Submission ID #: 61529
Scriptwriter Name: Bridget Colvin
Project Page Link: https://www.jove.com/account/file-uploader?src=18766393

Title: Chromatographic Fingerprinting by Template Matching for Data Collected by Comprehensive Two-Dimensional Gas Chromatography

Authors and Affiliations: Federico Stilo[1], Chiara Cordero[1], Carlo Bicchi[1], Daniela Peroni[2], Qingping Tao[3], and Stephen E. Reichenbach[3,4]

[1]Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino
[2]SRA Instruments
[3]GC Image, LLC
[4]Computer Science and Engineering Department, University of Nebraska

Corresponding Authors:
Chiara Cordero
chiara.cordero@unito.it

Stephen E. Reichenbach
reich@cse.unl.edu

Co-Authors:
federico.stilo@unito.it
carlo.bicchi@unito.it
peroni@srainstruments.com
qtao@gcimage.com

Author Questionnaire

1. **Microscopy**: Does your protocol require the use of a dissecting or stereomicroscope for performing a complex dissection, microinjection technique, or similar? **N**

2. **Software:** Does the part of your protocol being filmed demonstrate software usage? **Y**

   Videographer: All screen capture files provided; do not film

3. **Interview statements:** Considering the Covid-19-imposed mask-wearing and social distancing recommendations, which interview statement filming option is the most appropriate for your group? **Please select one**.

   ☒ Interviewees wear masks until the videographer steps away (≥6 ft/2 m) and begins filming. The interviewee then removes the mask for line delivery only. When the shot is acquired, the interviewee puts the mask back on. Statements can be filmed outside if weather permits.

   ☒ Interviewees self-record interview statements outside of the filming date. JoVE can provide support for this option.

4. **Filming location:** Will the filming need to take place in multiple locations (greater than walking distance)? **N**

Protocol Length
Number of Shots: **41**

Introduction

1. Introductory Interview Statements

REQUIRED:

1.1. **Chiara Cordero**: Chromatographic fingerprinting by template matching is significant, because it can automatically extract rich untargeted and targeted chemical information from GC-by-GC chromatograms across multiple samples with a consistent cross-alignment **[1]**.

1.1.1. LAB MEDIA: Named talent says the statement above in an interview-style shot, looking slightly off-camera

REQUIRED:

1.2. **Stephen E. Reichenbach**: The main advantages of this technique are its comprehensive coverage of untargeted and targeted analytes and its consistent matching of chromatographic features across multiple chromatograms and samples by cross-alignment **[1]**.

1.2.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera

OPTIONAL:

1.3. **Federico Stilo**: The technique has broad applications, but here we apply it to foodomics - specifically, to characterize the chemical information encrypted in the volatile fraction, or volatilome, of extra-virgin olive oils **[1]**.

1.3.1. LAB MEDIA: Named talent says the statement above in an interview-style shot, looking slightly off-camera

OPTIONAL:

1.4. **Stephen E. Reichenbach**: The technique is foundational for research using machine learning to discover complex relationships between sample chemistry and external factors, for example, in food, using sensory quality and production methods **[1]**.

1.4.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera

OPTIONAL:

1.5. **Federico Stilo**: A good quality chromatography is essential. If the separations are poor, the power of the technique is limited, but if the chromatograms are good, the method is fairly straightforward **[1]**.

1.5.1. LAB MEDIA: Named talent says the statement above in an interview-style shot, looking slightly off-camera

OPTIONAL:

1.6. **Chiara Cordero**: GC-by-GC visualizes structure-retention relationships in images. The human eye is well adapted to perceive two-dimensional patterns, so seeing the method demonstrates its power in a very direct manner **[1]**.

1.6.1. LAB MEDIA: Named talent says the statement above in an interview-style shot, looking slightly off-camera

Protocol

2. Importing Raw Data

2.1. To create a two-dimensional raster array for visualization and processing, launch the image software **[1]** and select **File** and **Import** to open the raw data files of interest **[2]**.

2.1.1. WIDE: Talent launching software

2.1.2. SCREEN: 2-1-2: 00:03-00:09

2.2. In the **Import** dialog, set the **Modulation Period** to 3.5 seconds, click **OK**, and select **File** and **Save Image As [1]**.

2.2.1. SCREEN: 2-2-1: 00:01-00:32 *Video Editor: please speed up*

2.3. Navigate to the desired folder, enter the file name, and click **Save [1]**.

2.3.1. SCREEN: 2-3-1: 00:01-00:10

2.4. To shift the modulation phase, select **Processing** and **Shift Phase** and set the **Shift Amount** to minus 0.8 seconds **[1]**.

2.4.1. SCREEN: 2-4-1: 00:04-00:14

3. Baseline Correction

3.1. For baseline correction, select **Graphic** and **Draw Rectangle [1]** and click-and-drag to draw a rectangle in the image in which no peaks are detected **[2]**.

3.1.1. WIDE: Talent selecting Graphic and Draw Rectangle, with monitor visible in frame

3.1.2. SCREEN: 3-1-2 and 3-2-1 and 3-3-1: 00:09-00:13

3.2. Select **Tools** and **Visualize Data** and note the mean and standard deviation of the detector signal **[1]**.

3.2.1. SCREEN: 3-1-2 and 3-2-1 and 3-3-1: 00:14-00:23 *Video Editor: please emphasize mean and standard deviation when mentioned*

3.3. Then close the tool and selecting **Processing** and **Correct Baseline [1]**.

3.3.1. SCREEN: 3-1-2 and 3-2-1 and 3-3-1: 00:24-00:38 *Video Editor: can speed up*

4. Chromatographic Image Coloring and 2D Peak Analyte Detection

4.1. To color the chromatographic images using value and color maps, select **View** and **Colorize [1]**.

4.1.1. WIDE: Talent selecting View and Colorize, with monitor visible in frame

4.2. Open the **Import-Export** tab and select the **number AAAA** custom color map and click **Import [1]**.

4.2.1. SCREEN: 4-2-1 and 4-3-1: 00:03-00:18

4.3. Then, in the **Value Mapping** controls, set the value range to the minimum and maximum values and click **OK [1]**.

4.3.1. SCREEN: 4-2-1 and 4-3-1: 00:20-00:42 *Video Editor: please speed up*

4.4. For 2D peak analyte detection, select **Processing** and **Detect Blobs**. Some peaks will be split and some spurious detections will be observed **[1]**.

4.4.1. SCREEN: 4-4-1 and 4-5-1- and 4-6-1: 00:02-00:12 *Video Editor: please emphasize split peaks and spurious detections when mentioned*

4.5. Select **Configure**, **Settings**, and **Blob Detection**. Set **Smoothing** to 0.1 for 1D samples and to 2 for 2D modulations and set the **Minimum** to 1 x 10^6 **[1]**.

4.5.1. SCREEN: 4-4-1 and 4-5-1- and 4-6-1: 00:14-00:36 *Video Editor: please speed up*

4.6. Then click **OK**, select **Processing** and **Detect Blobs**, and observe the improvements **[1]**.

4.6.1. SCREEN: 4-4-1 and 4-5-1- and 4-6-1: 00:37-00:46

5. 2D Peak Filtration and Peak Spectra Location

5.1. To remove meaningless detections, first select **Processing** and **Interactive Blob Detection [1]**.

5.1.1. WIDE: Talent selecting Processing and Interactive Blob Detection

5.2. Note the blob detection settings and click **Detect [1]**.

5.2.1. SCREEN: 5-2-1 and 5-3-1 and 5-4-1: 00:02-00:10

5.3. In the **Advanced Filter** builder, click **Add**. In the **New Constraint** window, select **Retention two** and click **OK [1]**.

5.3.1. SCREEN: 5-2-1 and 5-3-1 and 5-4-1: 00:11-00:15

5.4. Then, in the **Constraint** sliders, set the minimum and maximum 2D retention times for the filter to reduce the number of false peaks without losing the true peaks and click **Apply** and **Yes [1]**.

5.4.1. SCREEN: 5-2-1 and 5-3-1 and 5-4-1: 00:16-36 *Video Editor: please speed up*

5.5. To search for the peak spectra in an NIST MS (nist M-S) library of interest, select **Configure**, **Settings**, and **Search Library [1]** and set the **Type of Spectrum** to Peak MS, the **Intensity Threshold** to 100, the **NIST Search Type** to Simple Similarity, the **NIST Retention Indices Column Type** to Standard Polar, and the **NIST Retention Indices Tolerance** to 10 **[2]**.

5.5.1. SCREEN: 5-5-1: 00:01-00:09 **TEXT: NIST: National Institute of Standards and Technology; MS: mass spectrometry**
5.5.2. SCREEN: 5-5-1: 00:10-00:36 *Video Editor: please speed up*

5.6. Then click **OK** and select **Processing** and **Search Library for All Blobs [1]**.

5.6.1. SCREEN: 5-6-1: 00:05-00:10

6. Targeted Peak Template Creation and Template Matching and Application

6.1. To create a template with targeted peaks, in the image view **[1]**, click on the first peak and control click on any additional peaks to select the peaks of interest **[2]**.

6.1.1. WIDE: Talent opening image view, with monitor visible in frame
6.1.2. SCREEN: 6-1-2 and 6-2-1: 00:02-00:12

6.2. Click **Add to Template** and select **File** and **Save Template**. Then specify the folder and file name and click **Save [1]**.

6.2.1. SCREEN: 6-1-2 and 6-2-1: 00:20-00:54 *Video Editor: please speed up*

6.3. To match and apply the template, select **File** and **Open Image** and navigate to and open the chromatogram file of interest **[1]**.

6.3.1. SCREEN: 6-3-1 up to 6-9-1: 00:02-00:14

6.4. Set the cursor mode to **Template** and **Select Objects** and select **Template** and **Load Template [1]**.

6.4.1. SCREEN: 6-3-1 up to 6-9-1: 00:18-00:23

6.5. In the **Load Template** window, click **Browse** and open the targeted peaks template. The click **Load** and **Dismiss [1]**.

6.5.1. SCREEN: 6-3-1 up to 6-9-1: 00:24-00:31

6.6. In the Image view, right-click on a template peak to inspect its object properties, including the qualifier chemical logic expression or qCLIC (Q-click) and reference MS **[1]**.

6.6.1. SCREEN: 6-3-1 up to 6-9-1: 00:32-00:50 *Video Editor: please speed up*

6.7. Select **Template** and **Interactive Match and Transform Template**. In the **Match Template** window, click **Match All** and review the matching results in the table and in the image. Each template peak will be marked with an unfilled circle. If a match is made, there will be a link to a filled circle for the detected peak **[1]**.

6.7.1. SCREEN: 6-3-1 up to 6-9-1: 01:06-01:30 *Editor: please emphasize filled circle(s) when mentioned*

6.8. Then edit the matches as appropriate and click **Apply** to transfer the metadata from the template to the chromatogram **[1]**.

6.8.1. SCREEN: 6-3-1 up to 6-9-1: 01:50-01:57

7. Template Transformation (Optional)

7.1. If chromatographic conditions vary substantially, causing the template to be misaligned with a new chromatogram, load the **Targeted template 2.bt** file **[1]** and select **Template** and **Interactive Match Template** and click **Edit Transform [2]**.

7.1.1. WIDE: Talent loading Targeted template 2.bt file, with monitor visible in frame
7.1.2. SCREEN: 7-1-2 and 7-2-1 and 7-3-1: 00:29-00:32

7.2. In the **Transform Template** interface, vary the 1D and 2D scales, translations, and shears

to better align the template with the detected peaks and click **Transform Template [1]**.

7.2.1. SCREEN: 7-1-2 and 7-2-1 and 7-3-1: 00:32-00:54 *Video Editor: please speed up*

7.3. When the template has been transformed, click **Edit Match** and match the template peaks as demonstrated **[1]**.

7.3.1. SCREEN: 7-1-2 and 7-2-1 and 7-3-1: 00:54-01:02

8. Untargeted and Targeted Chromatogram Analysis

8.1. To establish correspondences between untargeted and targeted analytes, select **File** and **Open analysis [1]** and open the file as indicated **[2]**.

8.1.1. WIDE: Talent selecting File and Open analysis, with monitor visible in frame
8.1.2. SCREEN: 8-1-2- and 8-2-1 up to 8-4-1: 00:03-00:23 *Video Editor: please speed up*

8.2. Open the **Compounds** tab to review the metric values and statistics for specific analytes or untargeted analytes with identifiers aligned across all of the chromatograms **[1]**.

8.2.1. SCREEN: 8-1-2- and 8-2-1 up to 8-4-1: 00:36-00:43

8.3. Open the **Attributes** tab to review values and statistics for specific metrics across the chromatograms **[1]** and open the **Statistical Summary** tab to review the summary statistics for both compounds and features **[1]**.

8.3.1. SCREEN: 8-1-2- and 8-2-1 up to 8-4-1: 00:44-01:08 *Video Editor: please speed up*
8.3.2. SCREEN: 8-1-2- and 8-2-1 up to 8-4-1: 01:09-01:40 *Video Editor: please speed up*

8.4. If the chromatograms are from different classes, the **Summary** tab will list Fisher ratio statistics that provide insight into the features for discriminating between classes **[1]**.

8.4.1. SCREEN: 8-1-2- and 8-2-1 up to 8-4-1: 01:53-02:00

Protocol Script Questions

A. Which steps from the protocol are the most important for viewers to see?
n/a

B. What is the single most difficult aspect of this procedure and what do you do to ensure success?
4.4.1. Detecting blobs is fundamental to delineating 2D patterns and capturing metadata to create a consistent template for cross-samples analysis.
6.8.1. Template transformation is an interactive procedure that requires analyst supervision to understand the actual misalignment between a reference template and a sample's pattern. The tuning of transform parameters ensures proper re-alignment of the template peaks and a low false-negative matching rate.

Results

9. Results: Representative Chromatographic Fingerprinting

9.1. In these graphs, relative retention patterns for homologous series and classes are shown **[1]** with annotations for linear saturated hydrocarbons **[2]**, unsaturated hydrocarbons **[3]**, linear saturated aldehydes **[4]**, mono-unsaturated aldehydes **[5]**, polyunsaturated aldehydes **[6]**, primary alcohols **[7]**, and short-chain fatty acids **[8]**.

9.1.1. LAB MEDIA: Figure 1A
9.1.2. LAB MEDIA: Figure 1A *Video Editor: please emphasize black circles*
9.1.3. LAB MEDIA: Figure 1A *Video Editor: please emphasize yellow circles*
9.1.4. LAB MEDIA: Figure 1A *Video Editor: please emphasize blue circles*
9.1.5. LAB MEDIA: Figure 1A *Video Editor: please emphasize red circles*
9.1.6. LAB MEDIA: Figure 1 *Video Editor: please emphasize salmon circles*
9.1.7. LAB MEDIA: Figure 1 *Video Editor: please emphasize green circles*
9.1.8. LAB MEDIA: Figure 1 *Video Editor: please emphasize cyan circles*

9.2. Detected 2D peaks can be identified **[1]** by comparing the average MS spectrum extracted from the entire 2D peak **[2]** or from the largest spectrum **[3]**.

9.2.1. LAB MEDIA: Figure 2
9.2.2. LAB MEDIA: Figure 2 *Video Editor: please emphasize blob spectra*
9.2.3. LAB MEDIA: Figure 2 *Video Editor: please emphasize apex spectra*

9.3. The collection of identified 2D peaks can be adopted to build a template of targeted peaks to promptly establish reliable correspondences between the same compound across all of the sample chromatograms **[1]**.

9.3.1. LAB MEDIA: Figure 1B

9.4. In this analysis, a partial misalignment between the targeted template and the actual chromatogram can be observed **[1]**.

9.4.1. LAB MEDIA: Figure 3 top chromatogram

9.5. For minimal misalignments, interactive template transformations can reposition the template peaks for a better fit **[1]**.

9.5.1. LAB MEDIA: Figure 3 top chromatogram *Video Editor: please emphasize sequentially add/replace previous figure with second, third, and fourth chromatograms*

9.6. The untargeted feature template is composed of 2D peaks from analytes detected in the composite chromatogram that are matched by the reliable-peaks template **[1]**.

9.6.1. LAB MEDIA: Figure 5A

9.7. The mass spectra of the composite peaks, as well as their retention times, are then recorded in the feature template **[1]**.

9.7.1. LAB MEDIA: Figure 5B

9.8. When unsupervised pattern recognition by Principal Component Analysis is applied to targeted peaks distribution within the 20 analyzed samples **[1]**, Sicilian and Tuscany oils cluster separately **[2]**, suggesting that pedoclimatic conditions and terroir impact the relative prevalence of volatiles **[3]**.

9.8.1. LAB MEDIA: Figure 6 *Video Editor: please emphasize Figure 6B*
9.8.2. LAB MEDIA: Figure 6 *Video Editor: please emphasize Origin Sicily oval and text box in Figure 6A*
9.8.3. LAB MEDIA: Figure 6 *Video Editor: please emphasize Origin Sicily oval and text box in Figure 6B*

Conclusion

10. Conclusion Interview Statements

10.1. **Stephen E. Reichenbach**: When first learning the procedure, it is important to check the results at each step. If you have any questions, the people at GC Image are very responsive and helpful **[1]**.

10.1.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera (3.3., 4.6., 5.6., 6.8.)

10.2. **Chiara Cordero**: We are currently using the technique to analyze other food products, such as hazelnuts, cocoa, and teas, and bio-fluids, such as saliva, urine, and feces, for nutrimetabolomic applications **[1]**.

10.2.1. LAB MEDIA: Named talent says the statement above in an interview-style shot, looking slightly off-camera