

Claire Gay¹, Benoit Foucan Perrafide², Précillia Cochard¹, Robin Azemar¹, Fabienne Moreau³, Julien Thouin³, Gerardo Cebrián-Torrejón², Benjamin Poirot¹

¹APINOV Research and Training Center, 10 rue Henri Bessemer, 17140 Lagord, France

²Covachim M2E Laboratory, EA 3592, Université des Antilles, BP 250, 97157 Pointe à Pitre Cedex, France

³ADNid – Qualtech Groupe, 830 avenue du Campus Agropolis, 34980 Montferrier-sur-Lez, France

Honey's geographic origin significantly impacts its economic value and consumer trust. However, honey adulteration, including geographic mislabeling, remains a widespread issue. This study employs environmental DNA (eDNA) analysis combined with multivariate statistical methods to discriminate the geographic origins of honeys from Guadeloupe, Saint-Lucia, and the United States of America. DNA traces of plant species in honey samples were analyzed using trnL primer and the Illumina MiSeq platform, and processed with the FROGS pipeline. Three statistical approaches were applied: Principal Component Analysis followed by Linear Discriminant Analysis (PCA-LDA), Discriminant Correspondence Analysis (DiCA), and Nonmetric Multi-Dimensional Scaling (NMDS). Our findings demonstrate that the presence-absence data of plant DNA reads achieves higher classification accuracy than abundance data. PCA-LDA models yielded an average accuracy of 90.74%, with robust regional discrimination. DiCA provided complete separation of samples by region when using abundance data. NMDS revealed distinct floral compositions across regions when using presence-absence data, particularly between Caribbean islands and the USA. This study underscores the effectiveness of eDNA in honey for detecting subtle floral and geographic differences. The results advocate for eDNA-based techniques as a complementary or alternative method to conventional approaches like mellissopalynology or spectrometry. Furthermore, the application of multivariate analyses amplifies the discriminatory power of eDNA data, offering a robust tool for honey authentication. These findings hold significant implications for combating honey fraud, supporting local producers, and enhancing consumer confidence in labeled products.



Determining the Geographic Origin of American Honeys: An Analytical Method through Multivariate Analyses

Gay, c “Geographic Origin of Caribbean Honeys: An Analysis of Plant Composition through DNA Metabarcoding” Article in prep

Dr Benjamin POIROT
APINOV Research Center, France



Why honey origin matters ?

- Origin determines price & consumer trust
- Fraud : mislabeling geographical origin
- New method: **floral fingerprint** with **environmental DNA (eDNA)**





Honey origin identification :

- Melissopalynology: microscopic pollen identification
- Chemical/spectral fingerprinting (proteomics, IR, Isotope ratio, MNR)
- eDNA: high taxonomic resolution, fast, effective on very small and degraded samples



STUDY OBJECTIVES

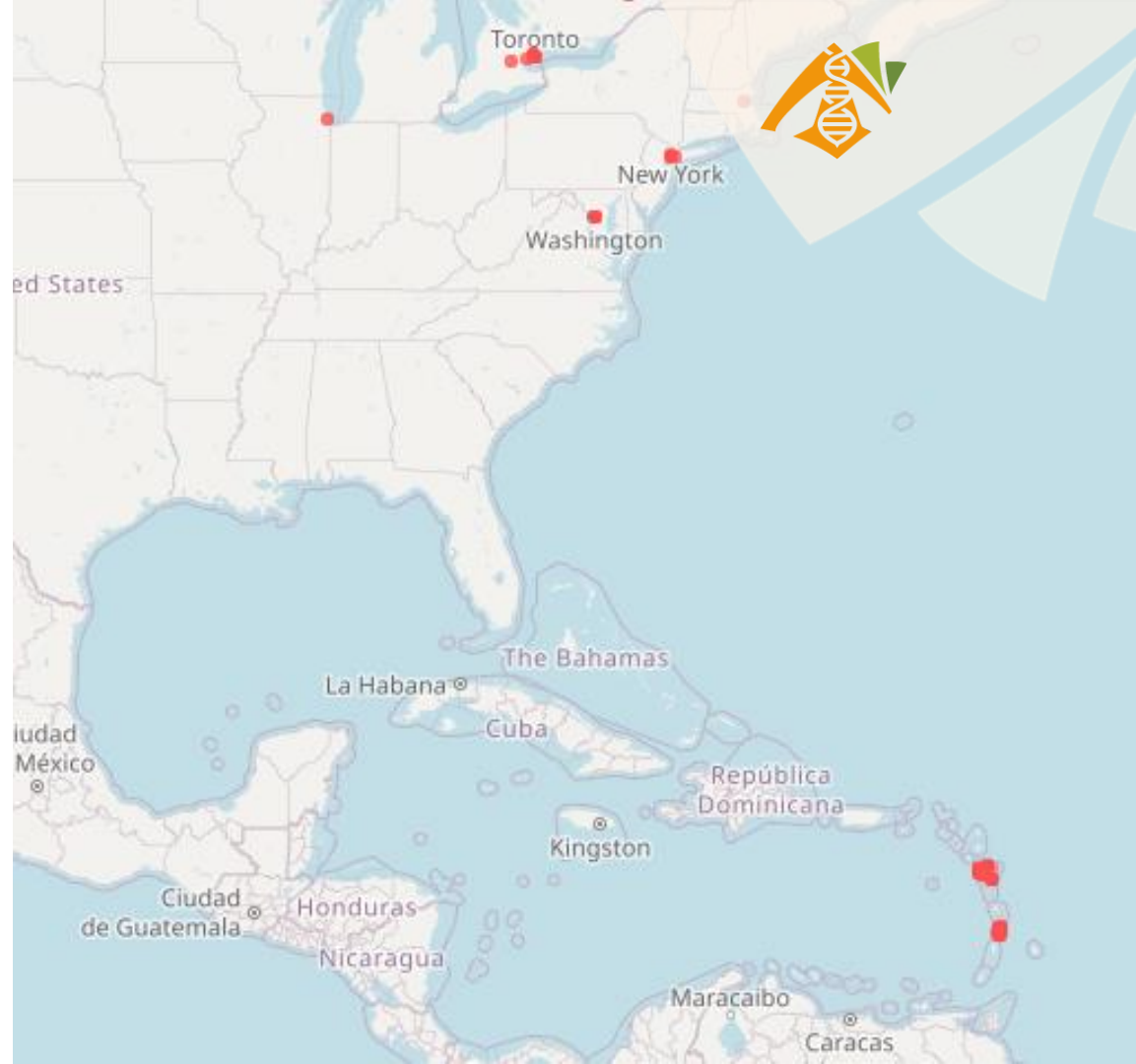
Authentication of the origin of honey from two neighboring Caribbean islands (Saint Lucia / Guadeloupe) using environmental DNA

- Compare plant **abundance** vs **presence–absence** data
- Apply **multivariate statistics models** to separate origins



Study Regions

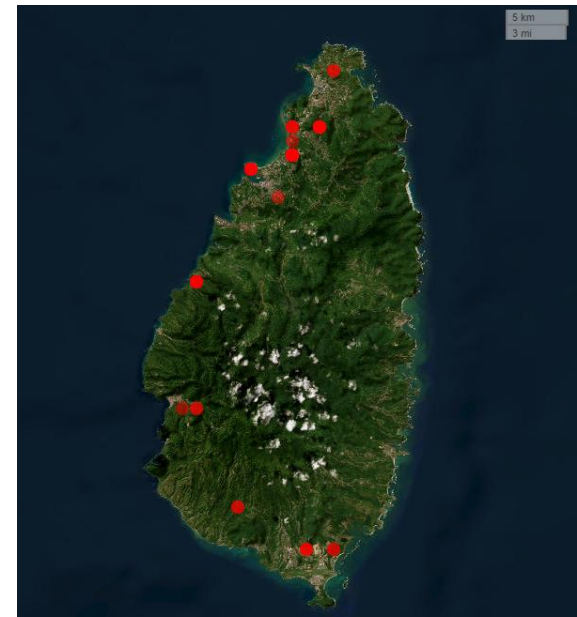
- Saint Lucia – 19 sites
- Guadeloupe – 19 sites
- North America – 19 sites



Saint Lucia / Guadeloupe islands

Sampling Map

- Multiple ecosystems: coastal forest, dry forest, mixed agriculture
- Different elevations and habitats to capture differences in floral availability.
- March–October sampling





North America Sampling Map

- Sites in USA & Canada:
Eastern Temperate
Forests
- April–October sampling





Sampling Protocol

- Honey collected directly from hive frames to ensures floral DNA signature reflects foraging environment





Lab Workflow

DNA extraction

PCR with **trnL g-h** and **trnL c-h** primers (markers)

Sequencing (Illumina MiSeq)

Species identification via reference database (NCBI Nt database)

Filters for impossible species : $<0.1\%$ abundance + % identity (95%) + E -value cut-off of $1e^{-10}$ + query cover (95%) + presence plausibility via literature

Botanical Composition: Most abundant species

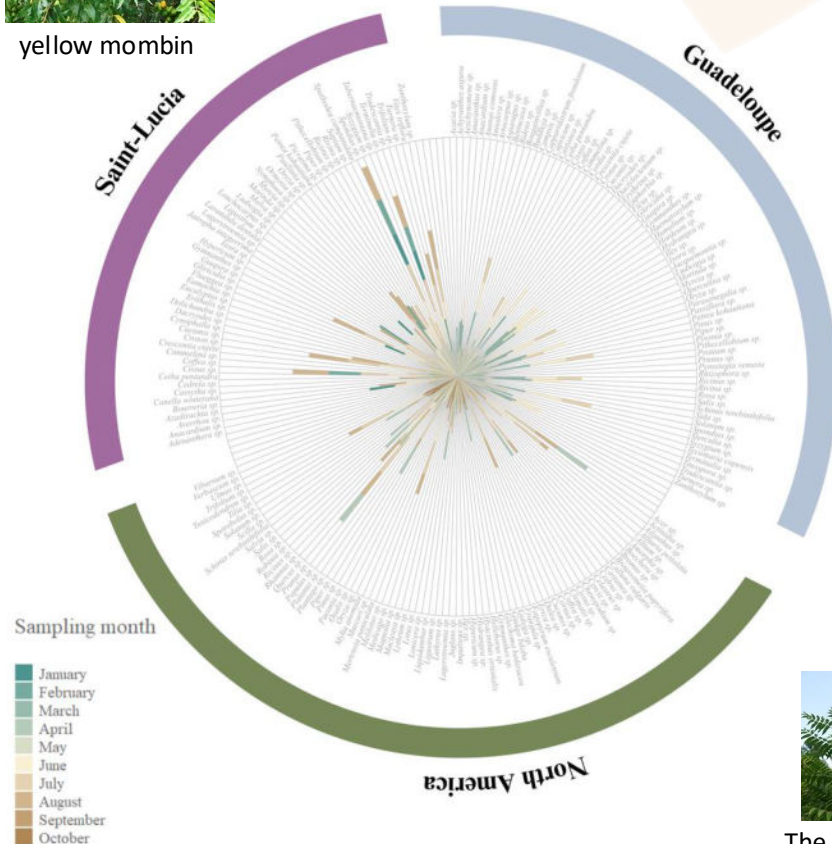
- **North America:** *Ailanthus altissima*, *Lythrum salicaria*
- **Guadeloupe:** *Bursera simaruba*, *Gymnanthes lucida*, *Homalium racemosum*
- **Saint Lucia:** *Spondias mombin*, *Guapira discolor*, *Croton hircinus*
- Some species common to all regions : melon (*Cucumis melo*)



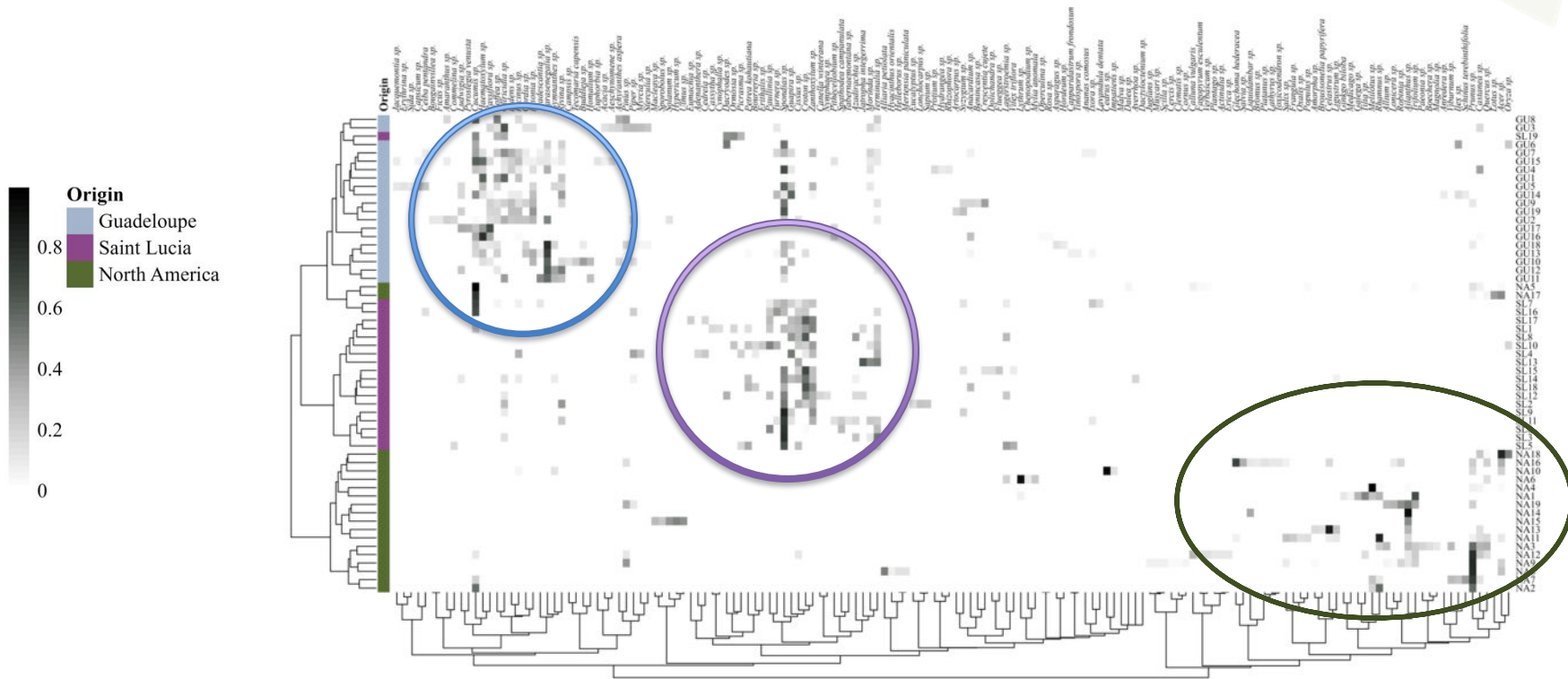
yellow mombin



gumbo-limbo



The tree of heaven

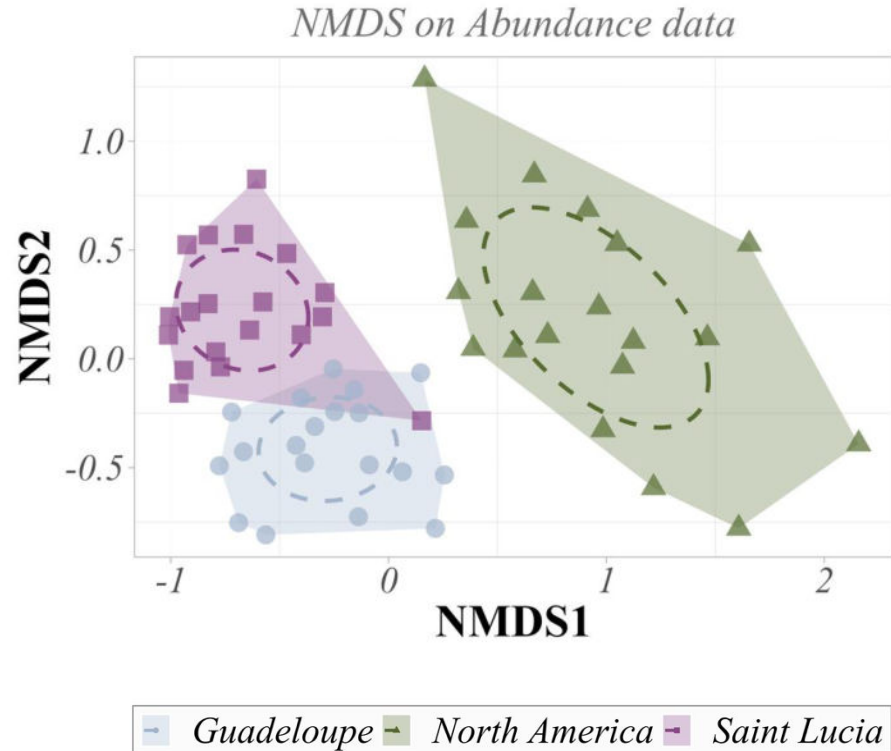




Multivariate statistics approach :

NMDS Analysis on abundance data

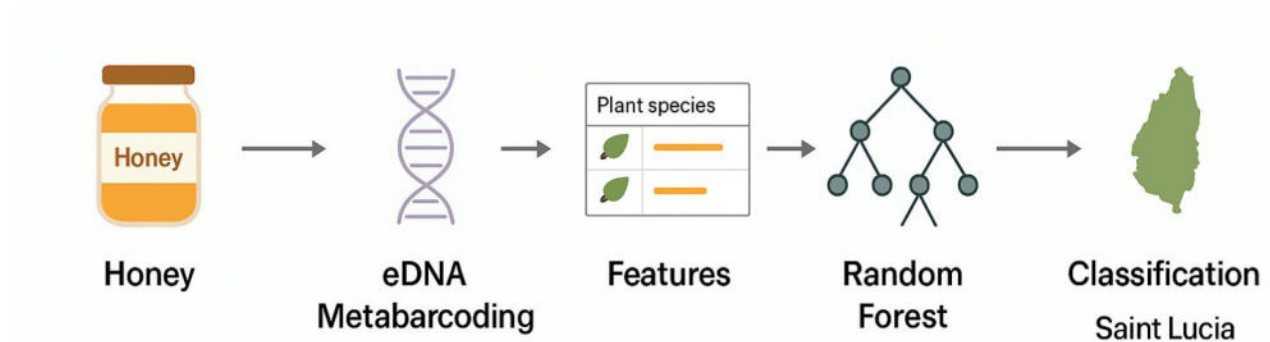
- NMDS model shows similarity/difference in plant compositions
- NA distinct, Guadeloupe and SL show a small overlaps
- Statistical tests (anosim) confirm significant differences





Next Step: Machine Learning for Honey Origin Authentication

- Train machine learning models on global eDNA data (>1000 samples).
- Tested algorithm: **Random Forest**
- Preliminary results : **72% correct classification** for continent and climate origine. Up to 83% for Caribbean samples





Implications



Science:

- improved regional discrimination
- new tool for fraud detection



Beekeepers:

- origin certification
- market value
- consumer trust



Conclusion

- Environmental DNA makes it possible to distinguish between honeys from Caribbean neighboring regions.
- Complementary to traditional methods and need to be integrated to international standards (Codex and ISO)
- A promising tool for assessing fraud and adulteration and certify labels

Thanks !



Funding: ODEADOM, GEF/UNDP, SLUNCF, Alvéole, Apinov



Alex McLean
Alveole



Richard Matthias
Iyanola Apiculture



Benoit Foucan-Pérafide
SCEA O'Miel



Claire Gay
Data scientist Apinov

Thanks to field teams and collaborators : Claire Gay*, Benoit Foucan-Pérafide, Précillia Cochard, Robin Azemar, Julien Thouin, Fabienne Moreau, Muriel Sylvestre, Marie-Noëlle Sylvestre, Richard Matthias, Gerardo Cebrián Torrejón