Metabolomics in agricultural research

Expanded applications and database capabilities for volatile compound capture and tracking

Kirsten Skogerson, Ph.D.
Analytical Biochemist, Monsanto
Outline

• Background
• Vineyard study design
• Sampling & detection methods
• Database construction
• Vineyard study results
Volatile Organic Compounds (VOCs)

- Small (C5-C20)
- Low boiling points & high vapor pressures

- Importance
  - Chemical ecology
  - Flavor / fragrance
  - Food authentication
  - Medical testing
Plant VOCs are complex mixtures

**Protection against:**
- Abiotic stress (thermotolerance and photoprotection)
- Herbivores
- Pathogens
- Competitors

**Attraction of:**
- Pollinators
- Mutualistic microbes
- Seed dispersers
- Predators/parasitoids

- Herbivores
- Parasitoids
- Pathogens

**Within-plant signaling**

**Allelopathy**

**Between-plant signaling**
Field applications for volatile biomarkers

- Real-time measures
- Early stress & disease detection
- Maturity & quality assessments
- Improved resource allocation

Small-scale studies: blueberry maturity, citrus canker disease detection, walnut tree drought stress, tomato plant damage
Traditional plant VOC trapping & detection

- Sampling systems require plant enclosure, pumps, plumbed gases & electricity
- Small sample sizes
- GC-MS detection
- ~20-80 compounds reported (identified)
Field study challenges

- Large sample numbers required
- GC-MS datasets complex (hundreds vs. dozens of peaks)
- Publically available VOC compound databases not available in searchable formats

BETTER SAMPLING, ANNOTATION, & DATABASE TOOLS NEEDED
Vineyard Volatiles: Cabernet Sauvignon fruit maturity

*Dominus Estate – Christian Moueix, Tod Mostero, Jean-Marie Maureze, Michelle Beyer*
Grape maturity

- Wine quality
- Viticultural practices
- Resource allocation during harvest

- Destructive measures are time/labor intensive
  - Physical – mass, deformability
  - Chemical – sugar, acids, phenolics
  - Sensory – color, aroma, flavor
Vine VOC Project

**VOC profiling (GC-MS)**

Standard grape maturity indices

**TRAIN E-NOSE TO DETECT BIOMARKERS**

**Maturity Biomarkers**
Study Design

- 3 Cabernet Sauvignon blocks
  - 6 vines per block & background air controls
- VOCs sampled daily for 12 weeks
- Grape maturity indices final 8 weeks

~1200 VOC samples (2008 season)
Volatile compound capture

Twister® Stir Bar

Glass
Magnet
PDMS

GERSTEL, Inc.

Solid Phase Micro Extraction
SPME Fiber

Twister® capacity **100-1000 fold greater** than polydimethylsiloxane (PDMS)-coated SPME fiber

Methods
Twister® Flexibility

Active / passive
Lab, greenhouse, field
(Intact fruit)
Easy to deploy off-site

Insert Twister
directly into flesh

Field sampler sent to collaborator

Greenhouse (bag enclosure)

HSSE

In-field Twister
VOC sampling solution – 24h passive sampling

Twister®
Thermal Desorption Unit

Methods

Inlet

Heated zone

Cryo-cooled zone

external markers loaded into capillary tube

Twister®

TDU graphics courtesy of GERSTEL, Inc.
Retention Index Markers

- Absolute RT shifts as function of column length, age & sample type (or injector errors)

- Fatty acid methyl ester (FAME) RI markers required for data auto-annotation (C4-C24)
VOC Analysis (GC-TOF-MS)

Methods

- Leco Pegasus IV / ChromaTOF software
- 35 min cycle time
Vineyard VOC Project – representative data
Challenges

• Hundreds, thousands of samples
• Complex chromatograms
• Manual annotation tedious, error-prone
• Databases of VOC compounds static and not easily searchable
• Unidentified peak information easily lost
Data annotation options

**Agilent Technologies**

Mass Profiler Professional Software

**Waters**

MetaboLynx XS
MarkerLynx

**Leco**

ChromaTOF

**SPECTCONNECT**

*Systematic Identification of Conserved Metabolites in GC/MS Data for Metabolomics Discovery*

Mark P. Styczynski, Joel F. Moxley, Lily V. Tong, Jason L. Walther, Kyle L. Jensen, and Gregory N. Stephanopoulos

Department of Chemical Engineering, Massachusetts Institute of Technology, Room 8-489c, Cambridge, Massachusetts 02139

**BIOINFORMATICS APPLICATIONS NOTE**

*Data and text mining*

MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data

Mikko Katajamaa¹, Jarkko Miettinen² and Matej Orešič²,*

¹Turku Centre for Biotechnology, Turku, Finland and VTT Technical Research Centre of Finland, Espoo, Finland

**TagFinder**

*Data and text mining*

TagFinder for the quantitative analysis of gas chromatography - mass spectrometry (GC-MS) based metabolite profiling experiments

Alexander Luedemann, Katrin Strassburg, Alexander Erban and Joachim Kopka*

Max Planck Institute of Molecular Plant Physiology, Department Prof. L. Willmitzer, Am Muehlenberg 1, D-14476 Potsdam-Golm, Germany.
Metabolite Database
http://binbase.sourceforge.net

• Only annotation software with databasing capabilities
  – Track & store unique molecules (dynamic system)
  – Comprehensive data annotations for all studies

• Assignments supported by Fiehn Metabolite Library (1000+ metabolites)

• BinBase framework adaptable for VOCs

Gert Wohlgemuth – BinBase Programmer
BinBase Algorithm

- RT → RI calculation
- Multi-tiered filtering system
  - RI
  - Unique ion
  - MS similarity
  - Peak purity
  - S/N
- New Bin generation possible with class filter
Database

Total ion intensity

Time (s)

300 500 700 900

Peak #122

Select ion intensity

Time (s)

400 404 408 412

deconvoluted MS of peak #122

RI = 446700
unique ion = 93
signal/noise = 2013
purity = 0.1137
apex masses = 98+31+39+...

1500 Bin database

RI Filter

8 possible matches

Unique ion filter

2 possible matches

Linalool
MS match 917

Terpinolene
MS match <500
### BinBase Reports

<table>
<thead>
<tr>
<th>compound name</th>
<th>retention index</th>
<th>PubChem CID</th>
<th>user provided data</th>
</tr>
</thead>
<tbody>
<tr>
<td>hexanol (2z)</td>
<td>390009 43</td>
<td>47178 36:68.0</td>
<td></td>
</tr>
<tr>
<td>heptanol (2-)</td>
<td>346144 45</td>
<td>45431 33:208.0 37</td>
<td>10976</td>
</tr>
<tr>
<td>linalool oxide (dehydr)</td>
<td>394631 68</td>
<td>47418 31:440.0 37</td>
<td>-</td>
</tr>
<tr>
<td>hepten-2-one (6-meth)</td>
<td>39009 39009</td>
<td>43718 36:68.0</td>
<td>-</td>
</tr>
<tr>
<td>octanone (2-)</td>
<td>392202 58</td>
<td>45340 36:2.0</td>
<td>-</td>
</tr>
<tr>
<td>pentyl furan (2-)</td>
<td>393793 81</td>
<td>45700 30:250.0</td>
<td>-</td>
</tr>
<tr>
<td>linalool oxide (dehyd)</td>
<td>394631 68</td>
<td>47418 31:440.0 37</td>
<td>-</td>
</tr>
<tr>
<td>hexyl acetate (3e)</td>
<td>398624 67</td>
<td>45926 32:6219.0 3 -</td>
<td>-</td>
</tr>
<tr>
<td>50987</td>
<td>400269 41</td>
<td>50987 35:308.0 31 -</td>
<td>-</td>
</tr>
</tbody>
</table>

**Note:**
- **VOC BinBase Name:** hexanol (2z), heptanol (2-), linalool oxide (dehyd), hepten-2-one (6-meth), octanone (2-), pentyl furan (2-), linalool oxide (dehydr), hexyl acetate (3e), 50987
- **SetupX Class ID:** VOC BB ID, mass spec, PubChem ID
- **Sample Prep:** sample prep, harvest timepoint, vineyard
- **Grape Variety Code:** Pinot grigio
- **Code:** PDA PDA PDA PDA PDA PDA PDB
- **PubChem CID:** CID 5283321, CID 8908, CID 19602, CID 8093, CID 9862, CID 18827, CID 637564, CID 10976
- **Mass Spec:** mass spec
- **First-Pass Annotation:** 'first-pass' annotation
- **Unique Bin ID:** unique Bin id
- **Ion Used for Quantification:** ion used for quantification
- **Replacement Values:** replaced values
BinBase ‘replaced’ values

- Absent peaks or those not passing thresholds yield zeros in results tables

- Raw data access allows for zero replacements

- Quality of replaced values is high
Compound Identification

- Purchased Robert Adams plant compound library
- Convert Adams RI to Fiehn FAME RI equivalent
- Test conversion by injecting pure standards
- 2200 compounds instantly searchable (both MS & RI data)
VOC BinBase Database Contents

- 1.6 million spectra from 3200 samples, but only ~1500 Bins present in the database

- Spectral similarity network + chemical similarity links (blue) for identified compounds (red nodes)

Dinesh Barupal (Fiehn Lab)

Vineyard Volatiles:
Cabernet Sauvignon fruit maturity

Dominus Estate – Christian Mouiex,
Tod Mostero, Jean-Marie Maureze, Michelle Beyer
Study objectives

• Evaluate Twister®-based sampling for large-scale field studies

• Test new VOC BinBase annotation and database tool

• Determine whether changes in vine canopy VOCs correlate with traditional maturity indices
2008-09 Study Design

~ 9-10 dates with winery maturity measures for each block

6 dates with winery maturity measures
Twister® detection
(Applaud pesticide)

Daily averages by block

- Block A
- Block B
- Block C
- Control

Block B
(all 2008 data)

Block B
(detail)
BinBase Annotation

- 2000 samples (2008-09)
- 900 Bins annotated – 171 identified
Principal Component Analysis (900 VOCs)

Unsupervised method shows data structure
Compound trends (terpene/norisoprenoid)

Vine VOC Project

Categ. Box & Whisker Plot: linalool

Box (Standard error) / Whiskers (1.96 x SE)
Compound trends (benzenoid/phenylpropanoid)

Vine VOC Project

Box (Standard error) / Whiskers (1.96 x SE)
Compound trends (esters/acids)

Vine VOC Project

Box (Standard error) / Whiskers (1.96 x SE)
Modeling work – Statistica

- Feature selection used to chose 20 Bin lists
- Modeled Sugar (Brix) & Sugar/TA ratios
- 2008 models and dual-year models constructed
- Backward variable selection for PLS (BVSPLS)
- Simultaneous modeling of multiple indices
Partial least squares regression analysis:
Sugar/TA ratio model (2008 data)

- Sugar/TA ratio ~linear
- Model constructed from 29 VOC profiles on days winery measure also taken
- Model tested with 82 VOC profiles collected on days where winery data not available
- Use interpolated values to assess quality of model predictions

- Prediction error acceptable (<10 Sugar/TA ratio units)
- Prediction error ~15 Sugar/TA units for 2009 samples!!
- Air control samples collected on property (50 m from vines) were not able to predict Sugar/TA values
Dual-year Sugar/TA Model Performance

- Model constructed from 2008 & 2009 winery data
- Model prediction errors similar to 2008 model
- Model predictions improved for 2009 VOC data
- (BVSP1LS best model with RMSEP = 7.87)
Summary

• Unique, large-scale study of plant VOCs to identify biomarkers for agricultural applications

• Twister® employed successfully as passive sampler

• Metabolite BinBase extended to complex volatile mixture analysis & Adams library integrated to speed compound identification

• Feasibility of an in-field grape maturity monitoring system demonstrated

Sampling / data analysis approaches developed should make large-scale field studies accessible
How close are we.....?

Conclusions

- Preliminary results from vineyard project are interesting.
- Project requires additional years data collection (multiple sites, varieties, vintages).
- Future efforts must focus on compound identification.
- Electronic-nose testing could begin with the identified compounds found in this study.
VOC BinBase Download


- .msp file format for uploading into vendor software
- Mass spectra provided with both Fiehn FAME RI & Kovats alkane RI values
- 220 identified, ~1000 unidentified (no artifacts)
Acknowledgements

- Oliver Fiehn
- **Fiehn Lab Members** (Gert Wohlgemuth, BinBase programmer)
- Sue Ebeler
- Florence Negre-Zakharov
- Roger Boulton
- John Yoder
- Dominus Team
  - Christian Moueix
  - Tod Mostero
  - Michelle Beyer
  - Jean-Marie Maureze

Funding

- Crosby Fellowship
- Dominus
- Bioinformatics Summer Grant
- Citrus Research Board
Questions
2009 VOC trends

Week 37
(23.2 Brix)
- ethyl decanoate
- ethyl octanoate
- nonanoic acid
- octanoic acid
- hexenyl isobutanoate (3z–)
- geranyl acetate
- linalyl acetate
- camphene
- acetonyl acetone

Week 38
(24.0 Brix)
- linalool oxides
- octadecanoic acid
- dodecanoic acid
- 4-hydroxy-4-methyl-2-pentanone

Week 39
(24.9 Brix)
- hexenyl butanoate (3z–)
- 6, 9-guaidiene
- α-thujene
- thujone
- ortho-cymene
- octanol (3–)

Week 40
(26.1 Brix)
- methyl biphenyl
- 1,8-cineole
- iso-menthone
- 2-octanone
- 2-heptanone

Dinesh Barupal (Fiehn Lab)
VOC BinBase: Compound ID & Tracking

- FAME RI markers added to each sample for auto-annotation & compound databasing
- MS-RI based compound assignments
  - Adams library 2000+
  - Std injections (coated capillaries)
- Current database stats:
  - 3600 samples (18 species)
  - 1.3 million spectra
  - 1500 unique spectra

Ability to track & catalog unknowns critical in area of complex volatile mixture analysis
Wine Grape Example: 2 varieties, 2 vineyards, 4 harvest dates

- Grape must (SBSE)
- 122 grape-related compounds
- PCA scores plots:
  - Site drives separation on PC1
  - Variety drives separation on PC3
- PCA loadings:
  - Pesticides are key in site differentiation
  - Terpenes key in varietal differentiation