

### 使用 Smart Metabolites Database™ 判别农作物的产地

伪造农作物产地是农业和食品业界普遍存在的问题，至今，为了判别农作物产地研发出了各种各样的方法。通常，很难根据农作物的天然生物学信息（遗传信息等）来进行判别，因此会使用农作物后天体内存在的物质信息进行判定。其中，以农作物从体外摄取的元素信息为基础进行判别的方法具有极好的稳定性，并且已经研究多年。

另一方面，在农作物体内后天产生的物质中，除了从体外摄取的物质之外，还存在体内产生的物质，例如氨基酸、有机酸、脂肪酸和糖。这些物质被统称为代谢物，是许多农作物中共含有的物质，但即使是相同品种的农作物，其代谢物浓度也会随着周围环境的变化而动态变化。如果可以全面测定此类作物中的代谢物，以找到特定农作物品种中各产地代谢物的浓度比率，则有可能将其用于农作物产地判别。

在本文中，使用可以同时测量 377 种亲水性代谢物成分的 Smart Metabolites Database，对共 106 种国内外芦笋样品中的代谢物进行了测量，并建立了判别国内外产品的模型，其判定精度可达 90% 左右。

T. Sakai



#### 芦笋代谢物的提取与衍生

准备了 58 个国内芦笋样品和 48 个国外芦笋样品。

将芦笋切成适当大小减小尺寸，并冷冻干燥。冷冻干燥后对其进行粉碎处理以获得粉末，使用基于 Bligh & Dyer 方法的预处理方案进行提取和衍生化。有关 Bligh & Dyer 的方法的详细信息，请参阅目录中的“代谢组学预处理手册”（C146-2181）。

另外，将 Ribitol 用作内标。

\* 样品由独立行政法人农林水产消费安全技术中心（FAMIC）提供，他们在预处理方面也给予了大力协助。

#### 衍生化亲水代谢物的测定

通过 GC-MS/MS 对衍生化后的样品溶液进行测量。分析条件基于我司 Smart Metabolites Database 的内容。表 1 显示了详细条件。

表 1 测定条件

GC-MS	: GCMS-TQ™8040 NX
色谱柱	: BPX-5 (30 m x 0.25 mm, 0.25 μm)
- GC -	
进样模式	: 分流 (30 : 1)
气化室温度	: 250 °C
柱温箱温度	: 60 °C (2 min) → (15 °C /min) → 330 °C (3min)
控制模式	: 线速度 (39.0 cm/sec)
吹扫流量	: 5.0 mL/min
- MS -	
采集模式	: MRM
离子源温度	: 200 °C
接口温度	: 280 °C
事件时间	: 0.25 sec

#### 峰检测

使用我司的分析软件 LabSolutions Insight™ (ver.3.5) 进行了峰检测。对峰检测设置了以下规则。

##### ★ 规则 1:

在保留指标预测的保留时间 ±0.08 分钟以内，如果定量离子和定性离子同时形成峰，且定量离子的峰高为 1000 以上，则定义为“检测到”（图 1）。

##### ★ 规则 2:

对于有半数以上数据显示“检测到”的化合物，即使某些数据不符合规则 1，也可以将接近预测保留时间的峰确定为“检测到”（为避免“未检测到”数据是丢失值或 0：图 1）。

##### ★ 规则 3:

如果在预测保留时间附近有两个以上应“检测到”的峰，原则上将更接近的一个峰定义为“检测到”，但如果因为样品而存在差异，则将半数以上样品中显示为“检测到”的峰定义为“检测到”（为了避免样品不同“检测到”不同峰：图 2）。

LabSolutions Insight 具有最佳的峰检测功能，即使需要进行大量数据（例如机器学习）的分析，也可以在短时间内完成。另外，此次通过使用 GC-MS/MS 进行测定，发现了许多峰型较好的峰，进而检测到许多物质（共 217 种成分）。

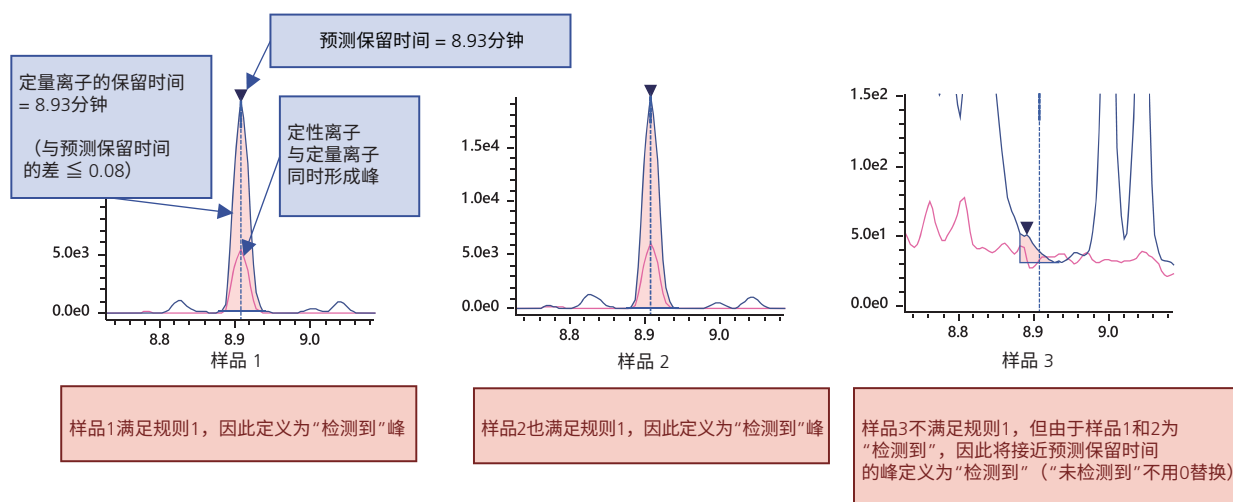


图1 规则1和规则2示例

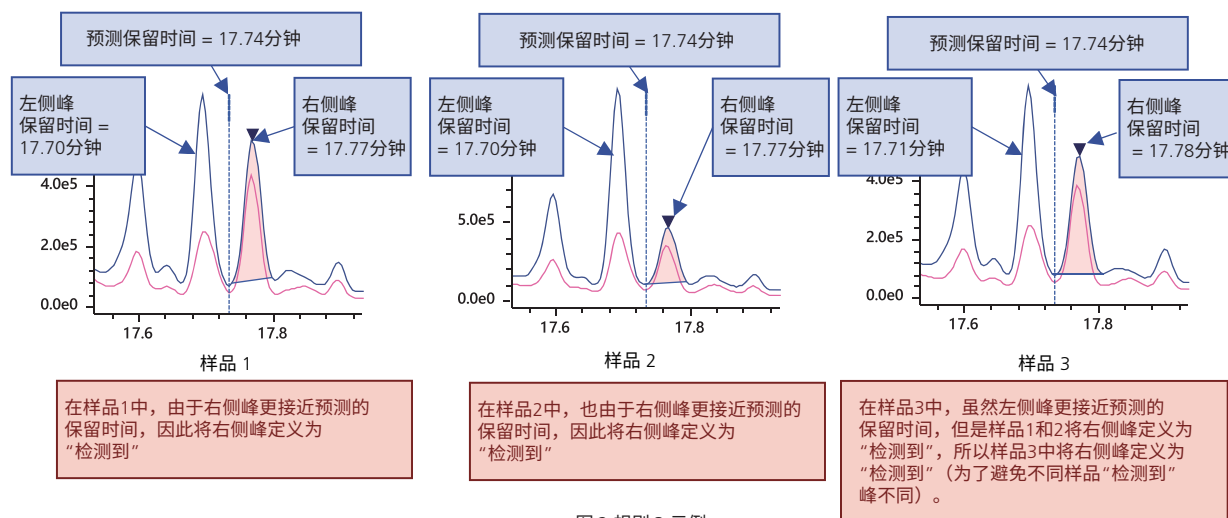


图2 规则3示例

## 创建产地判别模型

在完成峰检测之后, 将每个峰的高度作为数据矩阵输出, 并执行与应用报告 M282 中所述相同的数据预处理。使用上述规则进行峰检测的结果是, 该数据中没有包含缺失值的数据矩阵, 因此省略了缺失值处理。对于内标峰低于标准值的样品, 认为其在衍生化步骤中存在缺陷, 已从数据矩阵中删除。对于其他样品, 使用的是将每个峰面积值除以内标的峰面积值并通过 z 分数归一化而获得的数据。

与应用报告 M282 一样, 将所有样品随机分为训练组和测试组, 并确认了检测到峰的 217 种成分的箱形图和异常值分布, 选择了对判别有效的 13 种成分。使用这 13 种成分, 用随机森林算法创建了一个产地判别模型。将样品中的训练组和测试组随机交换 50 次, 并分别使用此处创建的模型计算预测精度, 发现模型精度的均值为 91.7%。图 3 显示了该模型典型的 Confusion Matrix 和 ROC 曲线。

该实验表明, 亲水性代谢物的全面测量可用于判别农作物的产地。

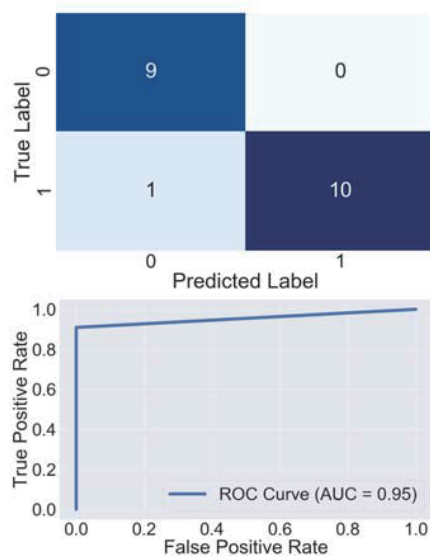


图3 芦笋产地判别模型中的 Confusion Matrix (上) 和 ROC 曲线 (下)

岛津应用云

Smart Metabolites Database、GCMS-TQ 和 LabSolutionsInsight 是岛津制作所株式会社在日本及其他国家所使用的商标。此外, 本文中出现的公司名称和产品名称是各公司的商标及注册商标。本文中可能对“TM”和“®”进行了省略。



岛津企业管理(中国)有限公司  
岛津(香港)有限公司

<http://www.shimadzu.com.cn>

用户服务热线电话: 800-810-0439  
400-650-0439

免责声明:

\* 本资料未经许可不得擅自修改、转载、销售;  
\* 本资料中的所有信息仅供参考, 不予任何保证。  
如有变动, 恕不另行通知。

第一版发行日: 2019年10月