Direct Analysis in Real Time-High Resolution Mass Spectrometry as a Tool for Rapid Plant Species Identification—Applications to Endangered Species Characterization

### **Development and Scaling of Innovative Technologies for Wood Identification**

Seattle Wood Identification Workshop February 28<sup>th</sup> 2017, Seattle, WA



Rabi Ann Musah, Ph.D. (rmusah@albany.edu) Department of Chemistry SUNY-Albany

# **Presentation Overview**

#### • Work on Species Identification

- Project year/Duration
- Funding sources
- Project partners and roles
- Taxa under scrutiny
- Are methods and data open-access and publically available for use?
- DART-HRMS as an Approach to Rapid Species Identification
  - Hypothesis
  - Approach
  - Results
  - Conclusions
- Challenges and Opportunities



### **Work on Species Identification**

- Project year/Duration: ~3 years
- Funding sources: (Indirectly: NSF and NIJ)
- Project partners and roles: Ed Espinoza
- Taxa under scrutiny: Anything...
- Are methods and data open-access and publically available for use? Partially



### Species Identification From Genome Analysis

- If the plant gene has been annotated, DNA analysis can provide species identity information
- The estimate of total plants on earth is ~400,000 species\*
- As of this year, ~180 plant species genomes have been published,\*\* which represents roughly 0.045% of estimated species!
  - Reliance on phenotypic characteristics for the identification of plant species.
  - Very challenging when distinguishing macroscopic features are absent (e.g. in finished products)
  - Very challenging when microscopic features are very similar (e.g. in members of the same genus)

\*http://www.bgci.org/policy/1521/?sec=ourwork&id=1521 <u>\*\*http://www.plabipd.de/portal/sequence-timeline</u>





# Species Identification From the Metabolome?

### Hypothesis:

Since the small molecule profile of an organism ultimately reflects the genes that distinguish it, the information content of the metabolome might be just as well suited as the genome for assessment of genetic relatedness between species, and species identification.



5

## Species Identification by Phytochemical "Fingerprinting"...





### Challenges to Metabolome Characterization

- Difficulty of acquiring a comprehensive small-molecule chemical map of an organism or its parts in real time
- Time-consuming nature of metabolome profiling by conventional methods
- Challenge of obtaining a faithful and consistent representation of defining chemical components or chemical component ratios that is divorced from biases or artifacts introduced by sample processing steps
- How to distinguish between chemicals that define a species, and those that do not provide discriminating information.

UNIVERSITY ATALBANY State University of New York

# Direct Analysis in Real Time-Mass Spectrometry (DART-MS)

### Instrumentation

- IonSense DART-SVP ion source
- JEOL AccuTOF mass spectrometer
   Features
- Open-air, ambient ionization method
- Millimass unit resolution
- Nanogram detection limit
- Direct analysis of samples
- In-source CID
- Only very small amounts of material are required



- Plant material sampled directly
  - Yields a diagnostic chemical signature
  - Enables rapid detection of fingerprint profile



## Direct Analysis in Real Time-Mass Spectrometry (DART-MS)







### Direct Analysis in Real Time-Mass Spectrometry (DART-MS)



### **Proton Transfer Ionization**

He  $(2^{3}S)$  + H<sub>2</sub>O  $\rightarrow$  H<sub>2</sub>O<sup>+•</sup> + He  $(1^{1}S)$  + e<sup>-1</sup>

#### 

# What is the Output of a DART-HRMS Experiment?



### **Proof of Concept: Species** Identification of *Datura* Seeds



WIVERSITY ATALBANY State University of New York

### **Species Identification Using Linear Discriminant Analysis**



## **Proof of Concept:** *Dalbergia* spp. Identification

a Mass spectral heat maps



b Kernel discriminant analysis



C Hierarchical clustering analysis





Musah, Espinoza, Cody, Lesiak, Christensen, Moore, Maleknia, Sci. Rep., 2015, http://www.nature.com/articles/srep11520

### **Proof of Concept: Species of Origin Determination of Biodiesel Feedstocks**

#### a Mass spectral heat maps



#### b Principal component analysis



15

C Hierarchical clustering analysis





Musah, Espinoza, Cody, Lesiak, Christensen, Moore, Maleknia, Sci. Rep., 2015, http://www.nature.com/articles/srepu520

### **Proof of Concept: Fly Species Determination From Puparial Cases**

С

#### a Mass spectral heat maps



#### b Kernel discriminant analysis



16



Musah, Espinoza, Cody, Lesiak, Christensen, Moore, Maleknia, Sci. Rep., 2015, http://www.nature.com/articles/srepu520

### **Proof of Principle: Eucalypt Species Determination From Plant Leaves**





#### b Kernel discriminant analysis



17

C Hierarchical clustering analysis





Musah, Espinoza, Cody, Lesiak, Christensen, Moore, Maleknia, Sci. Rep., 2015, http://www.nature.com/articles/srep11520

### **Creation of Databases of Plant Chemical Signatures Could Enable Rapid Species Identification**





# **Kernel Discriminant Analysis**



UNIVERSITY ATALBANY State University of New York

# **Hierarchical Clustering Analysis**



# **Hierarchical Clustering Analysis**



### Partial Least Squares- Discriminant Analysis



# **Featured Publications**

UNIVERSITY

ATALBANY

State University of New York



23

SCIE	INTIFIC <b>RE</b>	PORTS	ELSEV	Forensic Science International 242 (2014) 210-218 Contents lists available at ScienceDirect Forensic Science International journal homepage: www.elsevier.com/locate/forsciir	at
<b>OPEN</b> Received: 08 February 2015 Accepted: 29 May 2015 Published: 09 July 2015	A High Throughput A Spectrometric Approa Identification and Clar from Chemical Finger Signatures	mbient Mass ach to Species ssification print analytical	Rapid (DAR) specio Ashton <sup>*</sup> Departmen <sup>*</sup> JEOL USA I A R T I C L	I detection by direct analysis in real time-mass spectr T-MS) of psychoactive plant drugs of abuse: The case of osa aka "Kratom" A D. Lesiak <sup>a</sup> , Robert B. Cody <sup>b</sup> , A. John Dane <sup>b</sup> , Rabi A. Musah <sup>a,*</sup> at of Chemistry, University at Albary, State University of New York, 1400 Washington Avenue, Albary, NY 12222, USA Inc., 11 Dearborn Road, Peabody, MA 01960, USA LE INFO ABSTRACT Atticle pubsacs.org/ac	CrossM Mitragyna
Analytical Methods	Earl D. Christensen <sup>4</sup> , Hannah E. Moore <sup>5</sup> , Simin Maleknie A high throughput method for species identification and class processing of direct analysis in real time (DART) mass spectro has been developed. The method entails introduction of samp DART ion source and the mass spectrometer inlet, with the e subjected to unsupervised hierarchical clustering processing. chemotypes are instantaneously detected. The result is identi		entification from Chemical Fingerprints: A High- ion of Direct Analysis in Real Time Mass Phytochemical Por Phytochemical Explore this journal> Phytochemical Phytochemica		
CrossMark Center for determination of the species of origin of processed commercial products— application to psychoactive pepper supplements;		Info R C S J L Fi	Rapid Species-level Identification of <i>Salvias</i> by Chemometric Processing of Ambient Ionisation Spectrometry-derived Chemical Profiles ustine E. Giffen, Ashton D. Lesiak, A. John Dane, Robert B. Cody, Rabi A. Musah irst published: 14 December 2016 Full publication history	Mass View issue TO Volume 28, Issu January/February Pages 16-26	

State University of New York



### Laser Ablation Direct Analysis in Real Time Imaging Mass Spectrometry (LADI-MS)





Fowble, Teramoto, Cody, Edwards, Guarrera & Musah, *Analytical Chemistry*, 2017, *accepted* 

### Small-molecule Spatial Distribution Mapping of Plant Tissue





Fowble, Teramoto, Cody, Edwards, Guarrera & Musah, *Analytical Chemistry*, 2017, *accepted* 

### Conclusions: Advantageous Features of the DART-MS/Statistical Analysis Approach to Species Identification

- Species identification does not in any way require that the identity of the observed peaks be known
- Only very small amounts of material are needed
- The approach is extremely rapid
- Databases can easily be built and continuously expanded
- The approach has broad applicability
- The approach is scalable

### **Potential Challenges...**

Environmental influences

Physiological influences



### Proteome

Genome





- Funding
- Source of authenticated standards
- "Buy-in"



### Acknowledgements









Meghan Fogherty Allix Coon

Back Row: Kristen Fowble, Leif McGoldrick, Denny He, Cameron Longo

Front Row: Ashton Lesiak, Ph.D. Professor Rabi Musah, Justine Giffen

Ed Espinoza (USFWS)

**Funding Sources:** 



Dr. John Dane







UNIVERSITY ATALBANY State University of New York





# Questions?