New method developments in pollen DNA barcoding for forensic applications

Karen Leanne Bell, Emory University

Homeland Security Symposium Series, University of Texas El Paso, June 2016

Overview

- Background and aims
- Testing the DNA metabarcoding method
- Quantification
- Whole genome shotgun approach
- Removal of non-pollen material
- Ongoing method development

Pollen species identification

- Applications:
  - Forensics
  - Allergen monitoring
  - Pollination ecology
  - Determining plant community composition
- Typical sample:
  - Mixture of pollen species
  - May be low quantity

Pollen DNA metabarcoding

- Advantages over microscopic examination:
  - Greater taxonomic resolution

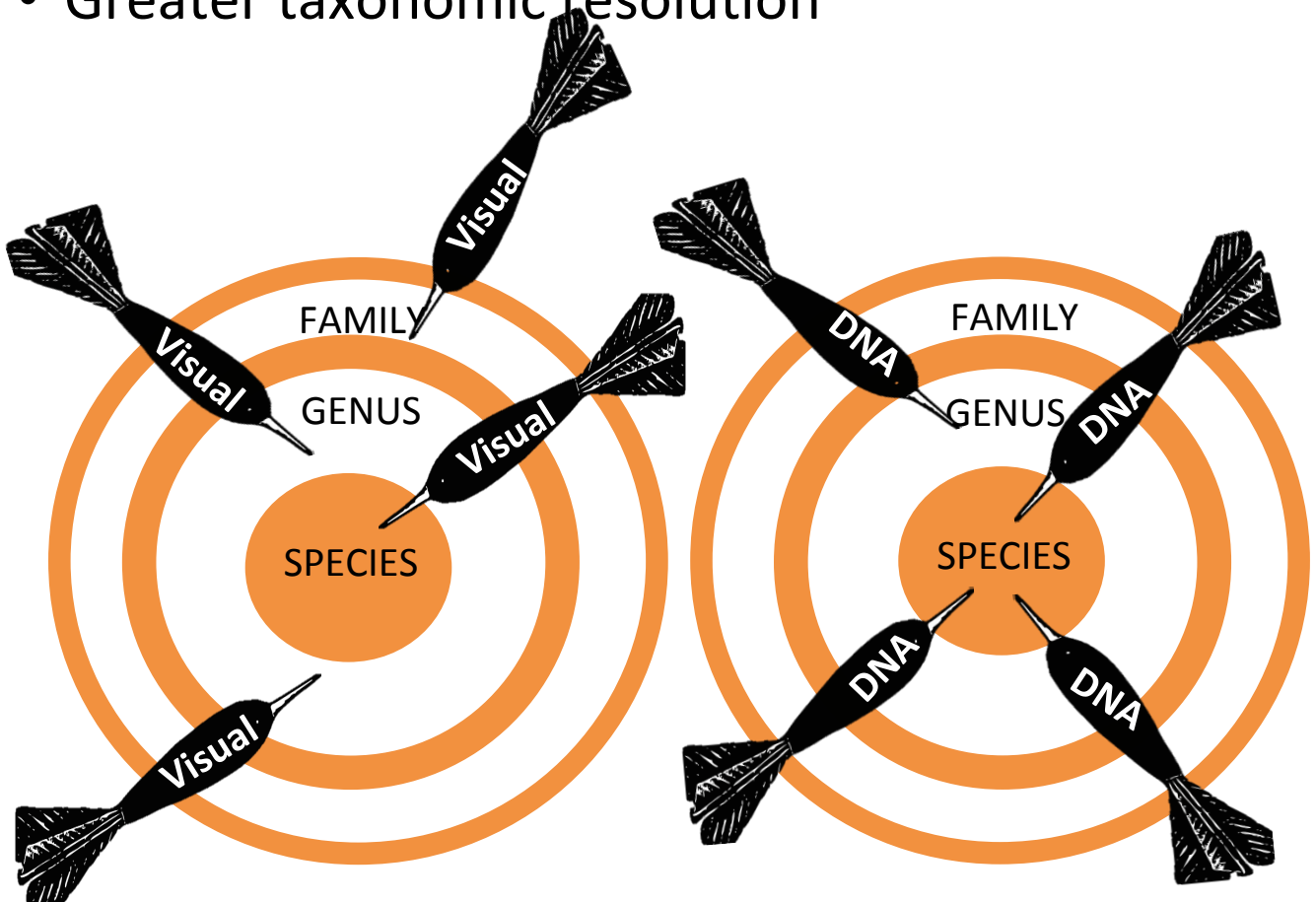


Image: C. Chu

Pollen DNA metabarcoding

- Advantages over microscopic examination:
  - Some species lack features on pollen
  - Few people trained in palynology
    - Many more could be trained in pollen DNA metabarcoding

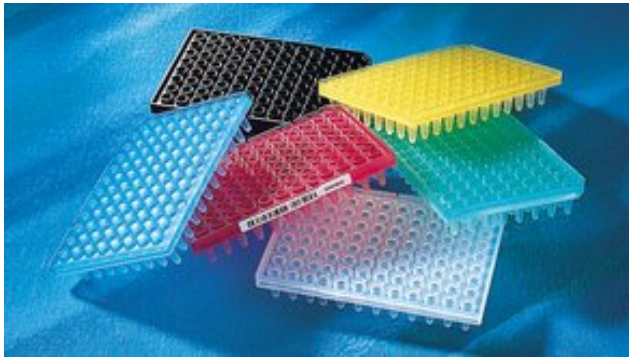
Fighting Crime, With Pollen

America's only full-time forensic pollen analyst has a lot of work on his hands.

"I am swamped with the amount of samples that come in."

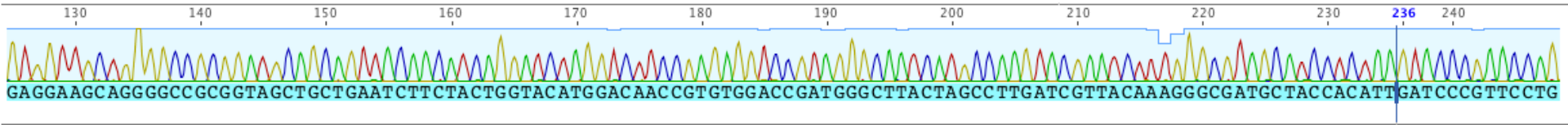
Pollen DNA metabarcoding

- Advantages over microscopic examination:
  - Some species lack features on pollen
  - Few people trained in palynology
  - Faster with large numbers of samples



Pollen DNA metabarcoding

- Advantages over microscopic examination:
  - Some species lack features on pollen
  - Few people trained in palynology
  - Faster with large numbers of samples
- Disadvantages:
  - Not quantitative
  - Species missing from databases
  - No standard bioinformatics pipeline

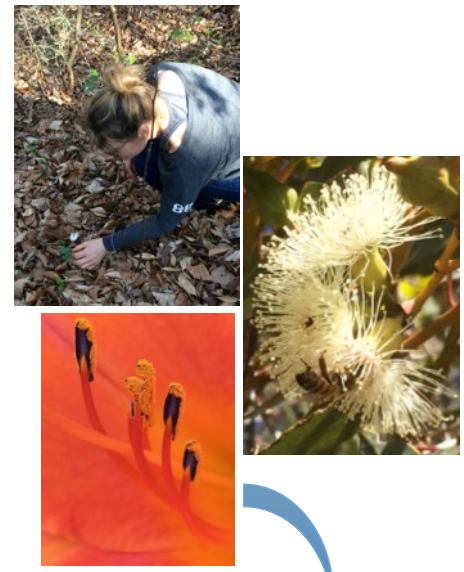


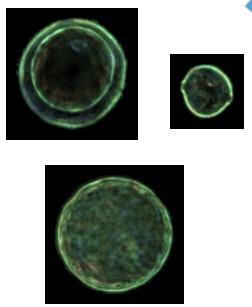
Overview

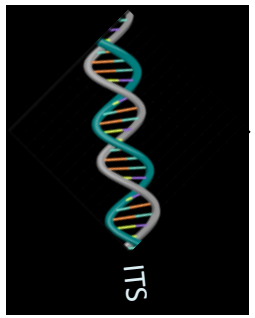
- Background and aims
- **Testing the DNA metabarcoding method**
- Quantification
- Whole genome shotgun approach
- Removal of non-pollen material
- Ongoing method development

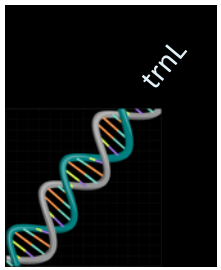
Pollen DNA metabarcoding

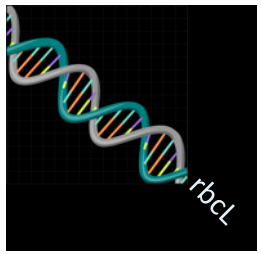
Testing the method with known pollen mixtures

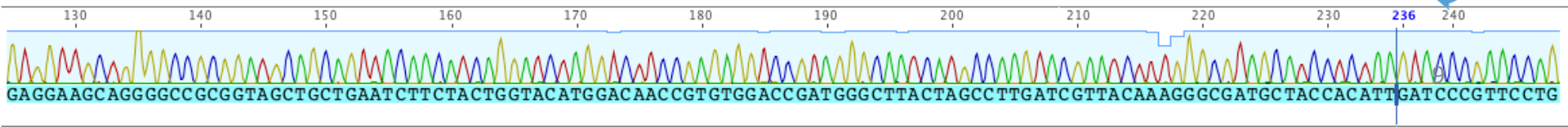








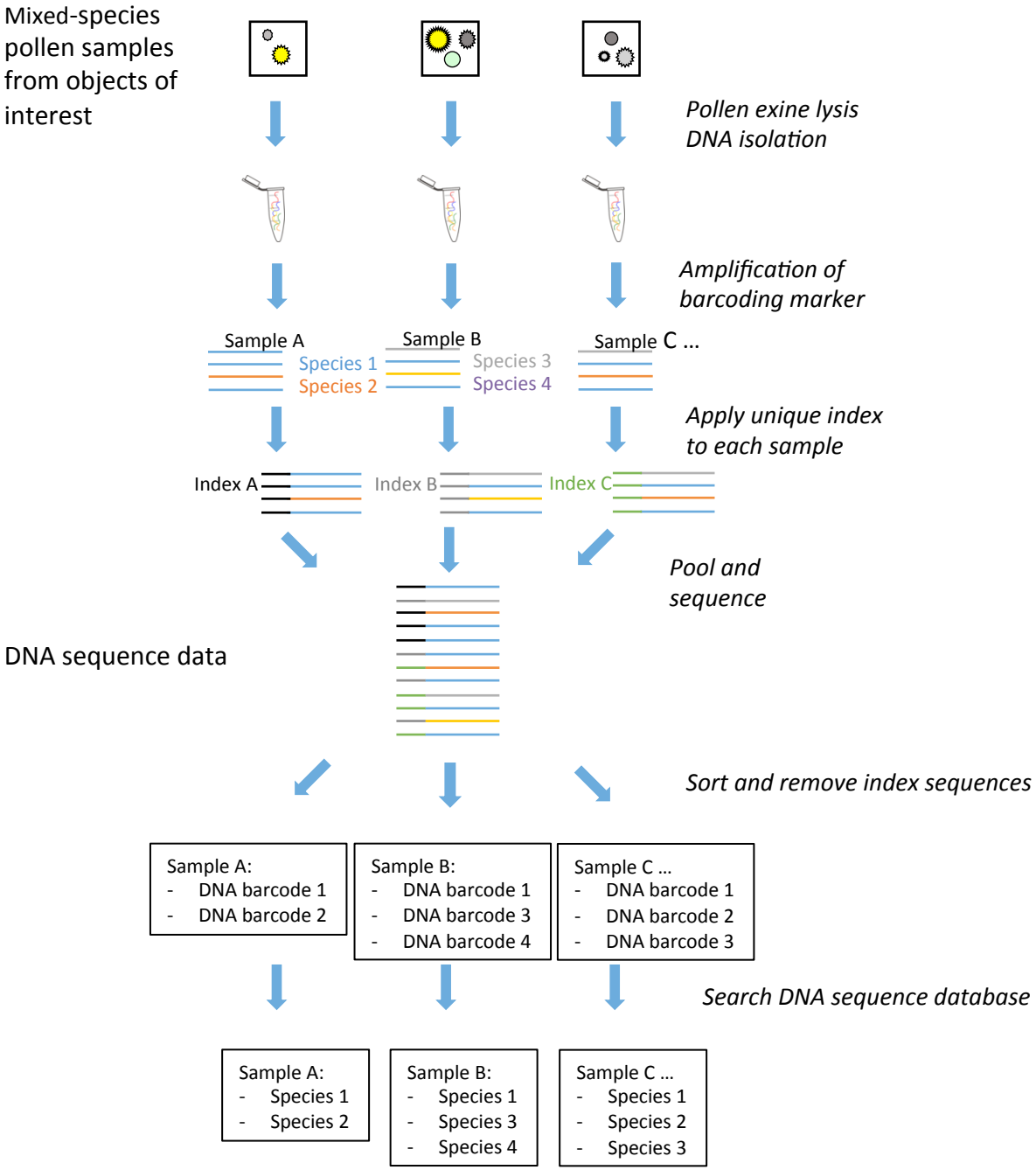




Pollen DNA metabarcoding: Concepts and definitions

- DNA barcoding:
  - The identification of species based on DNA sequence of a standard gene region

- DNA metabarcoding:
  - Simultaneous identification of all species in a mixture using high-throughput sequencing of a standard gene region



Bell, K.L., Burgess, K.S., Okamoto, K.C., Aranda, R., and Brosi, B.J. (2016). *Forensic Science International: Genetics*, 21: 110-116.

Testing the DNA metabarcoding method

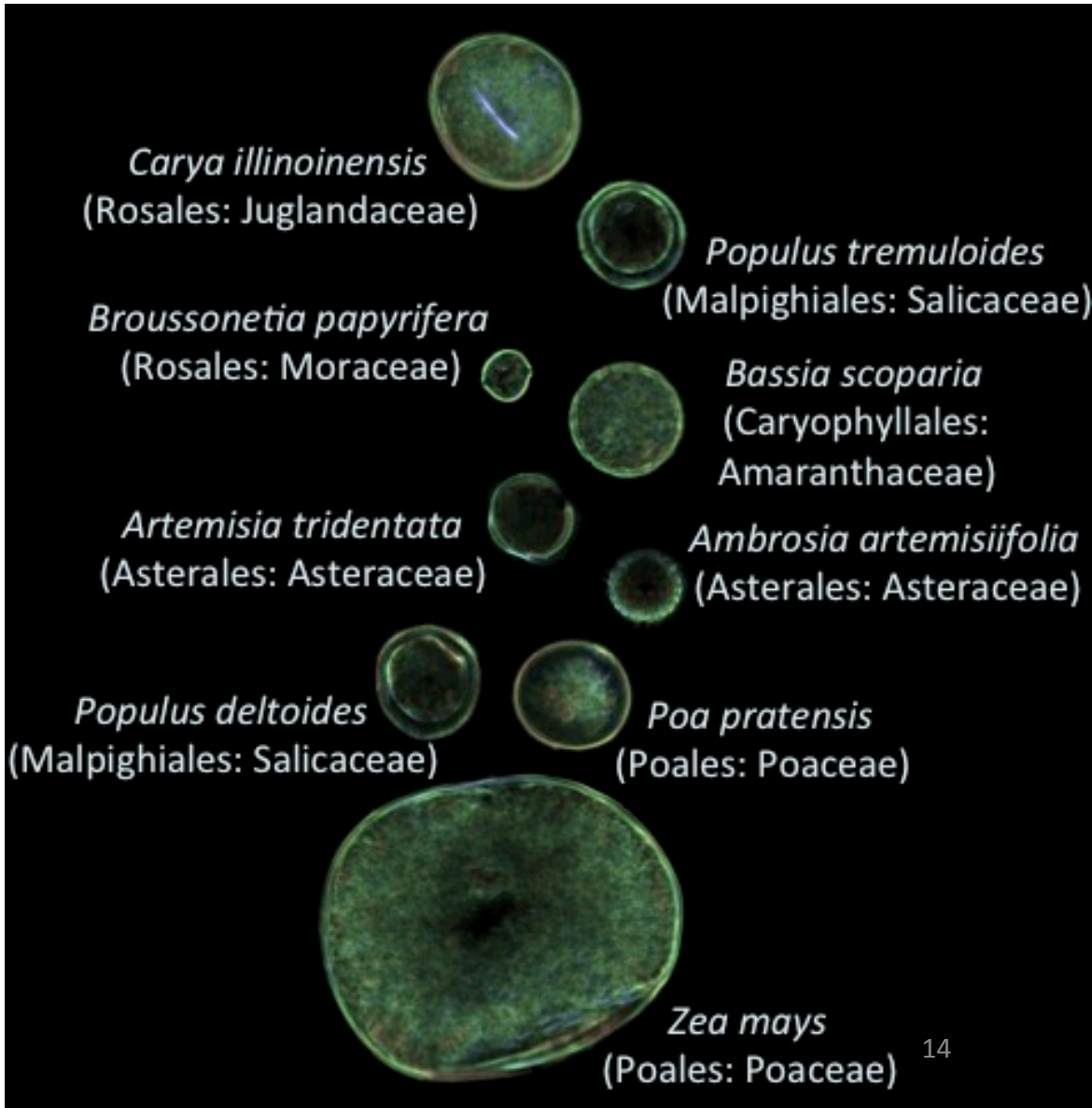
- Why test the method?
  - Don’t know how accurate the method is
  - Don’t know how many samples can be multiplexed
  - Don’t know if it is quantitative

Testing the basic pollen DNA metabarcoding method

- How many species can be detected within a sample?
- To what extent is this affected by taxonomic relatedness?
- How rare can a species be in a sample before it becomes undetectable?

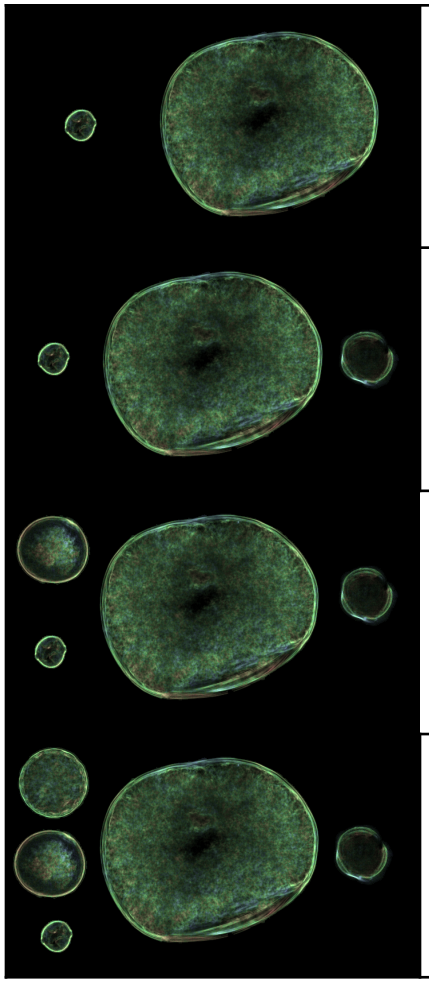
Testing the method with "knowns"

- Mixtures of varying complexity:
  - Number of species
  - Relatedness
  - Proportion of rarest species
- 450bp ITS2
- 500bp *rbcL*
- Illumina MiSeq:
  - Dual-index 96 samples/flow cell
  - Species identification with customized bioinformatics pipeline

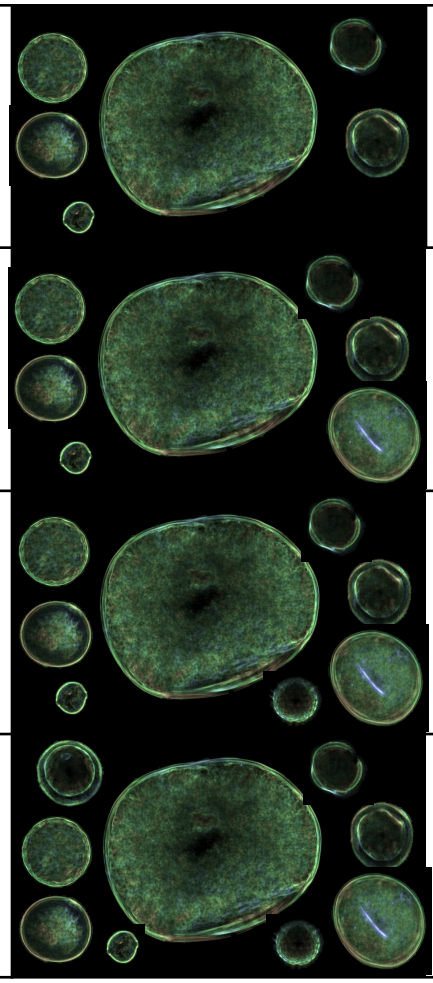


Bell, K.L., Burgess, K.S., Loeffler, V.M., Read, T.D., and Brosi, B.J. (in prep.). Testing the success of DNA metabarcoding of pollen for species identification and quantification with artificial mixtures of increasing complexity

ITS2 results: Increasing number of species

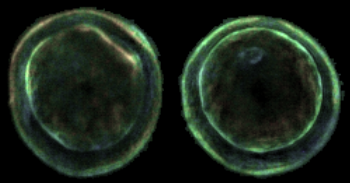


2 species detected	3 species identified to species, *Zea mays* undetected, 2 species identified to genus
Broussonetia papyrifera correctly identified, *Zea mays* undetected, *Artemisia tridentata* identified to genus	4 species identified to species, *Zea mays* undetected, 2 species identified to genus
2 species identified to species, *Zea mays* undetected, *Artemisia tridentata* identified to genus	5 species identified to species, *Zea mays* undetected, 2 species identified to genus
3 species identified to species, *Zea mays* undetected, *Artemisia tridentata* identified to genus	6 species identified to species, *Zea mays* undetected, 2 species identified to genus



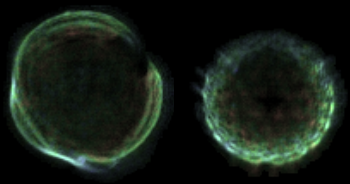
ITS2 results: Effect of taxonomic relatedness

Same genus



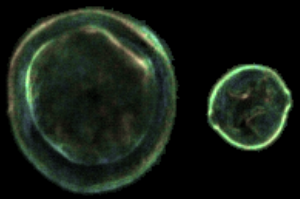
2 species detected

Same family



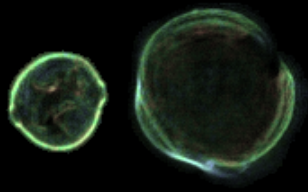
2 species detected

Same APGIII lineage

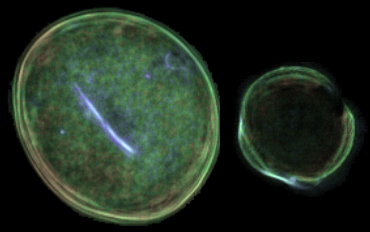


2 species detected

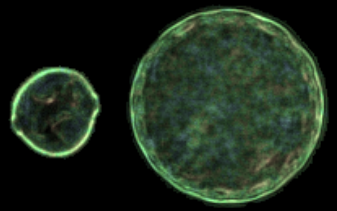
Different APGIII lineage



2 species detected

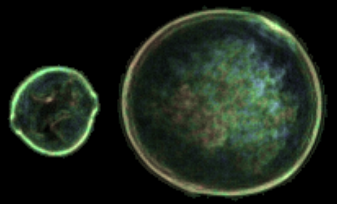


Zea mays undetected



2 species detected

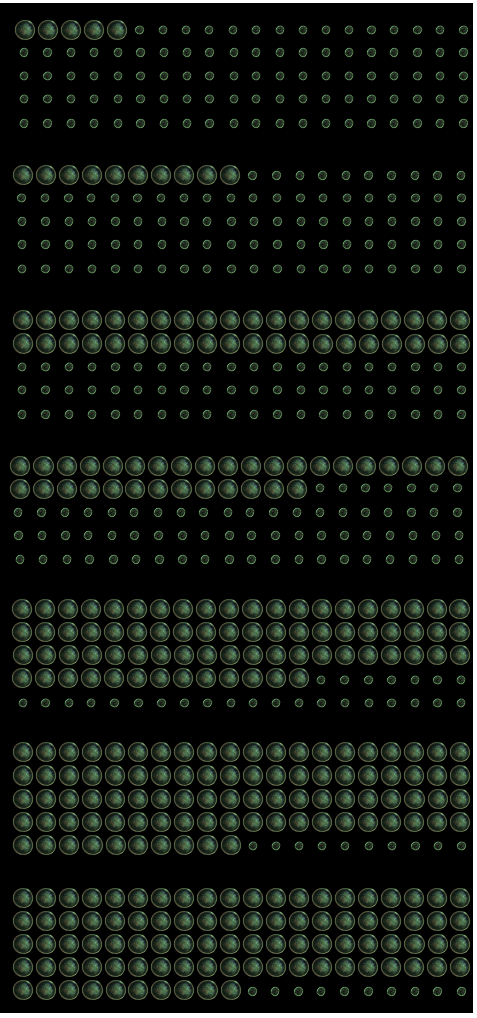
Monocot vs dicot



2 species detected

ITS2 results: Increasing rarity

Different families



2 species detected

2 species detected

2 species detected

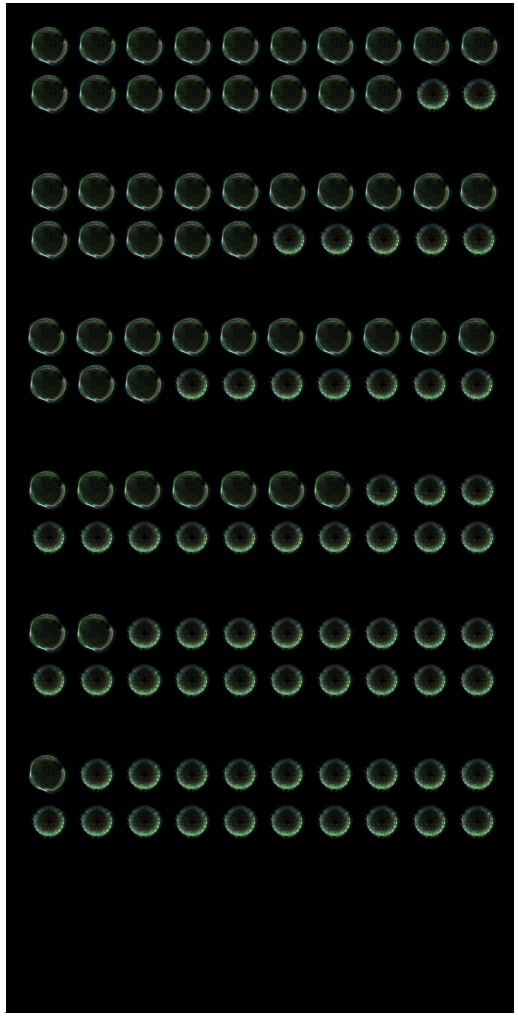
2 species detected

2 species detected

2 species detected

2 species detected

Same family



2 species detected

2 species detected

2 species detected

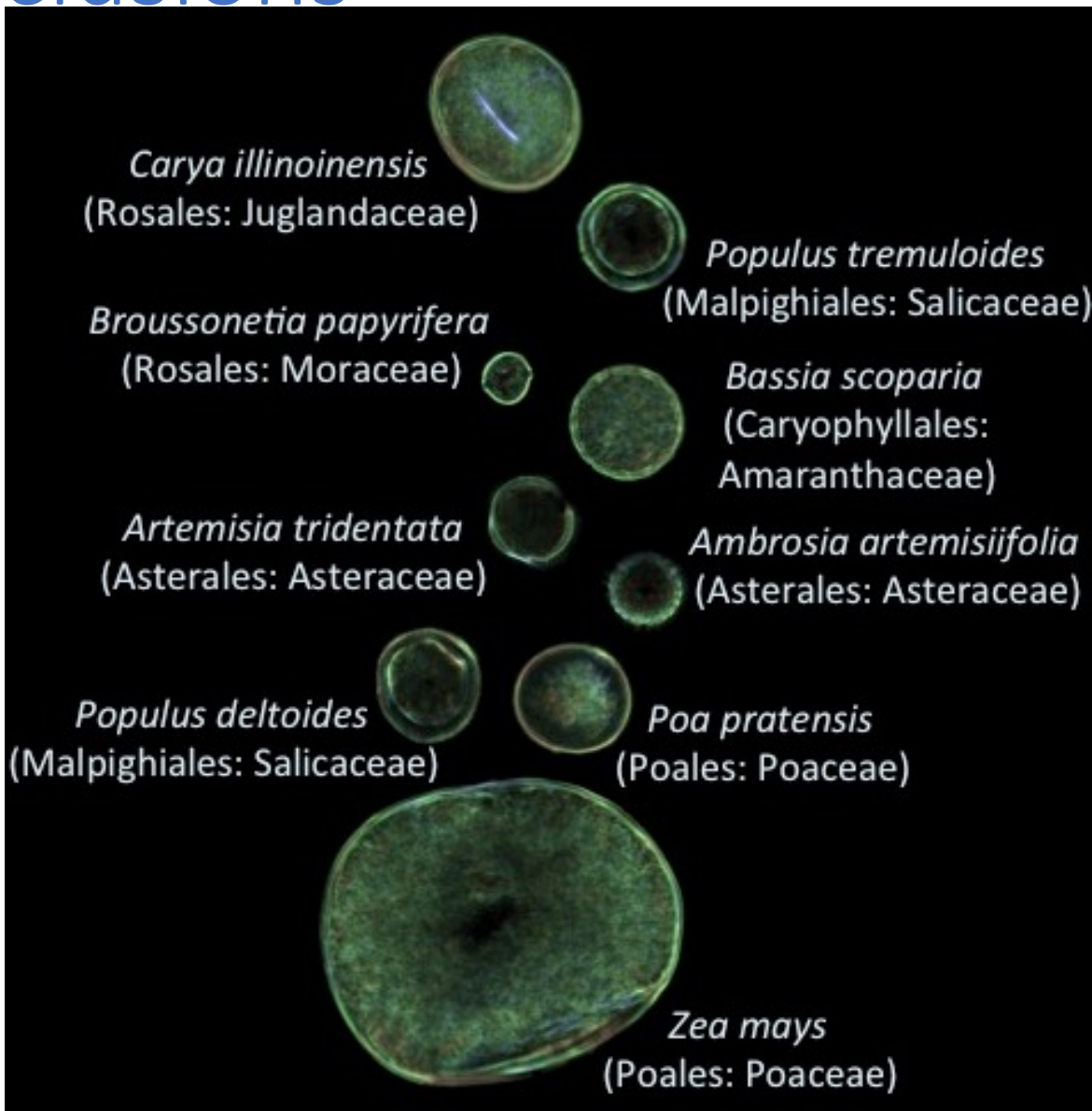
2 species detected

2 species detected

2 species detected

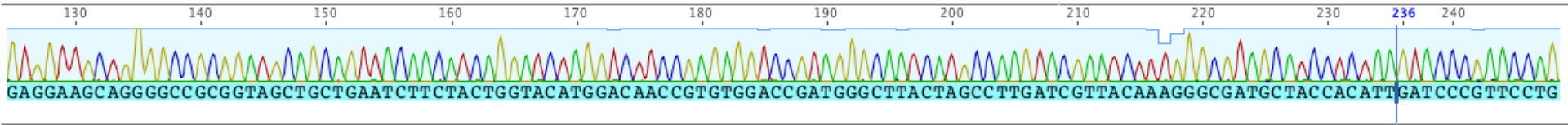
Testing the DNA metabarcoding method: conclusions

- Success varied across species
- Success dependent on rarity
- Not affected by:
  - Species richness
  - Taxonomic relatedness
- Not quantitative



DNA metabarcdoing: ongoing research

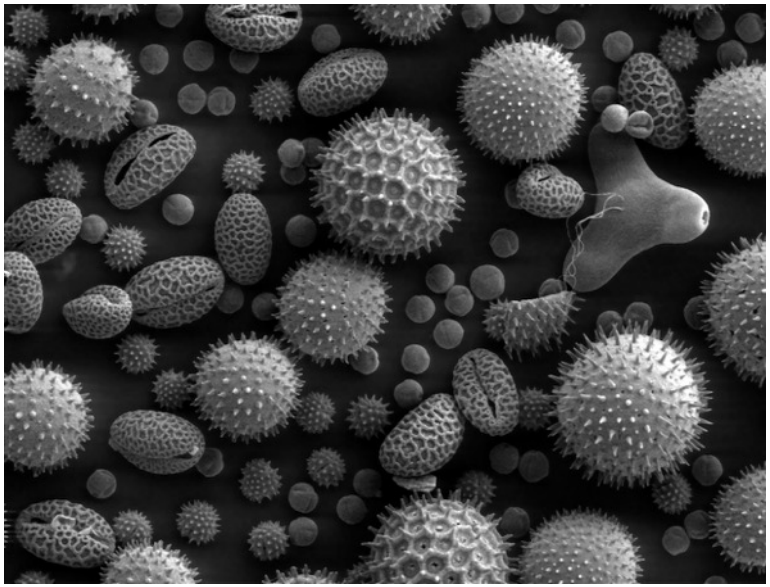
- Bioinformatics pipeline for mixed-amplicon DNA metabarcoding including *rbcL*
- Effect of sequencing depth
- Standard methodology



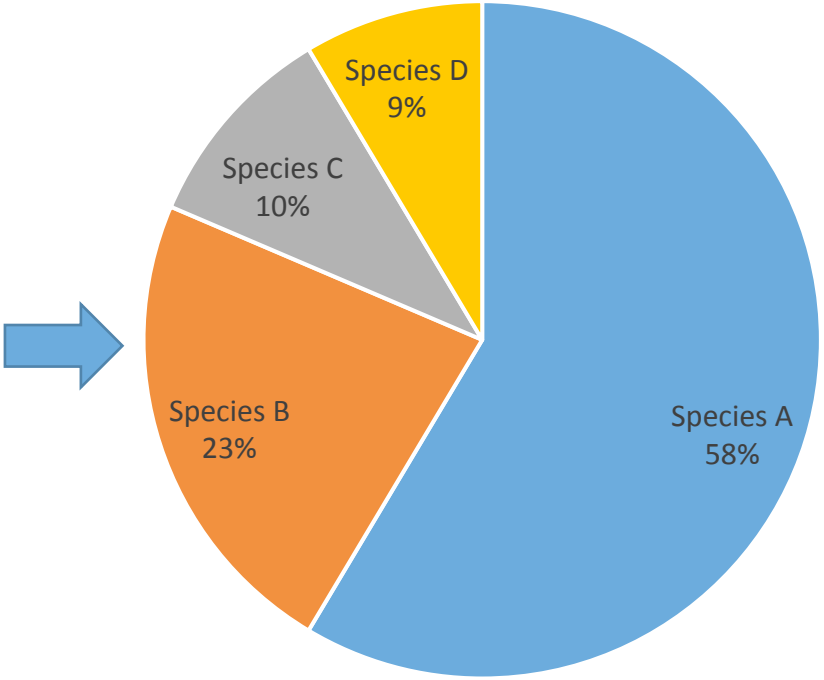
Overview

- Background and aims
- Testing the DNA metabarcoding method
- **Quantification**
- Whole genome shotgun approach
- Removal of non-pollen material
- Ongoing method development

Quantitative pollen DNA metabarcoding: goals



http://remf.dartmouth.edu/imagesindex.html



Quantitative pollen DNA metabarcoding: benefits

- Benefits for various applications
- Pollination biology:
  - Proportion of time on a particular plant species impacts pollination effectiveness
- Forensics:
  - Fine-scale geographic/temporal signature
  - May not differ qualitatively

Quantification & biases

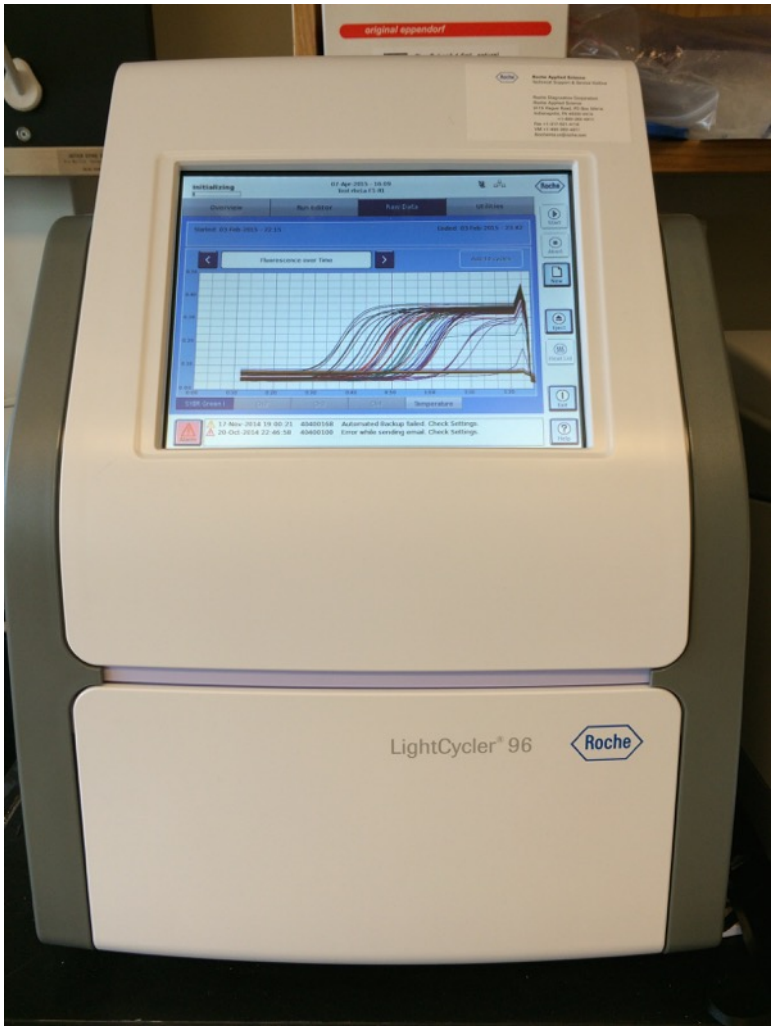
- Sources of bias:
  - Pollen preservation bias
  - DNA isolation bias
  - Amplification bias
  - Copy number bias
- Phylogenetic trends
- Correction

Quantification: testing for biases

- Pollen preservation bias:
  - Quantify DNA obtained from stored specimens
- DNA isolation bias:
  - Quantify DNA obtained from fresh specimens
- Amplification bias:
  - Context dependent
  - Testing on mixtures of known species composition
- Copy number bias:
  - *rbcL* and ITS2
  - qPCR methods

Quantification: copy number bias

- qPCR:
  - Follows PCR amplification in real time using fluorescent dye
  - Quantifies number of copies of gene of interest
- Determine copy number of chloroplast, ribosome and single-copy gene
- Look for phylogenetic patterns in copy number



Copy number bias: progress

- Primers developed for chloroplast for all major plant lineages
- Primers developed for ribosomal DNA for all major plant lineages
- Primers developed for single-copy nuclear gene for:
  - *Pinus taeda* – gymnosperm
  - *Magnolia grandiflora* – basal angiosperm
  - *Hibiscus syriacus* – eudicot

Copy number bias: qPCR analysis

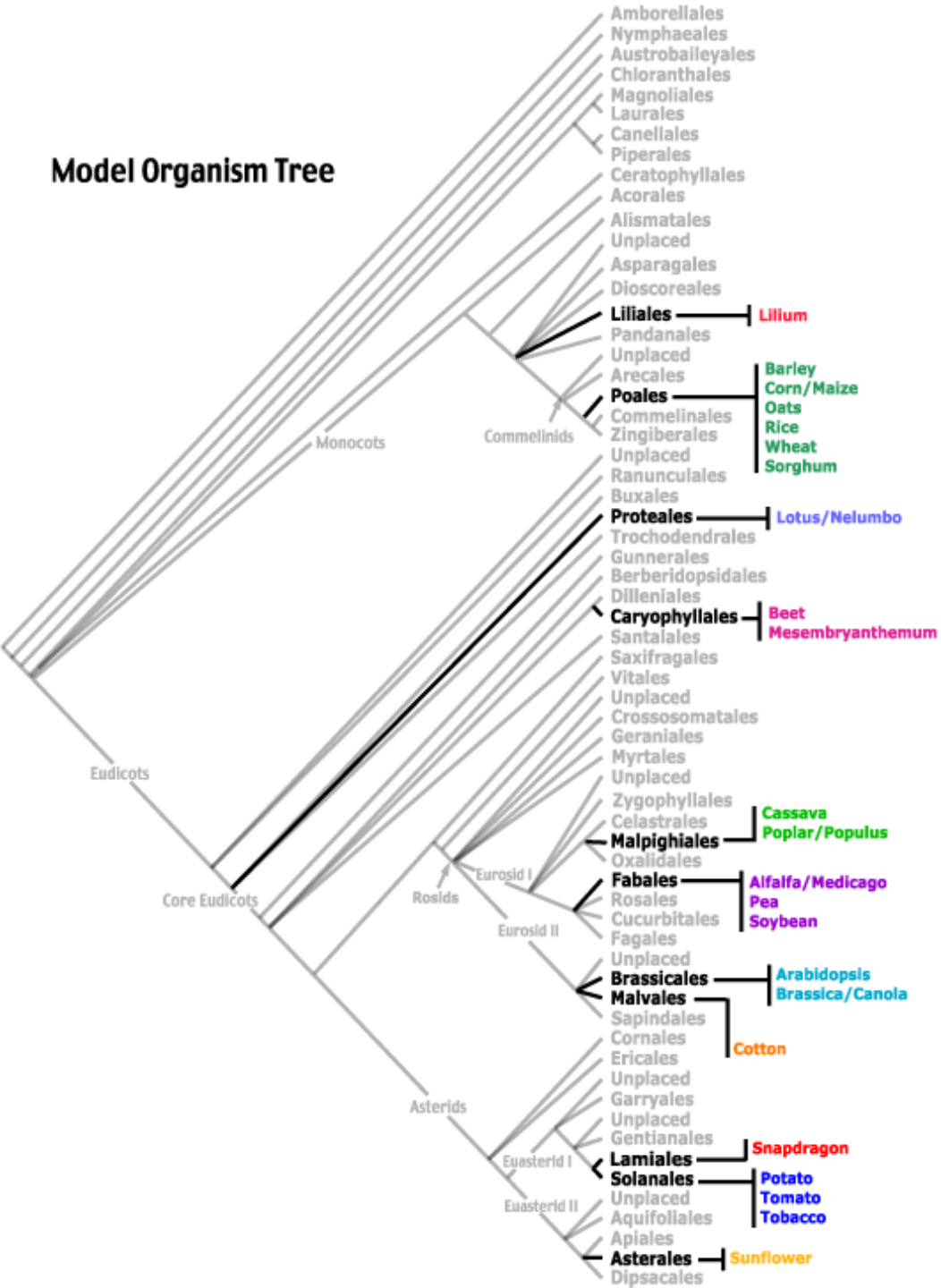
- Step 1: For 3 species with single-copy nuclear DNA primers:
  - Use qPCR to find copy number of chloroplast, ribosome, and single-copy nuclear gene
  - Calculate copy number per pollen grain for chloroplast & ribosome using:
    - qPCR copy number of 3 genes
    - Ploidy
    - Number of cells per pollen grain

Copy number bias: qPCR analysis

- Step 1: For 3 species with single-copy nuclear DNA primers:
  - Use qPCR to find copy number of chloroplast, ribosome, and single-copy nuclear gene
- Step 2: For all other species:
  - Measure DNA concentration
  - Use qPCR to find copy number of chloroplast & ribosome
  - Calculate copy number per pollen grain for chloroplast & ribosome using:
    - qPCR copy number of 3 genes
    - DNA concentration
    - Genome size
    - Number of cells per pollen grain

Copy number bias: qPCR analysis

- Step 1: For 3 species with single-copy nuclear DNA primers:
  - Use qPCR to find copy number of chloroplast, ribosome, and single-copy nuclear gene
- Step 2: For all other species:
  - Measure DNA concentration
  - Use qPCR to find copy number of chloroplast & ribosome
- For 3 species with single-copy nuclear DNA primers compare copy number calculation from both methods for accuracy



Copy number bias: sampling & phylogenetic analysis

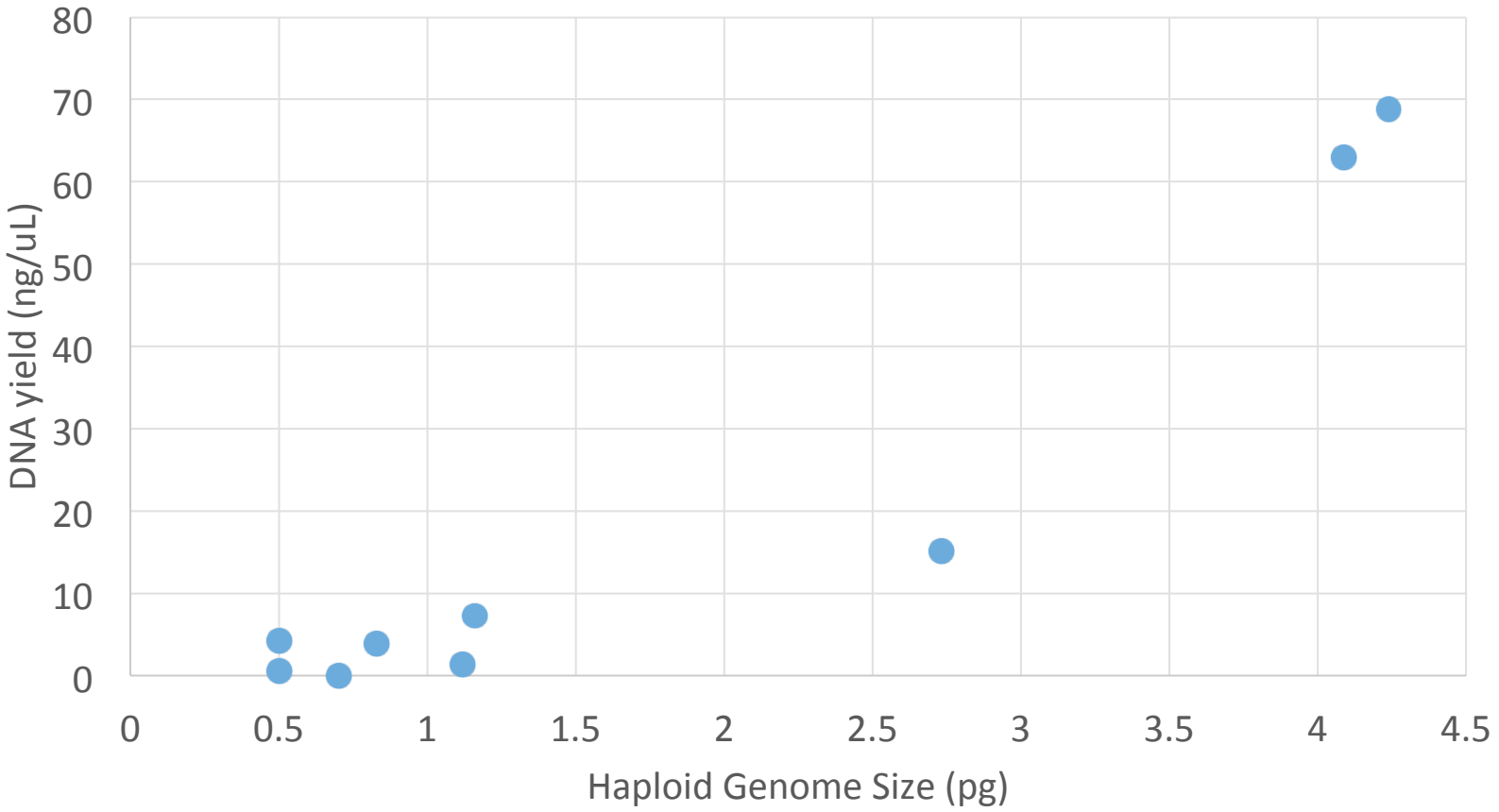
100 species for analysis:

- 25 orders, 3 species each in different families
- 5 orders, 2 species-pairs in same genus, one species from another family

Test for phylogenetic patterns in copy number

Make generalizations for lineage, order, or family

Quantification: DNA isolation bias



Quantification: other biases

- Pollen preservation bias:
  - Not examined
  - Could be examined using stored samples
- Amplification bias:
  - Could be examined with results of known mixtures
  - Results will only apply under same PCR conditions
  - Varies with other species in mixture

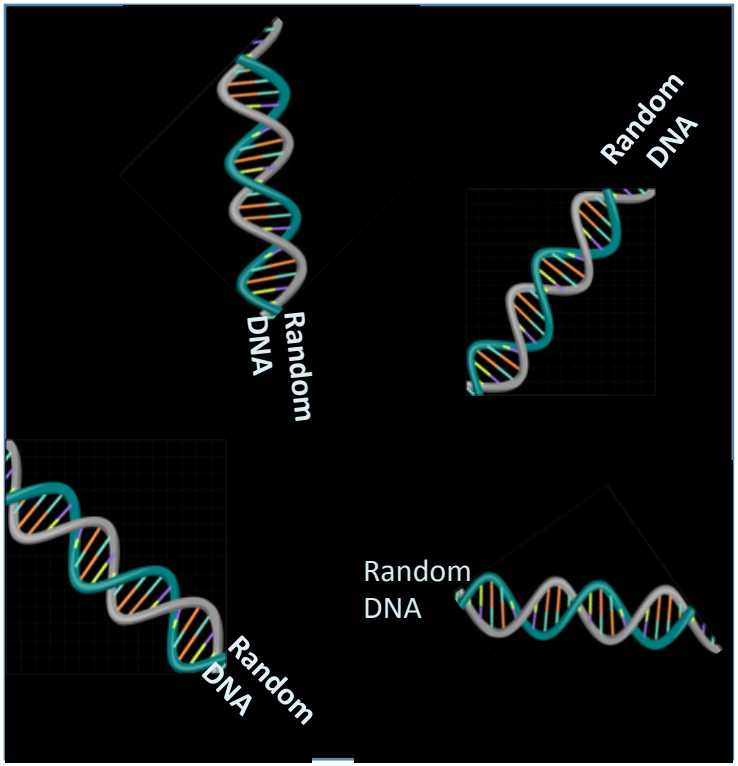
Quantification: ongoing research

- Test recently designed qPCR primers
- Quantify copy number of 100 species of pollen
- Determine phylogenetic trends for all biases
- Apply correction factors to convert from proportion of DNA sequencing reads to proportion of pollen grains
- Include correction factors in bioinformatics pipelines

Overview

- Background and aims
- Testing the DNA metabarcoding method
- Quantification
- **Whole genome shotgun approach**
- Removal of non-pollen material
- Ongoing method development

Whole Genome Shotgun (WGS) pollen DNA metabarcoding



WGS pollen DNA metabarcoding

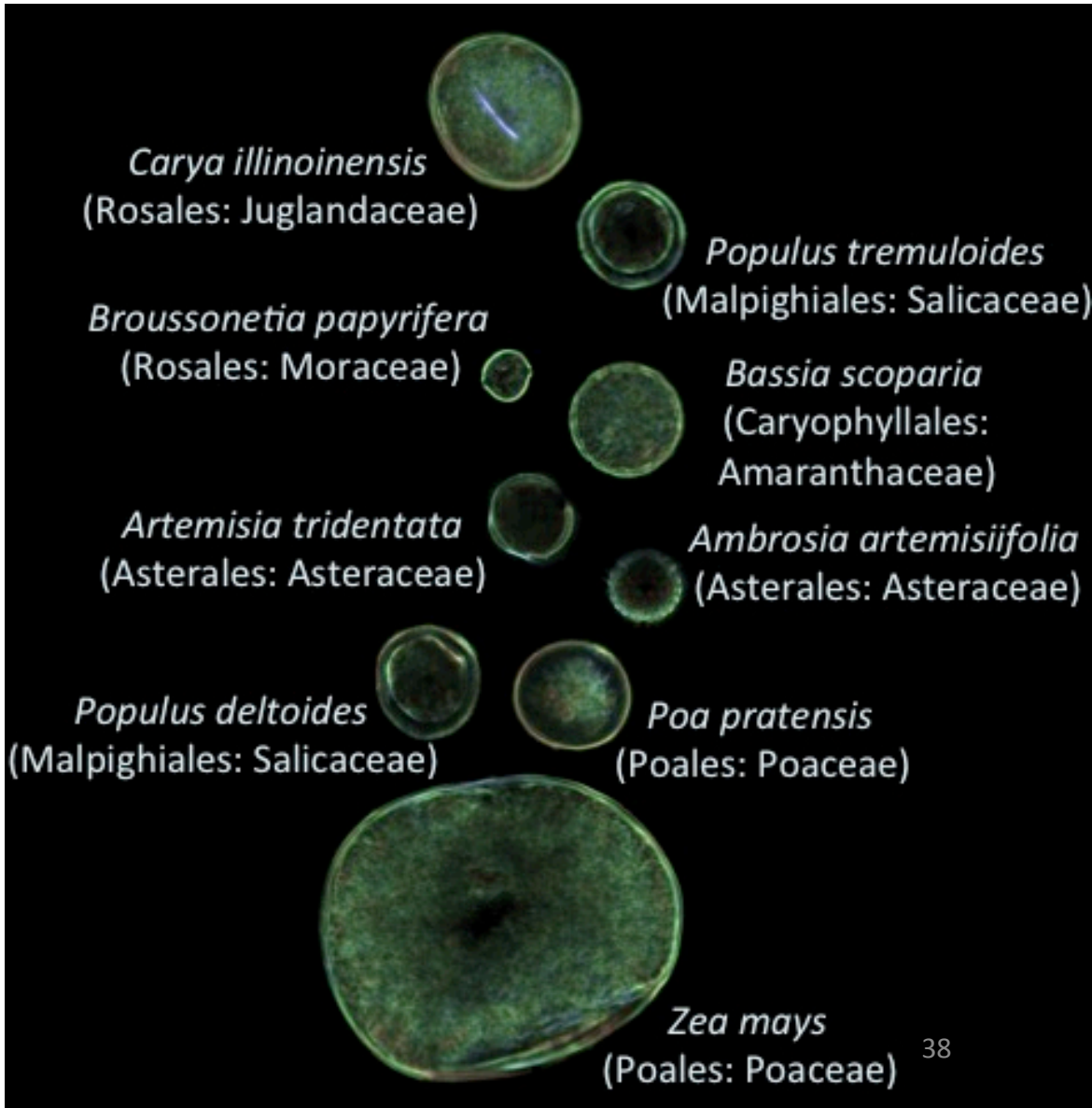
- Advantages over PCR-based approaches:
  - No amplification bias
  - No copy number bias
  - Finds variation between species identical in “barcode” genes
- Current challenges:
  - Genome size bias
  - Few species with genome data for reference library

Testing the WGS method

- How many species can be detected within a sample?
- To what extent is this affected by taxonomic relatedness?
- How rare can a species be in a sample before it becomes undetectable?
- How does it compare to mixed-amplicon sequencing?

Testing the method with "knowns"

- Subset of known mixtures used for mixed-amplicon sequencing
- Simulated data for same mixtures
- Illumina MiSeq:
  - 21 indexed samples/ flow cell
  - 24-30M paired reads of 150bp
  - Species identification with customized bioinformatics pipeline



Bell, K.L., Macpherson, M., Burgess, K.S., Read, T.D., and Brosi, B. (in prep.) Assessing the potential for whole genome shotgun methods in mixed-species pollen identification using artificial mixtures

WGS results: Simulated data

- Same mixtures as laboratory samples, with exception of those containing *Carya illinoinensis*
- Mixture complexities:
  - Up to 8 species
  - Pairs of varying relatedness
  - Down to 10% of pollen grains, 1% of DNA content

WGS results: Known pollen mixtures

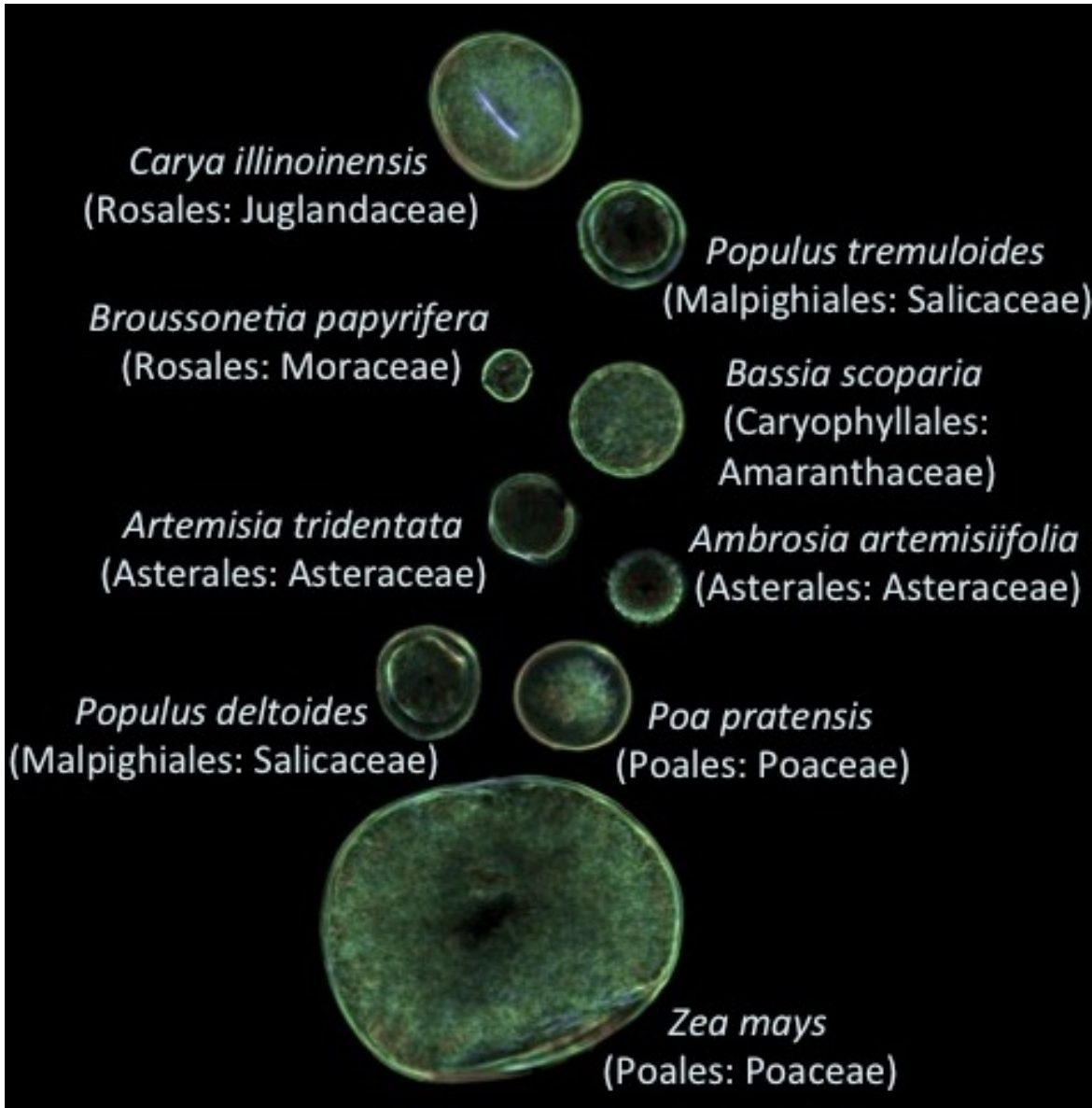
- Results better than simulated data
- Fewer false positives
- All species identified except *Carya illinoinensis*

WGS results: Overall success

- All species could be identified given presence in reference database
- Advantages over mixed-amplicon sequencing:
  - Species level identifications of species with identical ITS2 sequences
    - *Artemisia tridentata*
    - *Populus tremuloides*
  - Detection of species that amplified poorly with ITS2
    - *Zea mays*
- Disadvantages over mixed-amplicon sequencing:
  - Species absent from database:
    - *Carya illinoinensis*

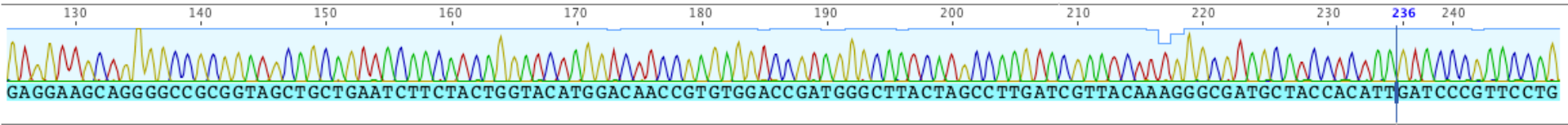
Testing the WGS DNA metabarcoding method: conclusions

- Demonstrated proof of concept
- Success varied only with database presence
- Not affected by:
  - Species richness
  - Taxonomic relatedness
  - Rarity
- Probably not quantitative



WGS: ongoing research

- Is it quantitative?
- Quantitative comparison to standard DNA metabarcoding
- Effect of sequencing depth
- Standard methodology



Overview

- Background and aims
- Testing the DNA metabarcoding method
- Quantification
- Whole genome shotgun approach
- Removal of non-pollen material
- **Ongoing method development**

Ongoing method development

- Improving bioinformatics pipeline for mixed-amplicon DNA metabarcoding:
  - Including multiple gene regions
  - Correcting for biases
- Improving methods to remove non-pollen plant material
- Developing standard methods for DNA metabarcoding and Whole Genome Shotgun methods

Possible future directions

- Developing a geographical database to assist with determining geographic origin of pollen samples
- Trialing standard methods on known mixtures that resemble forensic samples
- Developing standard methods for specific types of forensic and other analyses

Acknowledgements

Funding and Institutional Support:

- US Army Research Office (grants W911NF-13-1-0247 & W911NF-13-1-0100 to Brosi, Read, & Burgess)
- Emory University

Individuals:

- Berry Brosi (Emory)
- Kevin Burgess (Columbus State)
- Tim Read (Emory)
- Cindy Chu (Emory)
- Emily Dobbs (Emory)
- Julie Fowler (Emory)
- Virginia Loeffler (Emory)

Donated plant material:

- Atlanta Botanical Garden
- Missouri Botanical Garden
- New York Botanical Garden
- Montgomery Botanical Center
- Royal Botanic Gardens Victoria

Karen Leanne Bell
Emory University
karen.bell@emory.edu