

Smart Metabolites Database™による 農作物の産地判別

農作物の産地偽装は農業・食品業界における普遍的な課題であり、今日まで産地判別のためのさまざまな手法が開発されてきました。一般的に、遺伝情報など農作物が先天的に持つ生体情報をもとに判別を行うことは難しいため、後天的に農作物の体内に発生する物質の情報が用いられます。中でも農作物が体外から取り入れる元素の情報をもとに判別を行う手法は安定性に優れ、長年研究されてきました。

一方で、農作物体内に後天的に発生する物質の中には、体外から取り入れられる物質のほかに、アミノ酸、有機酸、脂肪酸、糖など、体内で産生される物質があります。これらは総称して代謝物と呼ばれ、多くの農作物に共通して含まれており、その存在濃度は、同じ品種の農作物でも周囲の環境に合わせてダイナミックに変化すると考えられます。このような農作物内の代謝物を網羅的に測定することで、特定の農作物品種における各産地の代謝物の存在比パターンを見出すことができれば、農作物の産地判別に利用できる可能性が考えられます。

本稿では、親水性の代謝物377成分を同時測定できるSmart Metabolites Databaseを用いて、国産および外国産のアスパラガス計106サンプル中の代謝物を測定し、国産と外国産を判別するモデルを作成し、約90%の精度で判別できることを確認しました。

T. Sakai



■アスパラガスからの代謝物の抽出と誘導体化

アスパラガスは、国産58サンプル、外国産48サンプルを準備しました。

適切な大きさにカットしたアスパラガスを縮分し、凍結乾燥させました。凍結乾燥後に粉碎処理を行い、得られた粉末に対し、Bligh & Dyerの方法をベースとした前処理プロトコルにて抽出・誘導体化しました。Bligh & Dyerの方法については当社カタログの「メタボロミクス前処理ハンドブック」(C146-2181)をご参照ください。

なお、内部標準にはRibitolを用いました。

* サンプルは独立行政法人農林水産消費安全技術センター (FAMIC) 様からご提供いただき、前処理に関してもご協力いただきました。

■誘導体化した親水性代謝物の測定

誘導体化後のサンプル溶液をGC-MS/MSで測定しました。分析条件は、当社Smart Metabolites Databaseの内容に準拠しました。表1に詳細な条件を示します。

表1 測定条件

GC-MS	: GCMS-TQ™8040 NX
分析カラム	: BPX-5 (30 m x 0.25 mm, 0.25 μm)
- GC -	
注入モード	: Split (30 : 1)
気化室温度	: 250 °C
カラムオープン温度	: 60 °C (2 min) → (15 °C/min) → 330 °C (3min)
制御モード	: 線速度 (39.0 cm/sec)
ページ流量	: 5.0 mL/min
- MS -	
測定モード	: MRM
イオン源温度	: 200 °C
インターフェース温度	: 280 °C
イベント時間	: 0.25 sec

■ピークの検出

当社解析ソフトウェアのLabSolutions Insight™ (ver. 3.5)を用いて、ピークの検出作業を行いました。ピーク検出に関しては下記のルールを設定しました。

★ルール1:

保持指標から予測された保持時間±0.08分以内に、定量トランジションと確認トランジションが同時にピークを形成しており、かつ定量イオンの高さが1000以上のものを「検出」とする (図1)。

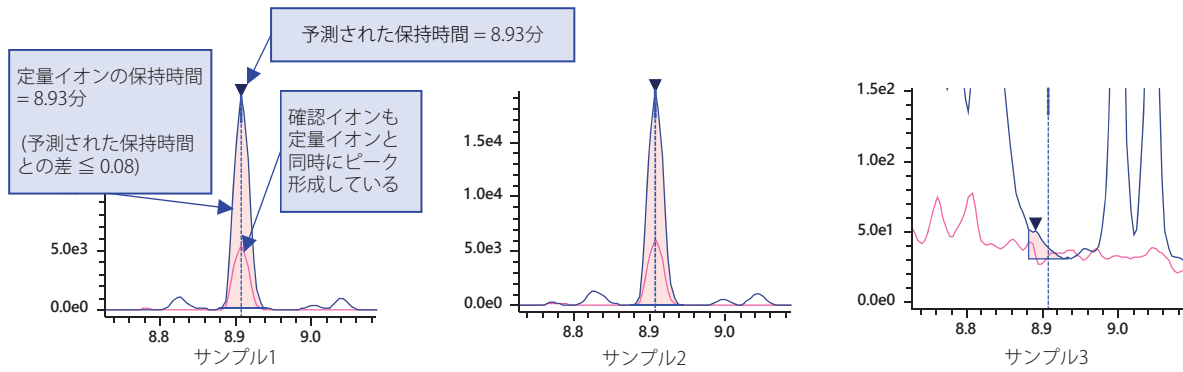
★ルール2:

半数以上のデータで「検出」となっている化合物に関しては、いくつかのデータでルール1を満たさない場合でも、予測された保持時間付近のピークを以て「検出」とする (「非検出」のデータが欠損値または0となることを回避するため: 図1)。

★ルール3:

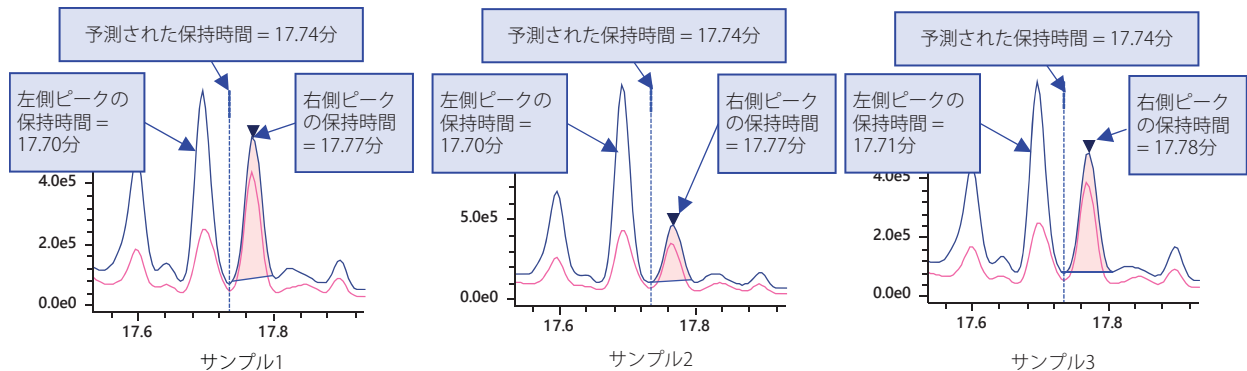
予測された保持時間付近に2本以上の「検出」とするべきピークがある場合は基本的には近い方のピークを「検出」とするが、それがサンプルによって異なる場合には、半数以上のサンプルで「検出」とされたピークの方を「検出」とする (サンプルによって異なるピークが「検出」とされるのを回避するため: 図2)。

LabSolutions Insightは、このようなピーク検出作業に最適な機能を有しており、機械学習など多数のデータが必要となる解析でも短い時間で作業が可能です。また、今回はGC-MS/MSで測定したことにより、形状の安定したピークが多く、結果として多くのピークを検出することができました (計217成分)。



サンプル1はルール1を満たすため、ピーク「検出」とする
 サンプル2もルール1を満たすため、ピーク「検出」とする
 サンプル3はルール1を満たさないが、サンプル1, 2で「検出」となっているため、予測された保持時間に近いピークを「検出」とする（「非検出」を0で代替しない）

図1 ルール1およびルール2の例



サンプル1では右側ピークのほうが予測された保持時間に近いため、右側ピークを「検出」とする
 サンプル2でも右側ピークのほうが予測された保持時間に近いため、右側ピークを「検出」とする
 サンプル3では左側ピークのほうが予測された保持時間に近いが、サンプル1, 2で右側ピークを「検出」としたため、右側ピークを「検出」とする（サンプルによって「検出」とするピークが異なるようにする）

図2 ルール3の例

産地判別モデルの作成

ピークの検出作業後、各ピークの高さをデータ行列として出力し、アプリケーションニュースM282に記載した内容と同様のデータ前処理を行いました。上記のルールでピーク検出を行った結果、当データでは欠損値を含むデータ列は存在していないので、欠損値処理は省略しました。内部標準のピークが基準値以下となったサンプルに関しては、誘導体化工程で不具合があったものと見なしてデータ行列から削除し、それ以外のサンプルに関しては、各ピーク面積値を内部標準のピーク面積値で除算した値を、zスコアにより標準化したものをデータとして使用しました。

アプリケーションニュースM282と同様に、全サンプルをランダムにトレーニングセットとテストセットに分けたうえで、ピークが検出された217成分に関して、ボックスプロットや外れ値の分布を確認し、判別に有効と思われる13成分を選択しました。この13成分でランダムフォレストのアルゴリズムを用いて、産地判別のモデルを作成しました。サンプル内のトレーニングセットとテストセットをランダムに入れ替える操作を50回行い、それぞれここで作成したモデルを適用して予測の精度を計算したところ、モデル精度の平均は91.7%となりました。このモデルによる代表的なConfusion MatrixおよびROC曲線を図3に記します。

当実験により、親水性代謝物の網羅測定が農作物の産地判別に応用できる可能性を示唆することができました。

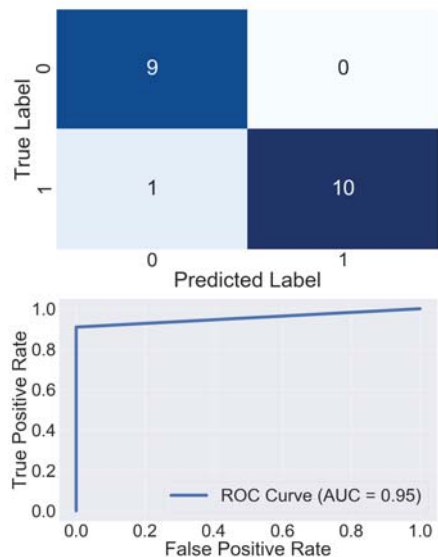


図3 アスパラガスの産地判別モデルにおけるConfusion Matrix(上)およびROC曲線(下)

Smart Metabolites Database、GCMS-TQおよびLabSolutions Insightは、株式会社島津製作所の日本およびその他の国における商標です。その他、本文中に記載されている会社名および製品名は、各社の商標および登録商標です。本文中では「TM」、「®」を明記していない場合があります。