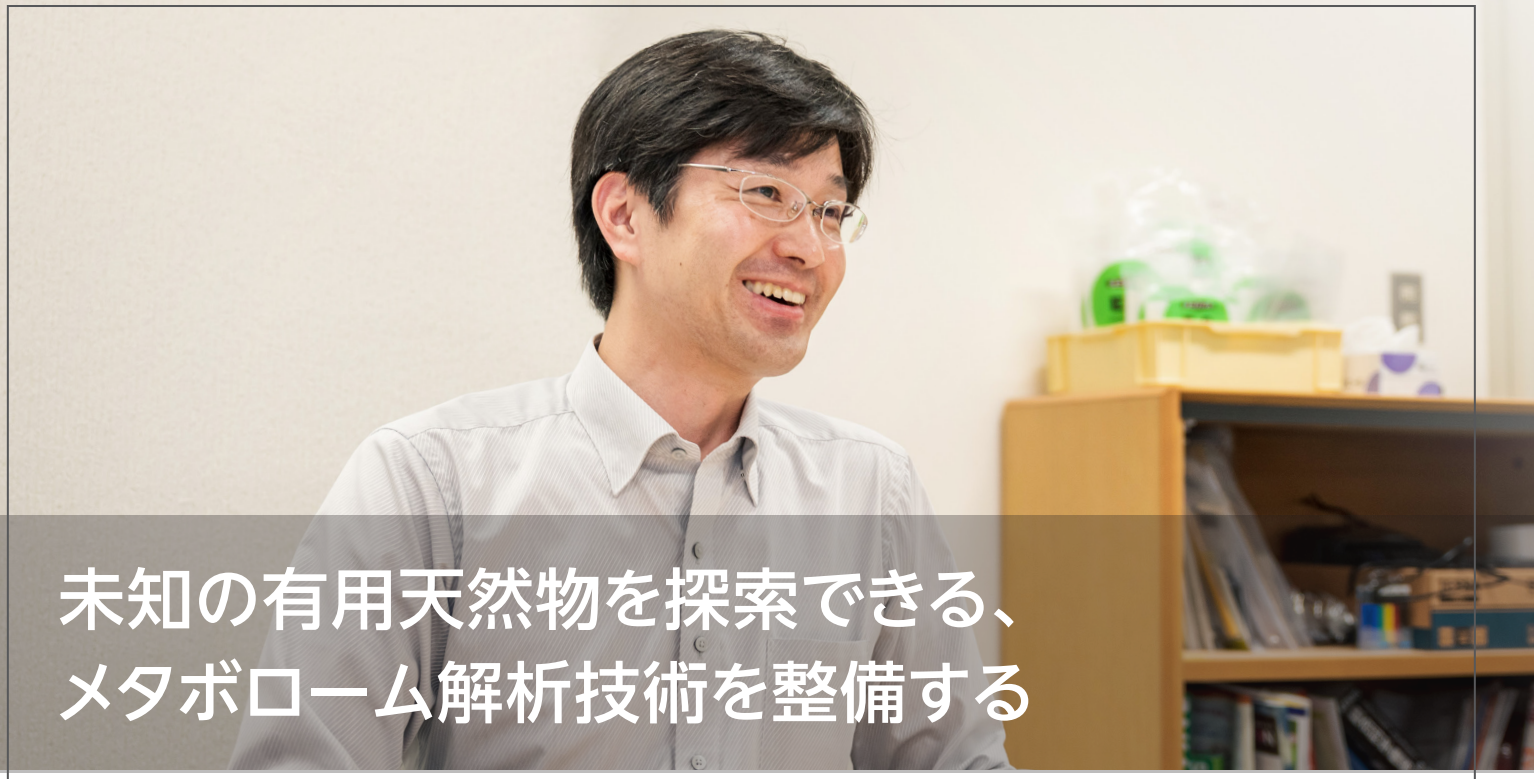


# 未知の有用天然物を探索できる、メタボローム解析技術を整備する

特任准教授　櫻井 望（さくらい のぞむ）　生命情報・DDBJ センター

東北大学農学部（山谷知行研究室）卒業、東北大学大学院農学研究科 博士課程修了。博士（農学）。公益財団法人 かずさDNA研究所 研究員、同メタボロミクスチーム チーム長を経て、2018年6月より、国立遺伝学研究所（遺伝研）特任准教授。2012年 日本植物細胞分子生物学会技術賞共同受賞。

**昨年6月に着任した櫻井 特任准教授。それまでは約17年にわたり、かずさDNA研究所でメタボローム解析の技術開発に携わってきた。作り上げた「解析対象とする化学成分を限定しない質量分析のツールやソフト」は、自らが管理するポータルサイトで公開している。未知成分の詳細な物質名が特定できなくても、化学的な特徴や候補物質のリストなどが得られることで、これまで埋もれていたデータを活用できると考えている。**

## かずさ DNA 研究所で始めたメタボローム解析

私は農学系の出身で、はじめは東北大学でイネの増産を目的に、窒素栄養の体内循環について研究していました。2001年に、かずさDNA研究所（以下、かずさDNA研）に赴任し、その翌年に立ち上がった「NEDO（新エネルギー・産業技術総合開発機構）の委託によるメタボローム解析プロジェクト」に参画することになりました。かずさDNA研は1994年にDNA専門の研究機関として設立され、ゲノム解読、DNA分析技術開発、遺伝子予測プログラム開発などを進めてきましたが、「生物をよりよく理解するためには生体内で代謝される物質の網羅的な情報も必要だろう」ということになり、メタボローム解析も始めることになったのです。当時は、さまざまな生物のゲノム解読が一気に進み、遺伝子発現などの網羅的（オーム）解析がちょっとしたブームになっていたという背景もあります。

NEDOのプロジェクトでは、工業原材料となる物質の増産や探索を目的に、主に植物を対象にしたメタボローム解析を行いました。私たちの他には、慶應義塾大学と理化学研究所が、それぞれに独自のメタボロームプロジェクトを立ち上げていました。

このような経緯で私はバイオインフォマティクス（生物情報学）の分野に入り、メタボローム解析と解析のためのコンピュータ技術開発を進めつつ、現在に至っています。

## 網羅的とはいえなかった初期のメタボローム解析

ただし、その頃のメタボローム解析は網羅的といえるようなものではありませんでした。質量分析装置を用いて特定の代謝物を調べる研究は行われていましたが、生データ全体を取り込んで網羅的に解析できるコンピュータ・ソフトがなく、試料に含まれる数千数万の代謝物の全貌を描き出すことなど、夢のまた夢だったの

です。今でも完全に実現できているわけではありませんが、技術はかなり進みました。

研究者はどうしていたかというと、市販されている「精製された単一物質（標品）」を購入して質量分析装置にかけ、得られたシグナルをデータ化して指標として使っていました。たとえば、A物質の指標と「試料を質量分析して得られたシグナル」とを照らし合わせれば、A物質については存在の有無や量を確定することが可能だからです。ただし、標品は限られた物質のものしかなく、最大でも1000種ほどでした。解析ソフトも成分情報を一つずつ見るものしかなく、標品がない成分を含め、データの中にある成分シグナルを網羅的にとらえ解析することは不可能だったのです。

そこでNEDOのプロジェクトでは、「質量分析装置が出す全てのシグナルを読み込んでデータ化できるソフト（PowerGet）」を開発しました。いまだに性能向上を目指して改良中ですが、このソフトを使えば、全シグナルについて、その質量と特徴などをリスト化することができます。出てきたリストを試料ごとに比較するなどして、ユーザーが「興味深い」と感じれば、その物質だけを取り出して詳しく解析することも可能です。つまり、標品をもとに特定の成分を対象にするのではなく、対象を定めない（ノンターゲットな）解析が可能となってきたのです。

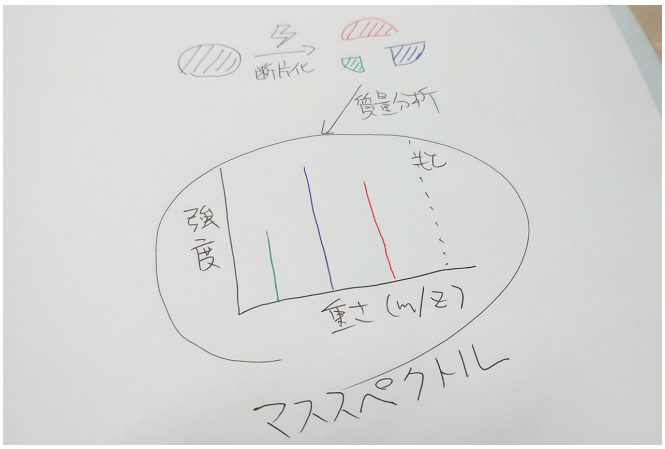
「DNAチップを使って遺伝子発現の全体像を俯瞰してから、特定の遺伝子に注目する」というトランスクリプトーム解析のストラテジーを、メタボローム解析でもある程度実現できたといえるでしょう。

◀PowerGet へのリンクはこちら

## ノンターゲットで7000種のフラボノイドを予測できるツール

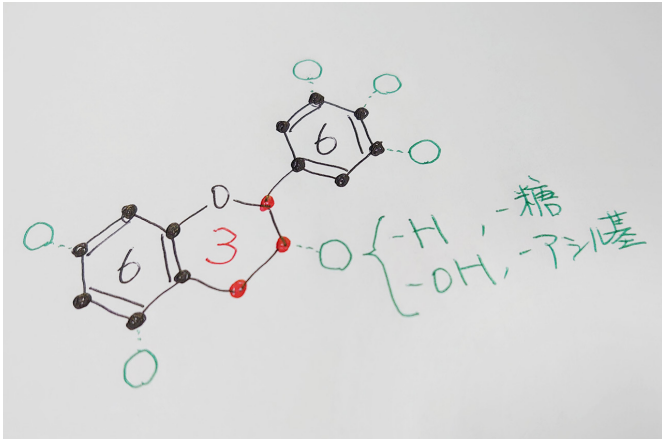
NEDOプロジェクト終了後も、「質量分析装置が拾っているはずの物質の情報を全て知りたい」というモチベーションの下、かずさDNA研において、メタボローム解析で最大の課題とされる「標品がない成分を推定するためのツール開発」を進めました。

一例をあげると、「フラボノイドを網羅的に検出できる技術（FlavonoidSearch）」を構築しました。フラボノイドは、植物が色素成分などとして作り出す一群の代謝物の総称で、7000種類以上あることがわかっています。抗酸化作用やホルモン様の作用をもつものが多いために機能性食品の成分としても注目されています。アントシアニン、茶カテキン、大豆イソフラボンなどが有名ですが、機能が未解明なものも多くあります。一方で、成分の同定に指標として使える市販のフラボノイド標品は100ほどしかありません。つまり、解析対象の試料に有用な生理活性をもつフラボノイドがあったとしても、検出できずにいることになります。




このような背景もあり、私は「液体クロマトグラフィー-質量分析（LC-MS）」という質量分析で得られるデータ（マススペクトル）を使って「フラボノイドと思われる物質」をすべて検出するツールを作りたいと考えました。マススペクトルというのは「成分の化学構造を反映したフィンガープリント」のようなものといえ、この情報をうまく使うと、化学構造の判別や推定がある程度可能になります。

◀FlavonoidSearch へのリンクはこちら




フラボノイドはC6-C3-C6の母核（ジフェニルプロパン構造）をもち、母核に糖やアシル基などの化学基が付加された構造になっ

ています。私たちはまず、市販の評品を入手して精密な質量分析を行い、正確なマススペクトルデータを得ました。次に、過去の文献情報なども参考にし、「どのようなフラボノイドの化学構造なら、どのようなスペクトルが得られるか」を予測するためのルールを構築していきました。この試みがうまくいったのは、当時、ポスドクとして一緒に開発に携わった、化学を専門とする秋元奈弓さん（現、かずさDNA研）が「職人的な離れ業」を発揮してくれたからだと思います。秋元さんは、私を含むスタッフの多くが理解できないほどの高いレベルで、わずかな化学構造の違いとマススペクトルとの関連性を見抜き、予測ルールを構築していきました。そして、実に3年をかけて7000種のフラボノイドの構造をマススペクトルデータに変換したのです。私はそのデータを使って、実際にLC-MSで測定されたスペクトルシグナルと比較できるプログラムを作り、幅広い生物種で網羅的にフラボノイドの候補を検出できるFlavonoidSearchを完成させました。

実際に、パセリに含まれるフラボノイドをFlavonoidSearchで網羅解析してみたところ、マススペクトルデータが得られた約1000の成分中、36個がフラボノイド候補であることが分かりました。そのなかには、母核にこれまでに知られていなかった化学基が付加されているもの、つまり新規のフラボノイドも含まれていました。これまでパセリで報告されていたフラボノイドは6個程度しかありませんので、FlavonoidSearchは新しい機能性フラボノイドの探索に活用できると期待できます。

FlavonoidSearchが完成した2016年は、AlphaGoというAIがプロの囲碁棋士に完勝したことが大きなニュースになりました。そのためか、「成分推定のためのツール開発にもAIを使えばよいのでは？」と思われがちなのですが、AIでは様々な化学構造を適切に処理できるパラメーターの設定が難しく、実用はまだ先のようです。

## 食品メタボロームレポジトリの構築と公開

2017年には、「食品メタボロームレポジトリ」という情報リソースを構築しました。略して「食レポ」とよんでいます。なぜ食品を対象にしたかというと、食品は植物、畜産物、魚とさまざまな生物を加工して作られるものなので、多様な代謝物を含んでいると考え、メタボローム解析全般に広く役立つと思ったからです。

◀食品メタボロームレポジトリへのリンクはこちら

私たちは、スーパーなどで買える一般的な222の食品をLC-MSで分析し、各成分の「LC溶出時間（水に溶けやすい成分であれば溶出時間は短く、溶けにくくなるほど溶出時間が長くなる）」と、「正確な質量値（質量電荷比）」のデータを集めました。結果的に、1食品あたり、平均して約4000もの物質を拾えました。食レポでは、マススペクトルが得られた成分についての「FlavonoidSearchでフラボノイドを予測した結果」も提供しています。さらに、横軸が溶出時間、縦軸が質量電荷比をあらわす2次元のマスクロマトグラムも提供しており、食品ごとの成分の違いを視覚的にとらえることも可能になっています。

日本食品標準成分表には、主な食品について、炭水化物、タンパク質、脂質、脂肪酸、アミノ酸、有機酸、ミネラル、ビタミン類など、約140の成分ごとに定量値が記載されています。ところが実際の食品には、成分表にはない数千もの物質が含まれており、そのなかには微量でも生体に有用なもの、原材料が同じでも製造過程で成分が大きく変化するものなどもあると思われます。このような未知の成分を明らかにして有効活用できればよいのですが、「どの未知成分にどうやって着目すればよいのか」が新たな問題となります。このようなとき、物質名まで特定できなくても、同じ成分が「ほかにどのような食品で検出されているのか」という情報があると、解析すべき物質の優先順位をつける手がかりになります。食レポを利用すると、このような情報を簡単に得ることができるのです。

想定される使い方を一つあげてみましょう。ヨーグルトの研究をしていて、ある乳酸菌を使った製品だけに「血圧を下げる効果」がみられたとします。PowerGetのようなソフトを使ってノンターゲット解析をしたところ、血圧を下げる効果と相関する成分候補が20物質も出てきたとします。これらをすべて詳細に解析するのは不可能なので、ターゲットを絞る必要があります。このようなとき、食レポでは「ユーザーがLC-MSにより得た*m/z*値（質量電荷比）」を使って、同じ未知成分がほかにどのような食品に含まれているかを検索し、リスト化できます。20候補のうちのA物質について、「ヨーグルトなどの乳酸菌で作った製品に特に多くみられ、原料の生乳やチーズでは検出できない」といったことがわかれば、Aは乳酸菌のはたらきで作られる有用成分の可能性が高く、優先順位が上がります。一方で、20候補のうちのB物質が、「野菜や魚などにもみられる、ありふれたもの」とわかれば、Bの優先順位は下がります。

実際に食レポが役立った私の研究例を紹介しましょう。複数の地域で「作物の根の周囲」の土壌を採取してノンターゲット解析したところ、「ある地域の土壌だけから抽出される物質」がありました。その物質を食レポのサーチにかけてみたところ、同じ物質を含む食品はみつかりませんでした。私は「みつからない」ということを

手がかりに、「この成分は植物ではなく、土壌細菌が作り出すものではないか」と推定して詳しく調べてみました。すると思惑どおり、土壌細菌由来のものであるとわかりました。このようなアプローチによる研究を進めれば、「新しい抗生物質のシーズ」などを探索することも可能だと思います。

ただし、食レポで採用したLC-MSの分析条件では、タンパク質、疎水性の強い物質、イオン化されにくい物質などは測定できない点を申し添えておきます。

## 産学連携によるメタボローム解析データのレポジトリ開発を

ここまでに3つのツールを紹介しました。いずれもKOMICS (Kazusa Metabolomics Database)というポータルサイトで公開しており、学術研究用であれば誰でも利用できるようになっています。

◀KOMICS (Kazusa Metabolomics Database) [ 日本 ] へのリンクはこちら

遺伝研での私のミッションは、かずさDNA研での研究開発を踏襲し、質量分析装置の種類などを問わない公共のメタボローム解析レポジトリを構築することにあります。
国内だけでなく、世界を見渡しても「本当に使えるメタボローム解析レポジトリ」は、まだありません。データベースとしては米国にMETABOLOMICSWORKBENCHが、ヨーロッパにMetaboLightsがありますが、単なる情報アーカイブのようになっており、比較解析できるツールとしては機能していません。

◀METABOLOMICSWORKBENCH [ 米国 ] へのリンクはこちら

◀MetaboLights [ ヨーロッパ ] へのリンクはこちら

その大きな要因は、分析装置と分析条件によって、試料の溶出時間、イオン化の特性、検出感度などが異なり、データを単純比較できないことにあります。私は、食レポの経験を生かして、質量分析の装置や条件を問わない、より使い勝手のよい「検索や比較ができるデータベース」の開発を進めたいと考えています。

メタボローム解析は、病気の予防、創薬、新たな治療法開発などにつながるポテンシャルを秘めています。たとえば、血液を質量分析すると1000以上のシグナルが出てきます。その大半は、食べ物、腸内細菌、肝臓などでの代謝物由来だと思われますが、どのような物質なのか特定できていません。もし、健康状態によって顕著に変動する物質(バイオマーカー)や漢方でいう未病(病気になる前段階)の状態を反映する物質が特定できれば、大いに役立つだろうと思います。

そのため、産学連携の推進もまた、私の大きなミッションになっています。すでに、遺伝研、かずさDNA研、ツムラが連携し、漢方薬成分の網羅的解析、成分の構造推定などを進めています。日本では漢方医学が独自に発展し、病気の治療における有用性もかなりわかっています。現代医療においても不可欠な存在といえますが、有効成分や作用メカニズムについては未解明な部分が多く残されています。ツムラとはツールやデータを共有して「漢方ミクス」とでもいうべき解析を進め、有益な知見を蓄積したいと考えています。

産学連携は特許や知的財産権が絡んでややこしい部分もありますが、うまく連携しないと成果を実用化につなげられません。私は、産も学もメリットを得られるように環境や法律を整備すべきだと考えています。

## 夢は安価な質量分析装置と新規センシング技術の開発

メタボローム解析では、「埋もれていたデータを生かして有用な未知物質を発見すること」にやりがいを感じていますが、解決すべき点もいろいろあると思っています。その一つに、質量分析装置が非常に高額だという問題があります。安くて1000万円、高額になると2億円にもなりますので。今は夢でしかありませんが、いつか、自分の手で安価な質量分析装置も開発したいと思っています。また、質量分析に代わる安価なセンシング技術を開発できたらと思っており、実際に取り組みも始めています。

私の研究開発により、「メタボローム解析はやりにくい」と敬遠する研究者が多い現状を打破できると嬉しいと思います。メタボロームでは解析対象が無限にあって複雑なのは確かですが、ゲノミクス研究者にも「やってみようかな」と思ってもらえるよう、使い勝手のよい技術の開発を進めていきたいと思っています。

聞　き　手：サイエンスライター　西村 尚子
写 真 撮 影：リサーチ・アドミニストレーター室　来栖 光彦

2019年7月