Implementation of a food origin verification system

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND
SAPERE AUDE

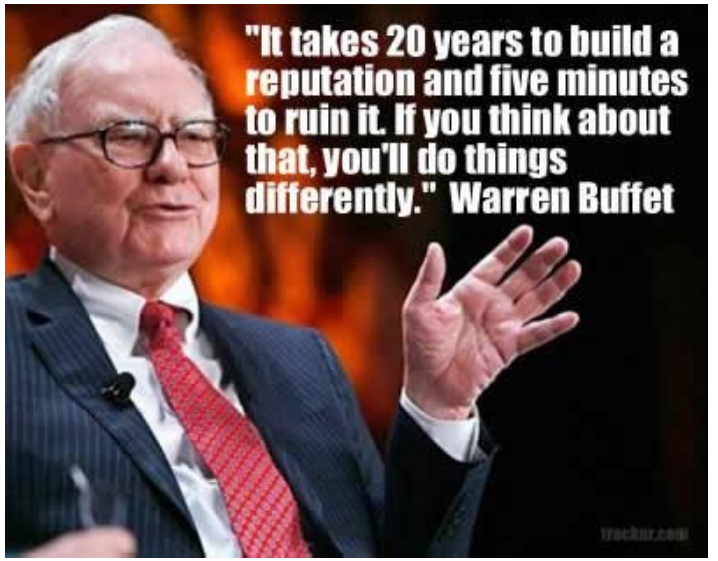
Why Implement Food Origin Systems?

- New Zealand is known for producing high quality food products
  - Manuka Honey
  - Dairy
  - Kiwifruit
  - Wine

- New Zealand is dependant on export of Agri and Hort products – maintaining a reputable brand and high quality is extremely important

Why Implement Food Origin Systems?

- New Zealand is known for producing high quality food products
  - Manuka Honey
  - Dairy
  - Kiwifruit
  - Wine



- New Zealand is dependant on export of Agri and Hort products – maintaining a reputable brand and high quality is extremely important

Before we start – the key message

IF YOU WANT TO BE SURE WHERE SOMETHING COMES FROM, YOU NEED TO TEST THE ACTUAL PRODUCT ITSELF.

Fingerprinting – Principles

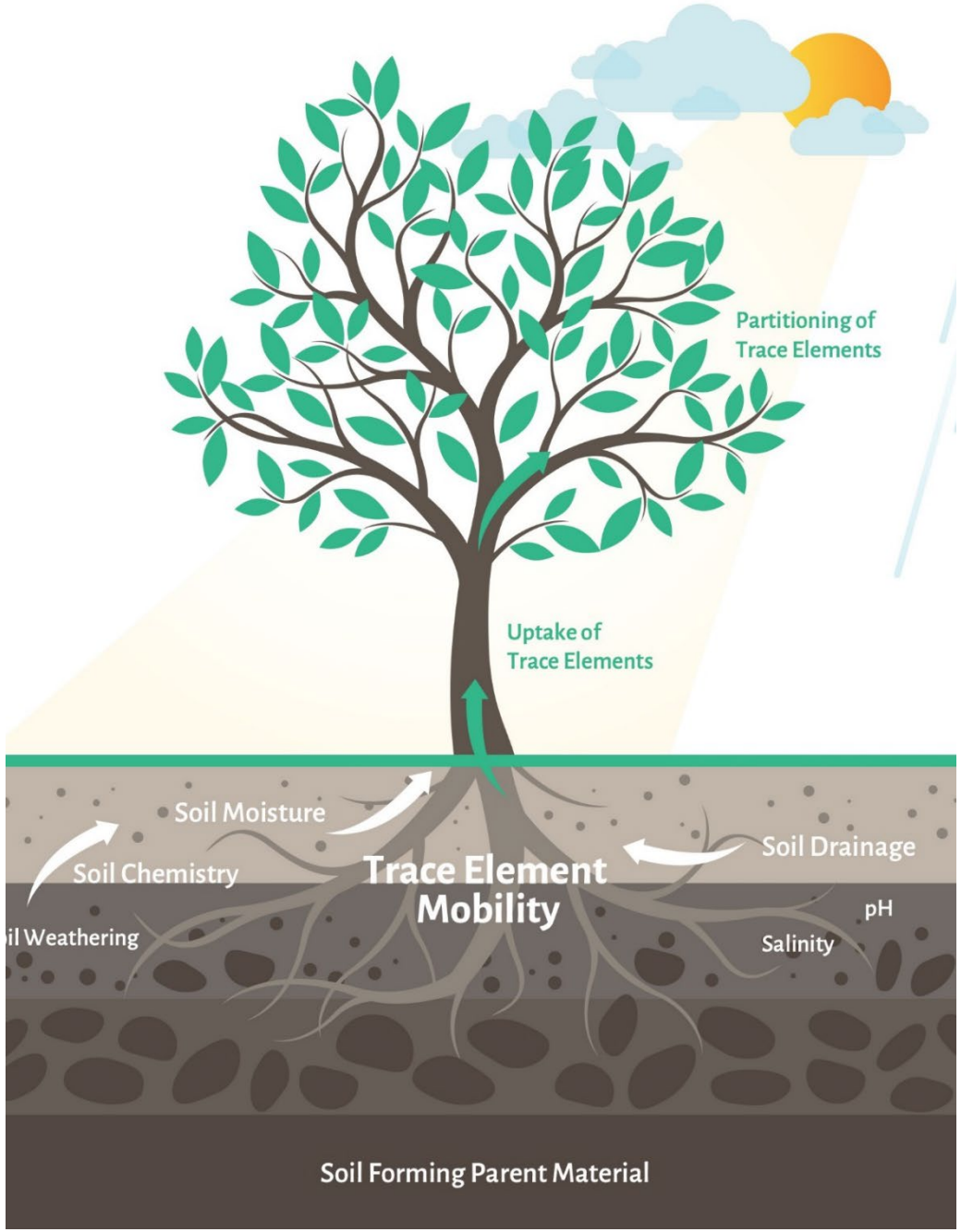
There are four basic principles required for a fingerprinting system to work.

1. Fingerprints must be individual (unique)

2. Fingerprints should remain the same over time

3. Fingerprints should be able to be systematically classified

4. Fingerprint markers should have direct causal relationship with the object to be identified



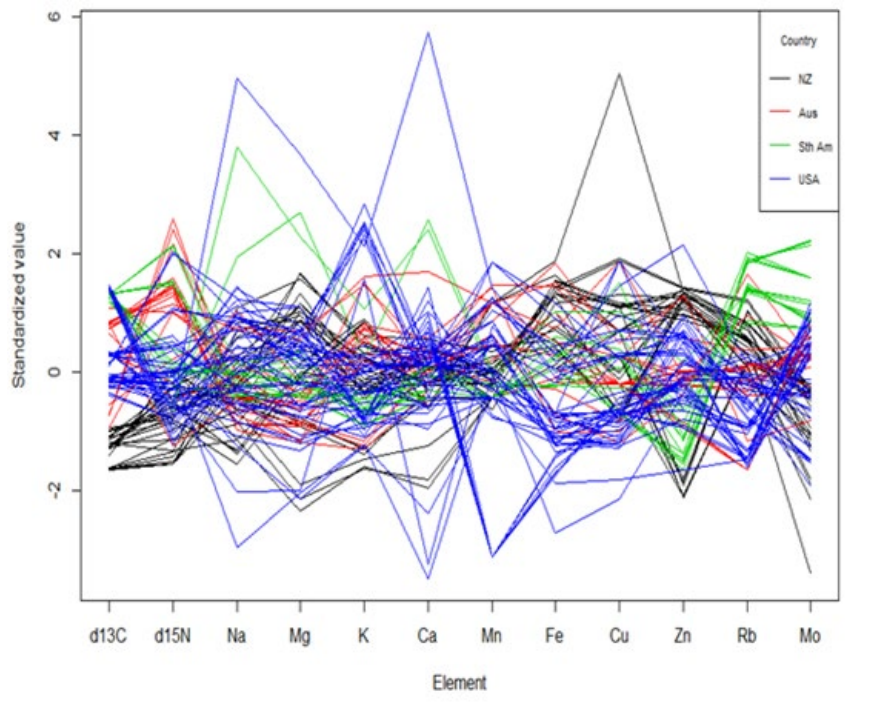
How Nature Creates a Unique Fingerprint

The product naturally accumulate elements and isotopes. This gives the product its unique origin fingerprint

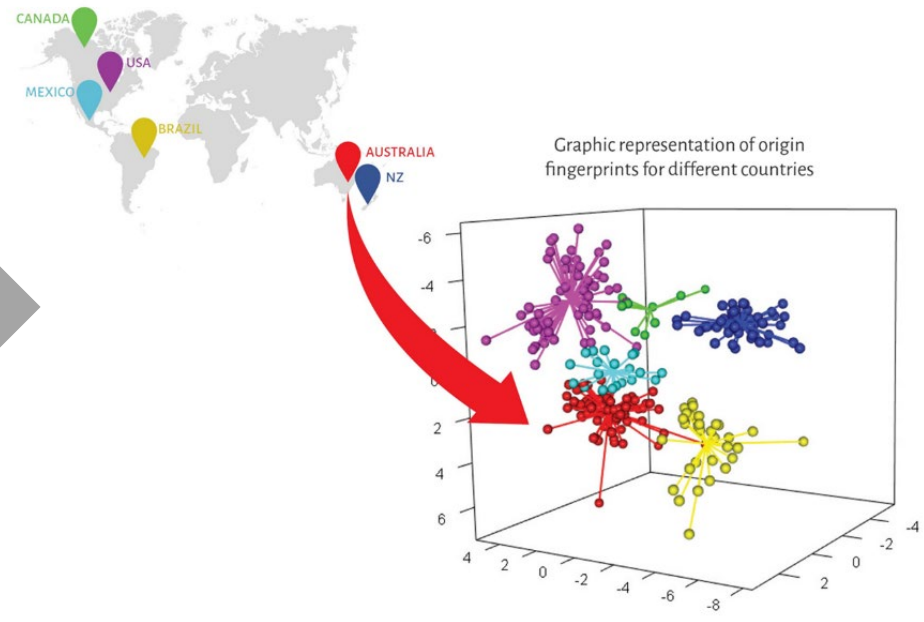
Origin Fingerprint

Oritain uses data science and statistics to interpret 1000's of sample/analyte data into something that can be comprehended.

We convert Raw Data from this



Into graphical representations such as this



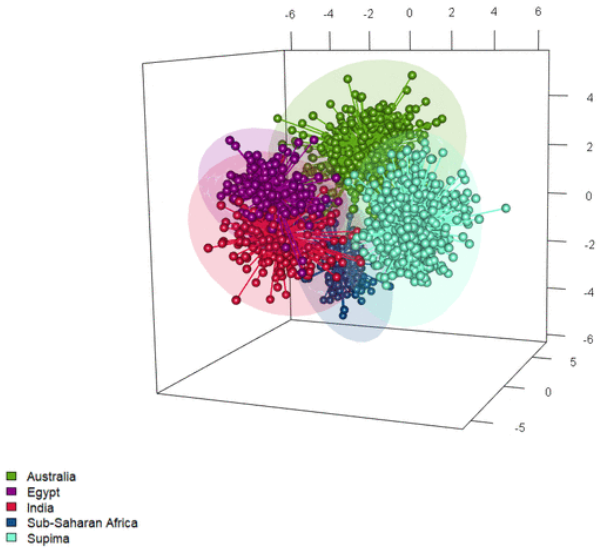
Resolution of the chemical fingerprint

The science can discriminate between properties with very close proximity if required

Resolution of the chemical fingerprint

The science can discriminate between properties with very close proximity if required

Resolution: Country

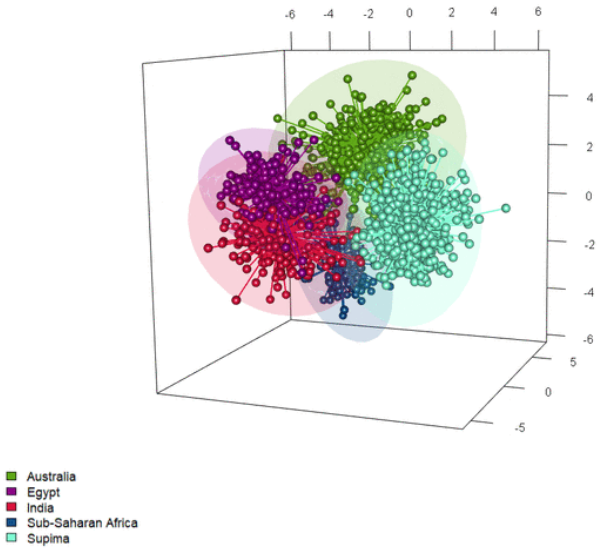


Slide

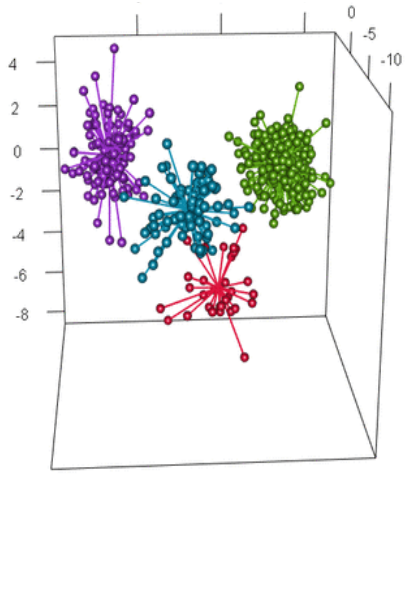
Resolution of the chemical fingerprint

The science can discriminate between properties with very close proximity if required

Resolution: Country



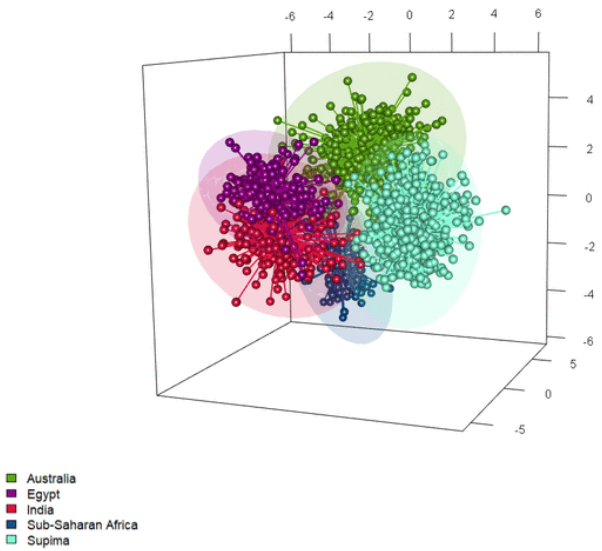
Resolution: Region



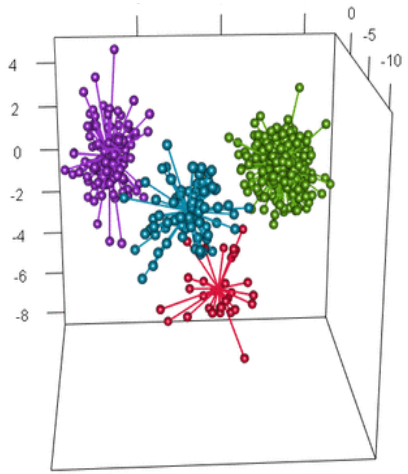
Resolution of the chemical fingerprint

The science can discriminate between properties with very close proximity if required

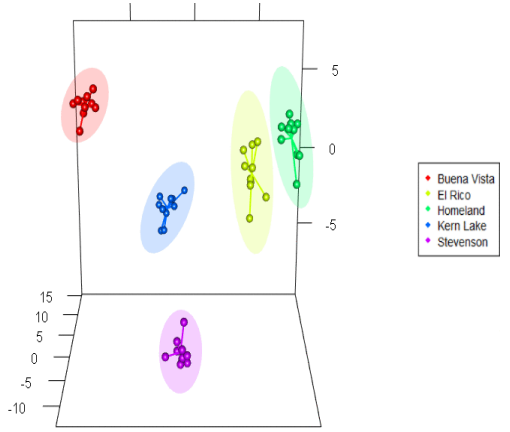
Resolution: Country

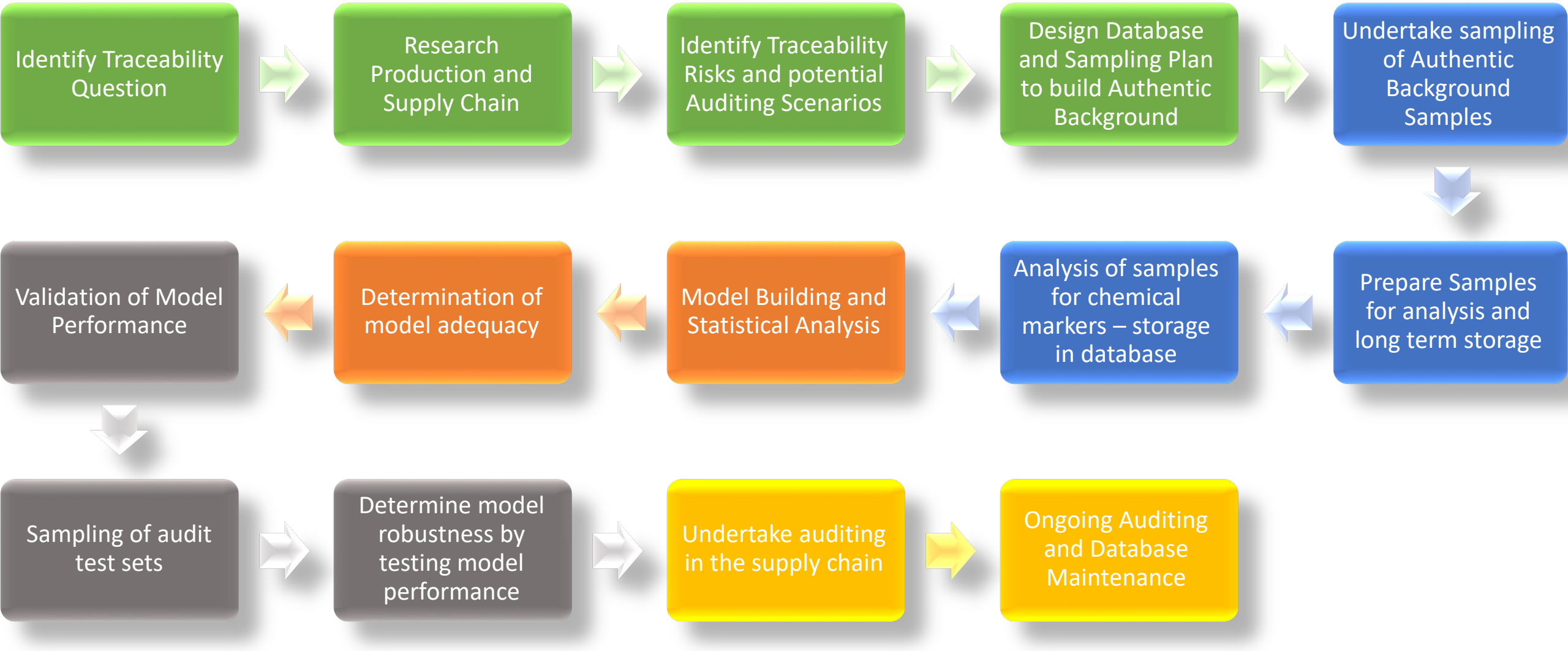


Resolution: Region



Resolution: Farm





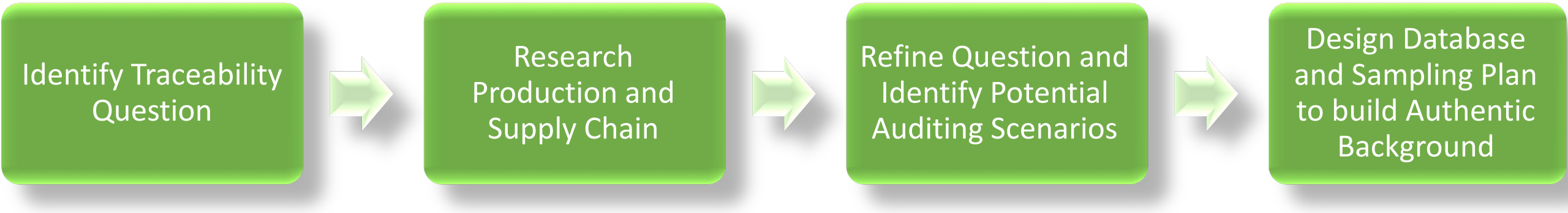
Design Phase

Chemical Analysis Phase

Statistical Modelling Phase

Fit for Purpose Phase

Active Auditing Phase



Design Phase | Chemical Analysis Phase | Statistical Modelling Phase | Fit for Purpose Phase | Active Auditing Phase

Identify Traceability Question

- This should always be the first milestone when setting up a verification system
- Many different traceability questions
    - Is the product from this location?
    - Is the product not from that location?
    - Where is this product from?
    - Has this product gone through the correct parts of the supply chain?
    - Has anything happened to this product while it was in the supply chain?
    - Is this product what it says it is on the label?

- **“Is the tea in the market consistent with the label of produced in Ceylon”**
    - **Region**
    - **Estate**
    - **Cultivar (flavour)**

- The statistical modelling used in chemical fingerprinting relies on the question asked being well defined and testable.

Research Production and Supply Chain

- Need to learn as much about the target products production as possible.
    - Processing steps can affect the chemical markers being measured – we want to understand how these could confound the chemical fingerprint.
    - Farming/Manufacturing processes are an influential part of the formation of the chemical fingerprint – the contributions to the variance of the chemical fingerprint are important to characterise.
    - For the best comparison it is optimal if very little processing or change occurs between background and audit samples.

- We want to understand enough to determine what our lowest sampling unit will be and where potential confounding variation could come into the chemical fingerprint.
    - **Cotton is packed into bales at the origin and is shipped for further manufacturing into thread, then greige fabric before bleaching, dyeing and being made into a product.**

Identify Risk, Auditing Scenarios

- Given knowledge of the production system and the supply chain, it is best to review the traceability question in light of potential risks.
- The need for origin verification is at the end of the supply chain at the final product. This is where audit samples are obtained for comparison.
- Is there a stage where product from other origins can be incorporated? It is this sort of fraud we want pick up by determining the origin.
- The manufacturing process becomes a source of confounding variation

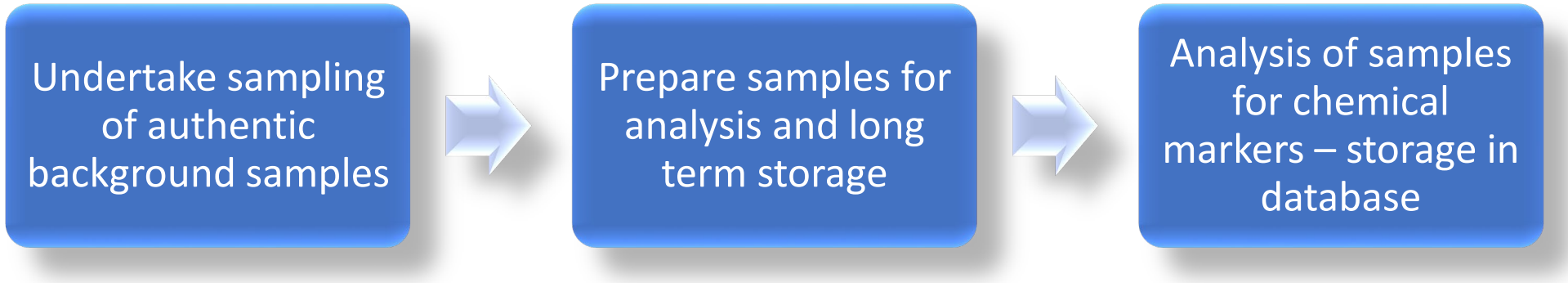
Design Database, Sampling Plan, build Authentic Background

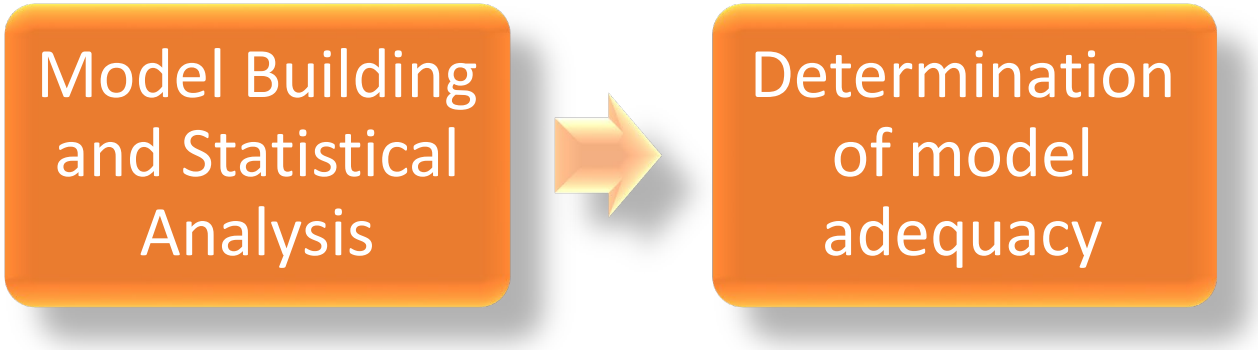
- What samples do we need, in what form and how do we ensure their integrity.

- What descriptive information (metadata) do we need to have with the samples to help us answer our question.
    - Location information – GPS coordinate, Country of Origin.

- How will we store and organise the information to keep track of it easily.
    - Information Management System.

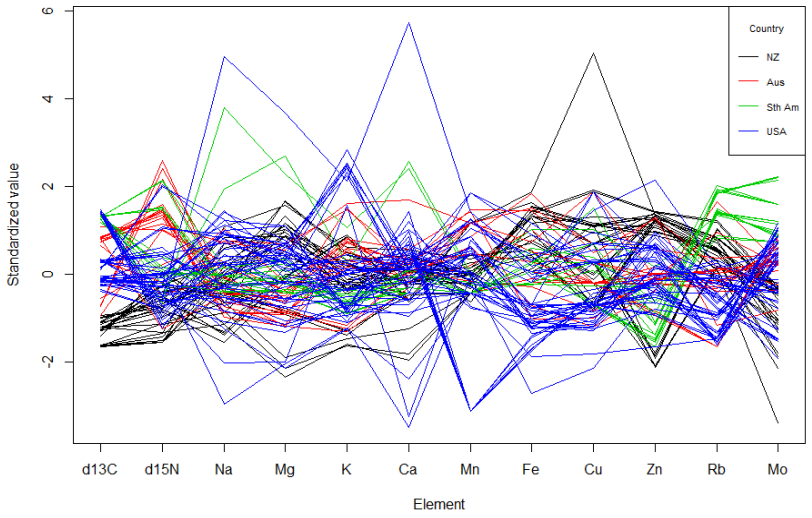
- **Good design and planning makes the logistics, analysis and modelling easier to undertake.**

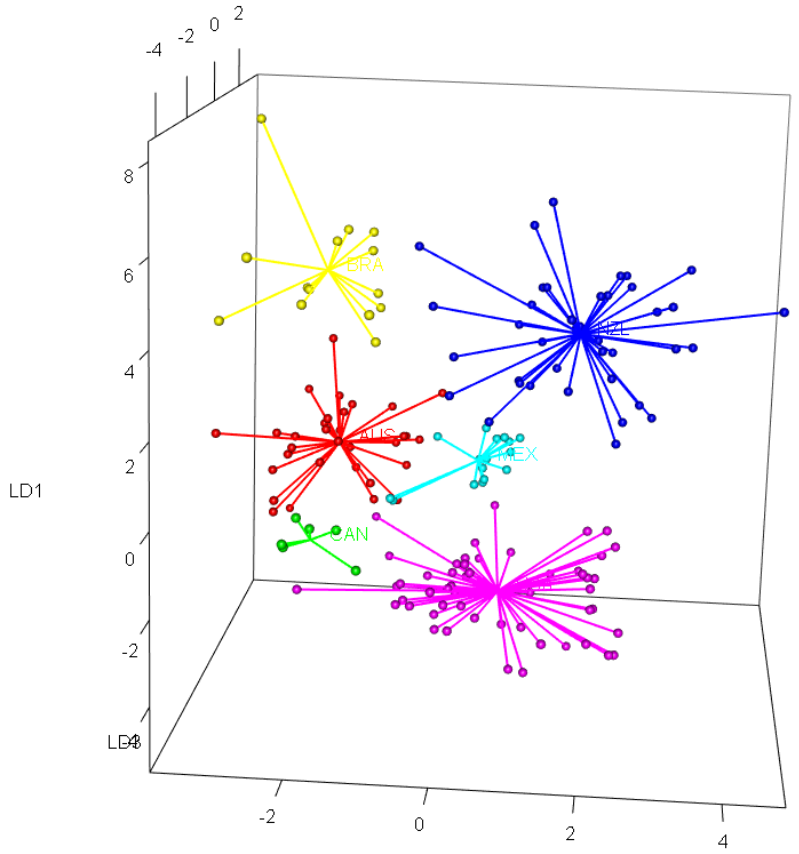

Undertake sampling of authentic background samples
Prepare samples for analysis and long term storage
Analysis of samples for chemical markers – storage in database


Model Building and Statistical Analysis
Determination of model adequacy

Model Building and Statistical Analysis

Statistics and Modelling – understanding the chemical fingerprint.





Determination of Adequacy

Follow a process when building a model for data:

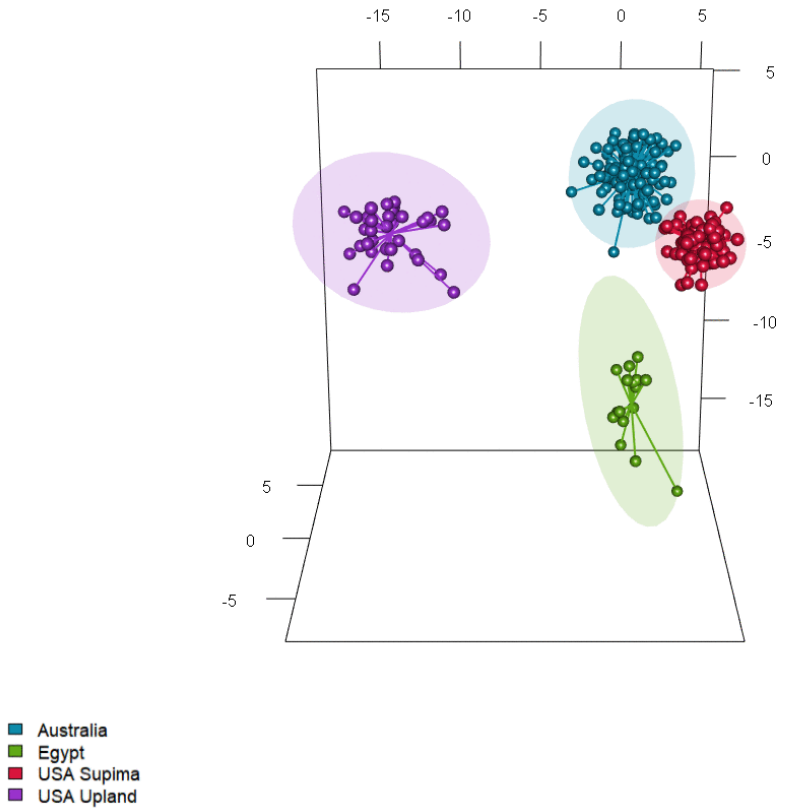
Build – create a construct to describe the data

PCA/LDA model of reference data

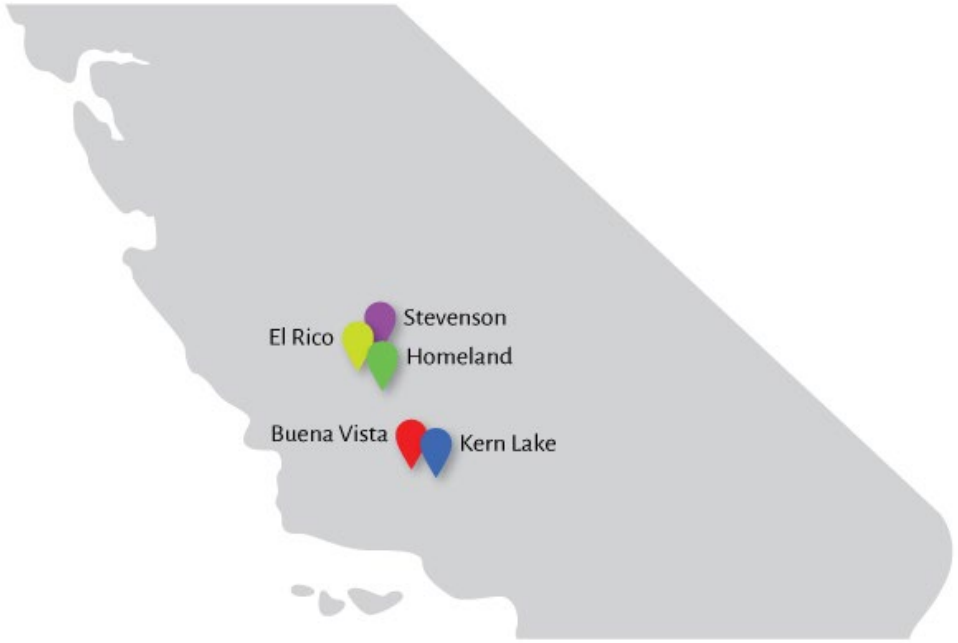
Calibrate – find out how well the construct is describing the data

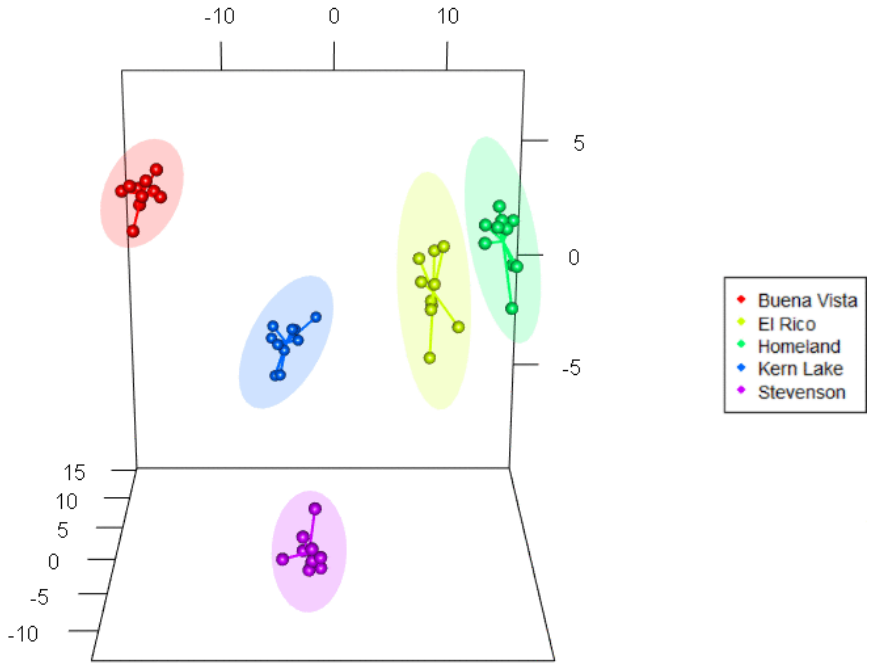
Cross-validation of model performance

Determination of Adequacy



Determination of Adequacy





Validation of Model Performance
Sampling of audit test sets
Determine model robustness by testing model performance

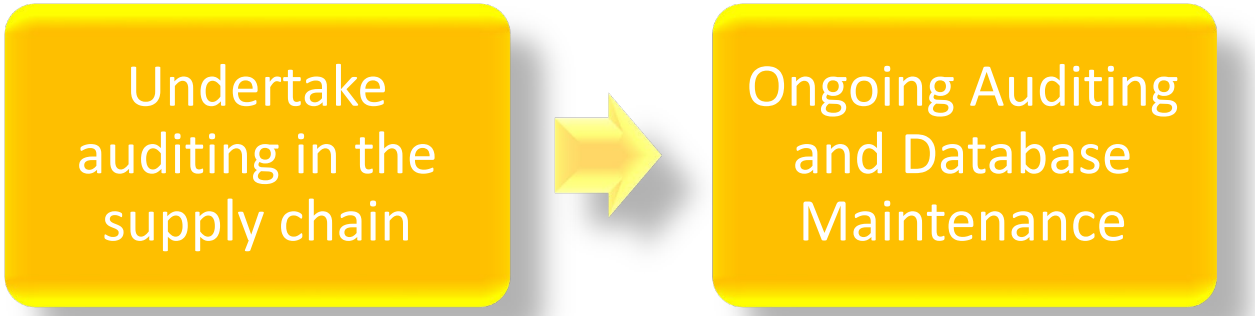
Validation of Performance

- Validity of models
  - Internal Cross Validation methods
  - External Cross Validation
    - Test data sets of known origin
    - Blind testing
- Determine error rates, specificity and sensitivity
- Make sure we have appropriate sample sizes

Model Performance Testing

- Stability of chemical fingerprint
  - Temporal changes across the system
  - Changes due to farm management systems
  - Changes in climate
  - Varietal effects

- Blind Audits

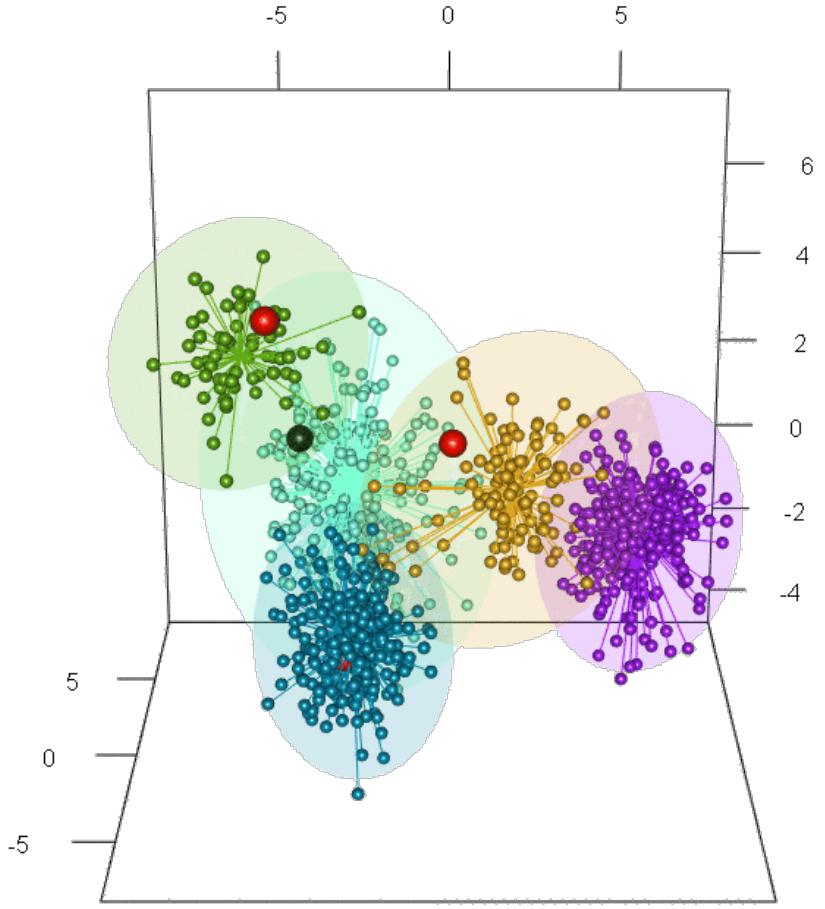

Undertake auditing in the supply chain
Ongoing Auditing and Database Maintenance

Origin Verification Testing

Based on specific origin, resolution, and type of product –

“Is sample X consistent with the origin fingerprint of the Claimed origin?”

- Brazil
- Egypt
- Pakistan
- Turkey
- USA
- **Audit**: Non-genuine (USA)
- **Audit:** Genuine (USA)



POSTGRADUATE

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

Origin Traceability of Coffee

Demonstration study

- Nescafé selected 22 farms that were tested and graded by the BSCA (Brazilian Specialty Coffee Association).

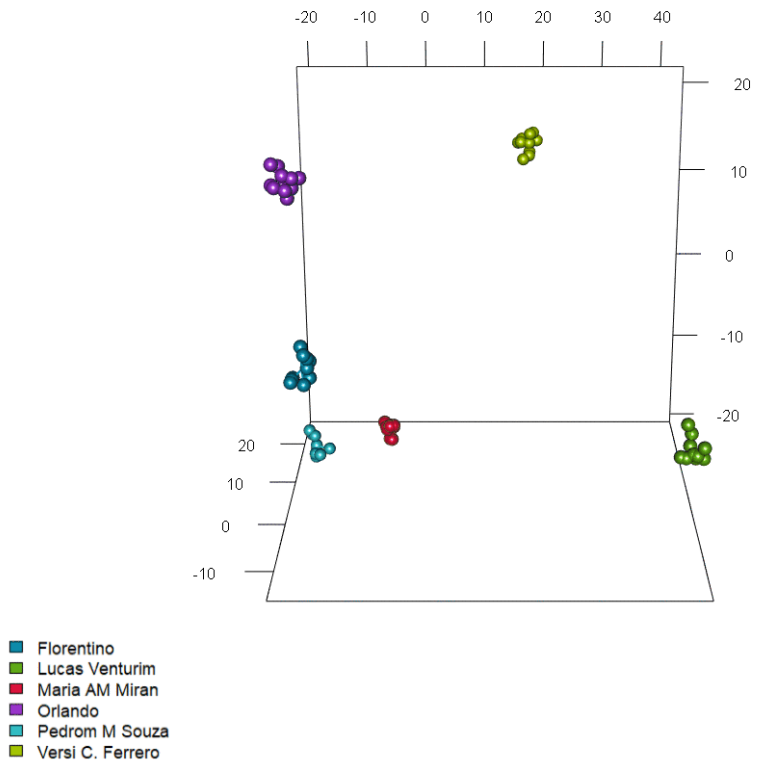
7 of these farm were selected as the ones producing the highest quality coffee beans.

Farm Name	Address
Florentino	Sitio Rancho Dantas, Montanhas Capixabas, Brejetuba, ES
Lucas Venturim	Fazenda Venturim, Conilon Capixaba, Sao Domingos do Norte, ES
Maria AM Miran	Fazenda Edio Miranda, Matas de Minas, Araponga, MG
Orlando	Fazenda Pilar, Norte Pioneiro do Parana, Cornelio Procopio, PR
Pedrom M Souza	Fazenda Horizonte, Chapada Diamantina, Piata, BA
Versi C Ferrero	Fazenda Bau, Sul de Minas, Itamogi, MG
Fabio P Abreu	Sitio Forquilha de Rio, Caparao, Espera Feliz, MG

Results – Differentiation by Farm

- The statistical models showed good differentiations between the supplied farms
  - However it must be noted that this database has not been approved as being commercially Fit for Purpose
  - For a commercial project we have Standard Operating Procedures to collect the right number of samples to ensure we capture all the natural variability of the farm
  - However given the dataset we had the models performed very well as per the below

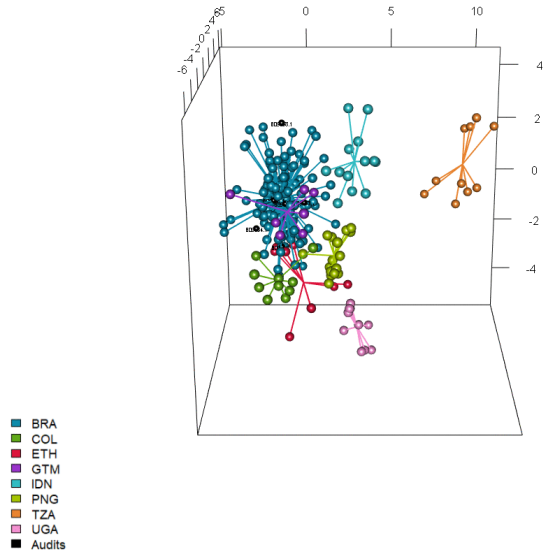
Performance rates	Fabio P Abreu	Florentino	Lucas Venturim	Orlando	Pedrom M Souza	Versi C Ferrero
True Positive	>95%	>95%	>95%	>95%	>95%	>95%
True Negative	>80%	>80%	>80%	>80%	>80%	>80%

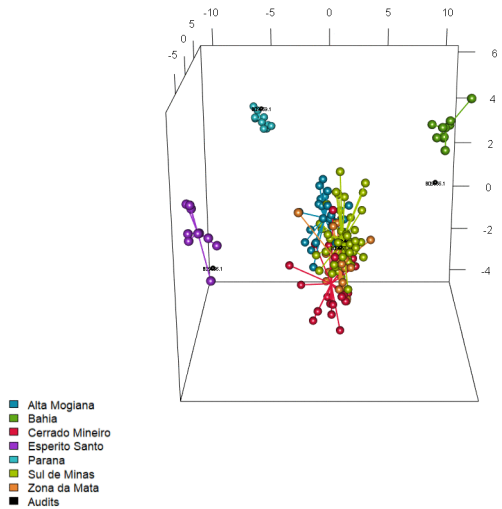


Differentiation by region

- While our Proof of concept was designed to differentiate at farm level we also compared the samples with our databases for being consistent with Brazil and the Regions.

Sample number	Purported Country of Origin	Conclusion	Purported Region of Origin	Conclusion
B09653.1	Brazil	*Pass*	Alta Mogiana	*Pass*
B09654.1	Brazil	*Pass*	Sul de Minas	*Pass*
B09655.1	Brazil	*Pass*	Bahia	*Pass*
B09656.1	Brazil	*Pass*	Esperito Santo	*Pass*
B09658.1	Brazil	*Pass*	Sul de Minas	*Pass*
B09659.1	Brazil	*Pass*	Parana	*Pass*



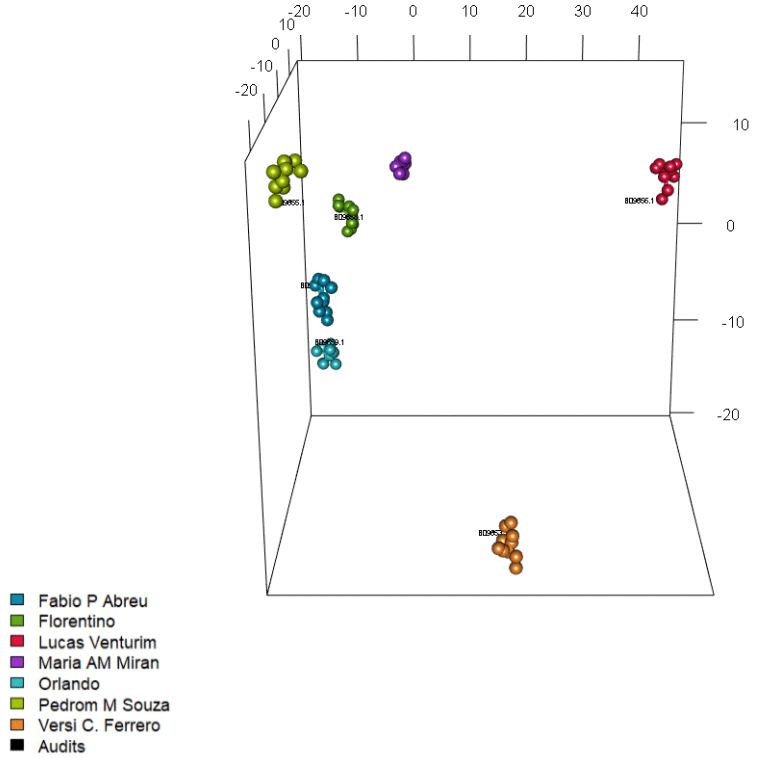


(1) To note is that some of the regions are in close proximity within the graph, this over lap is resolved at higher resolutions and using a set of different statistical models.
(2) To note is that we have different number of samples per region in this dataset, which affects the capability of the models to differentiate one region from another.
(3) The graphs are purely for representative reasons and it should be considered that this is only one statistical model of the many used to determine consistency/inconsistency with a purported origin.
(4) The databases are not Fit for Purpose hence for demonstration purposes only.

Samples tested against supplying farms

Additional samples were supplied for the audit demonstration

Sample number	Purported Origin Address	Purported Farm	Conclusion
B09653.1	Fazenda Bau, Sul de Minas, Itamogi, MG	Versi C Ferrero	*Consistent*
B09654.1	Sitio Forquilha de Rio, Caparao, Espera Feliz, MG	Fabio P Abreu	*Consistent*
B09655.1	Fazenda Horizonte, Chapada Diamantina, Piata, BA	Pedrom M Souza	*Consistent*
B09656.1	Fazenda Venturim, Conilon Capixaba, Sao Domingos do Norte, ES	Lucas Venturim	*Consistent*
B09658.1	Sitio Rancho Dantas, Montanhas Capixabas, Brejetuba, ES	Florentino	*Consistent*
B09659.1	Fazenda Pilar, Norte Pioneiro do Parana, Cornelio Procopio, PR	Orlando	*Consistent*



(1) To note is that some of the regions are in close proximity within the graph, this over lap is resolved at higher resolutions and using a set of different statistical models.
(2) To note is that we have different number of samples per region in this dataset, which affects the capacity of the modelsto differentiate one region from another.
(3) The graphs are purely for representative reasons and it should be considered that this is only one statistical model of the many used to determine consistency/inconsistency with a purported origin.
(4) The databases are not Fit for Purpose hence for demonstration purposes only.
(5) The graphs displayed in this presentation only represent the 7 farms of interest to Nescafé. The low number of samples will not capture all variation of a farm. Additionally, no high-risk origins

POSTGRADUATE

Rapid Origin Traceability of Coffee

Joy SIM
University of Otago
PhD Candidate

Russell FREW, Indrawati OEY, Cushla MCGOVERIN, Biniam KEBEDE

Research Objective: To build an effective and rapid toolbox for coffee origin traceability

ORITAIN™
PROVING ORIGIN, PROTECTING REPUTATIONS

Geochemistry

Stable isotopes
Trace elements

Research Objective: To build an effective and rapid toolbox for coffee origin traceability

UNIVERSITY of OTĀGO
Te Whare Wānanga o Otāgo
NEW ZEALAND
SAPERE AUDE

ORITAIN™
PROVING ORIGIN, PROTECTING REPUTATIONS

Geochemistry

Stable isotopes
Trace elements

Metabolomics

(Non) volatile

Research Objective: To build an effective and rapid toolbox for coffee origin traceability

ORITAIN™
PROVING ORIGIN, PROTECTING REPUTATIONS

Geochemistry

Stable isotopes
Trace elements

Destructive

Destructive

Metabolomics

(Non) volatile

Research Objective: To build an effective and rapid toolbox for coffee origin traceability

ORITAIN™
PROVING ORIGIN, PROTECTING REPUTATIONS

Geochemistry

Stable isotopes
Trace elements

Destructive

Rapid/Non-destructive

Vibrational spectroscopy

Near infrared
Mid infrared
Raman
Hyperspectral

Destructive

Metabolomics

(Non) volatile

Research Objective: To build an effective and rapid toolbox for coffee origin traceability

Research Strategy

ORITAIN™
PROVING ORIGIN, PROTECTING REPUTATIONS

Geochemistry

Stable isotopes
Trace elements

Destructive

Rapid/Non-destructive

Vibrational spectroscopy

Near infrared
Mid infrared
Raman
Hyperspectral

Multi-omics Machine Learning

Destructive

Metabolomics

(Non) volatile

Research Objective: To build an effective and rapid toolbox for coffee origin traceability

Research Strategy

ORITAIN™
PROVING ORIGIN, PROTECTING REPUTATIONS

Geochemistry

Stable isotopes
Trace elements

Destructive

Rapid/Non-destructive

Vibrational spectroscopy

Near infrared
Mid infrared
Raman
Hyperspectral

Multi-omics Machine Learning

Destructive

Metabolomics

(Non) volatile

Research Objective: To build an effective and rapid toolbox for coffee origin traceability

Research Design

1. Sample/Processing

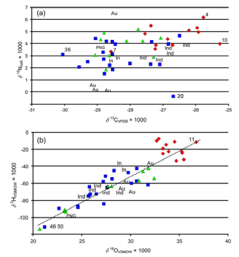
ORITAIN™
PROVING ORIGIN, PROTECTING REPUTATIONS

Green bean, Ground/ Whole

2. Analysis

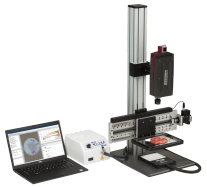
Multi-omics

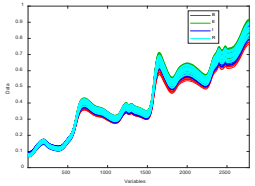
(1) Geochemical



Stable isotopes
Trace elements

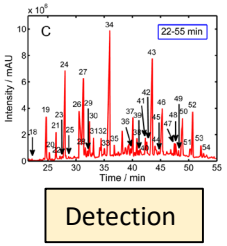
(2) Vibrational Spectroscopy





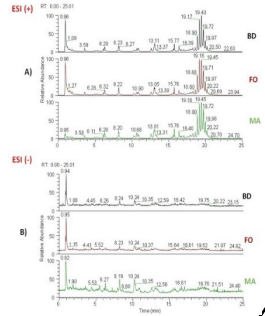
(3) Metabolomics

Sample preparation



Detection

Separation

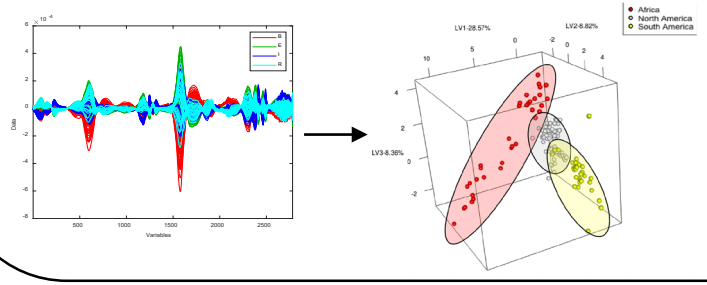


3. Data analysis

Machine Learning

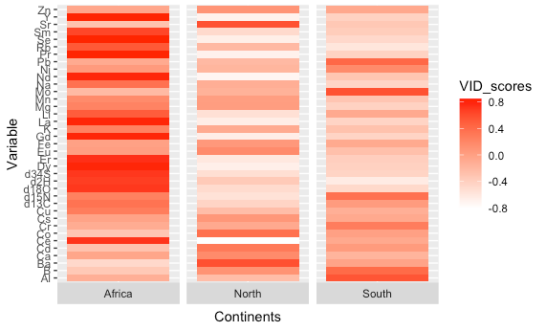
(1) Development of individual models

Experiments (1) to (3) data are pre-processed individually using various machine learning tools (linear/non-linear). Effective, rapid, and non-destructive models developed



(2) Feature identification

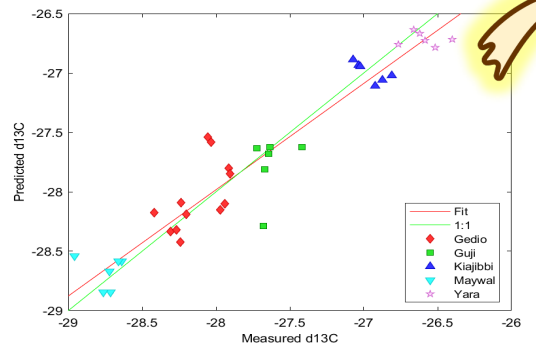
Determine features or markers important for origin verification



Decision making

(3) Determine relationships for prediction

Using rapid methods to predict concentrations of discriminating features



Q1: Green/Roasted

Q2: Which instrument?

Q3: Data mining

Q1: Green/Roasted

Q2: Which instrument?

Q3: Data mining

Q1: Green/Roasted

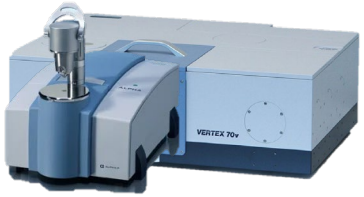
SCA

Q2: Which instrument?

Raman

Dispersive NIR (near-infrared)

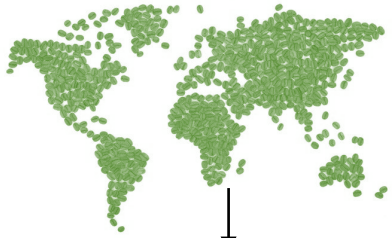
Hyperspectral-NIR



ATR-FTIR (mid-infrared)

Q3: Data mining

Q1: Green/Roasted

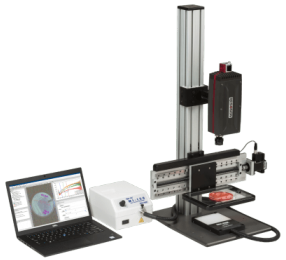


SCA

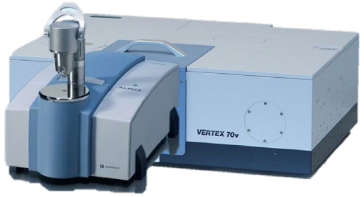
Q2: Which instrument?

Raman

Dispersive NIR (near-infrared)



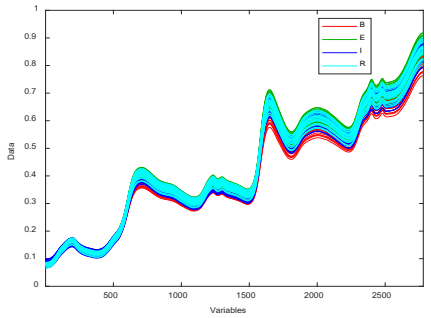
Hyperspectral-NIR

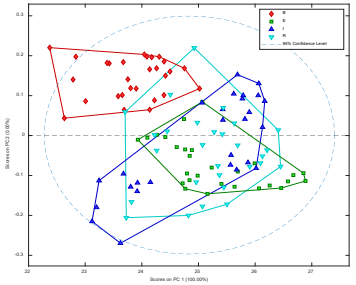


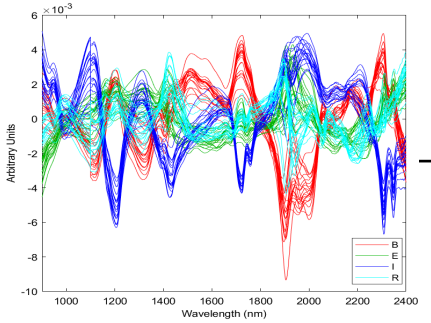
ATR-FTIR (mid-infrared)

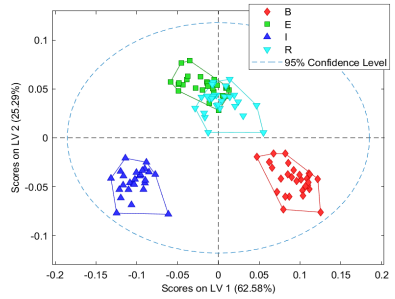
Q3: Data mining

PREPROCESSING









UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND
SAPERE AUDE

Q1: Green/Roasted

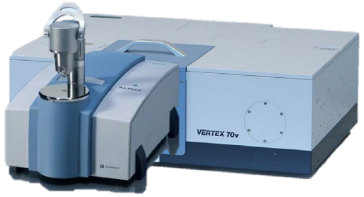
SCA

Q2: Which instrument?

Raman

Dispersive NIR (near-infrared)

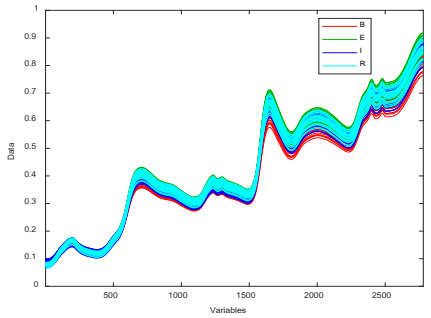
Hyperspectral-NIR

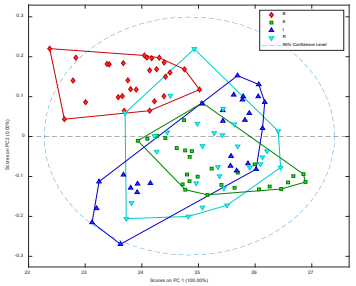


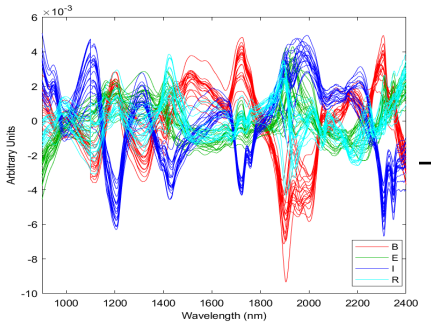
ATR-FTIR (mid-infrared)

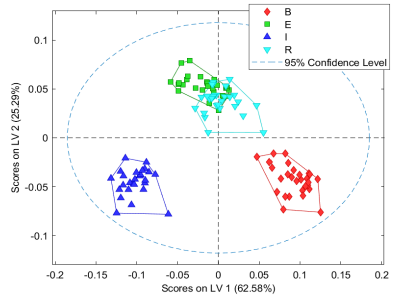
Q3: Data mining

PREPROCESSING









Q1: Green/Roasted

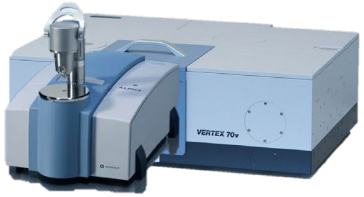
SCA

Q2: Which instrument?

Raman

Dispersive NIR (near-infrared)

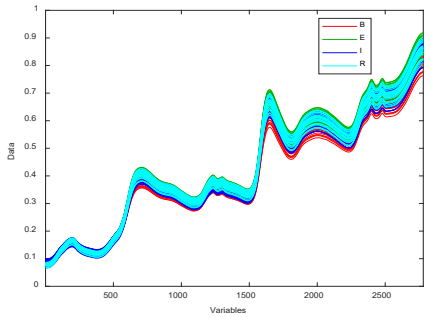
Hyperspectral-NIR

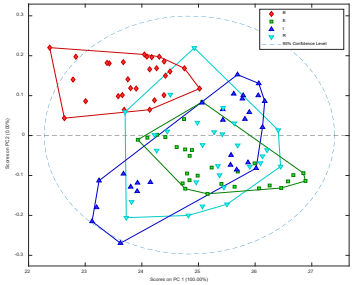


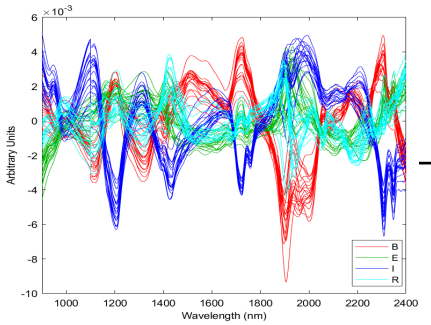
ATR-FTIR (mid-infrared)

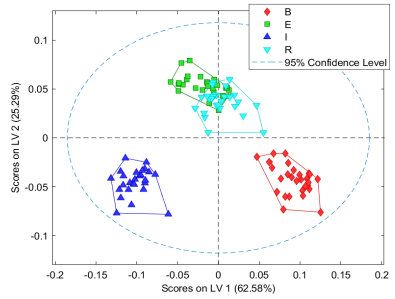
Q3: Data mining

PREPROCESSING









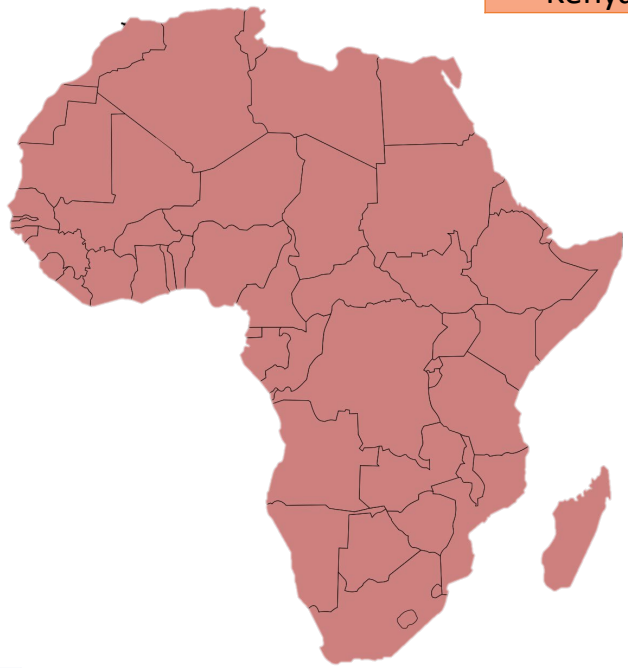
Central America

Costa Rica	West Valley
Costa Rica	Tarrazu
Costa Rica	Santa Maria de Dota
Guatemala	El Progreso
Guatemala	Guatemala
Guatemala	Huehuetenango
Mexico	Oaxaca
Mexico	Puebla
Mexico	Estado De Mexico
Nicaragua	Estelí
Nicaragua	Matagalpa
Nicaragua	Nueva Segovia

South America

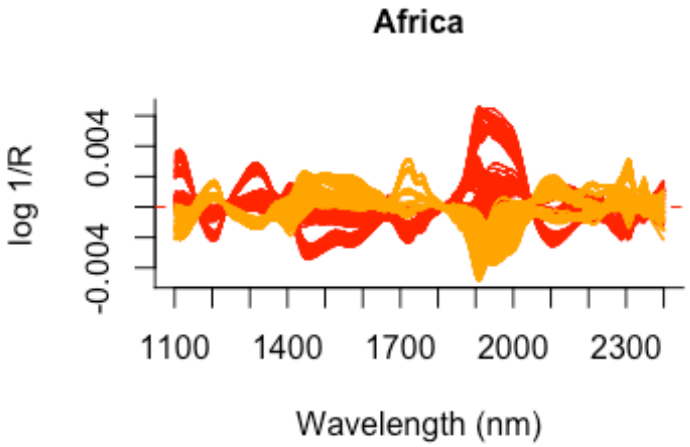
Colombia	Antioquia
Colombia	Huila
Colombia	Narino
Peru	Cusco
Peru	Cusco
Peru	Pasco

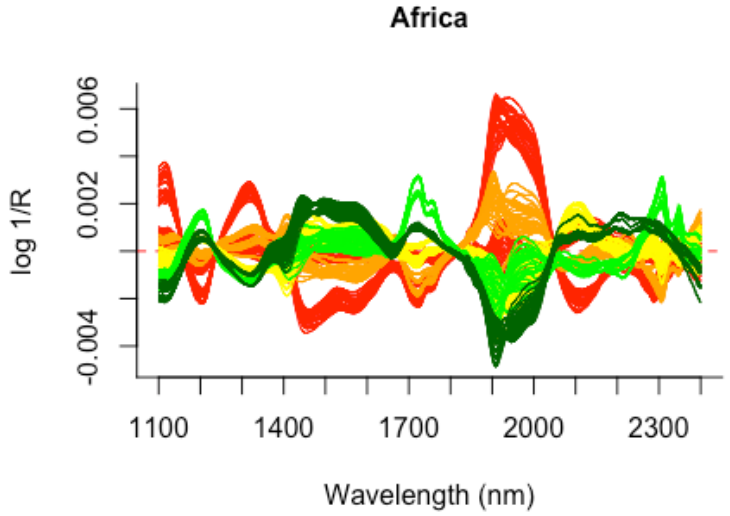
Africa



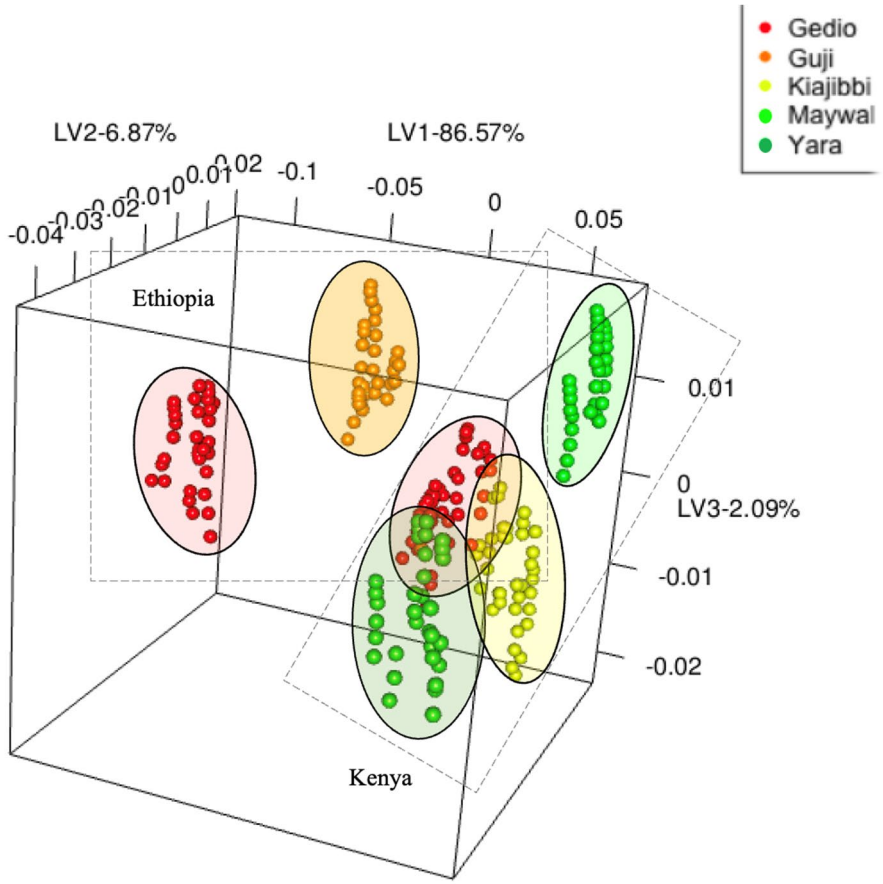
Ethiopia	Gedio
Ethiopia	Gedio
Ethiopia	Guji
Kenya	Maywal
Kenya	Kiajibbi
Kenya	Yara

Africa

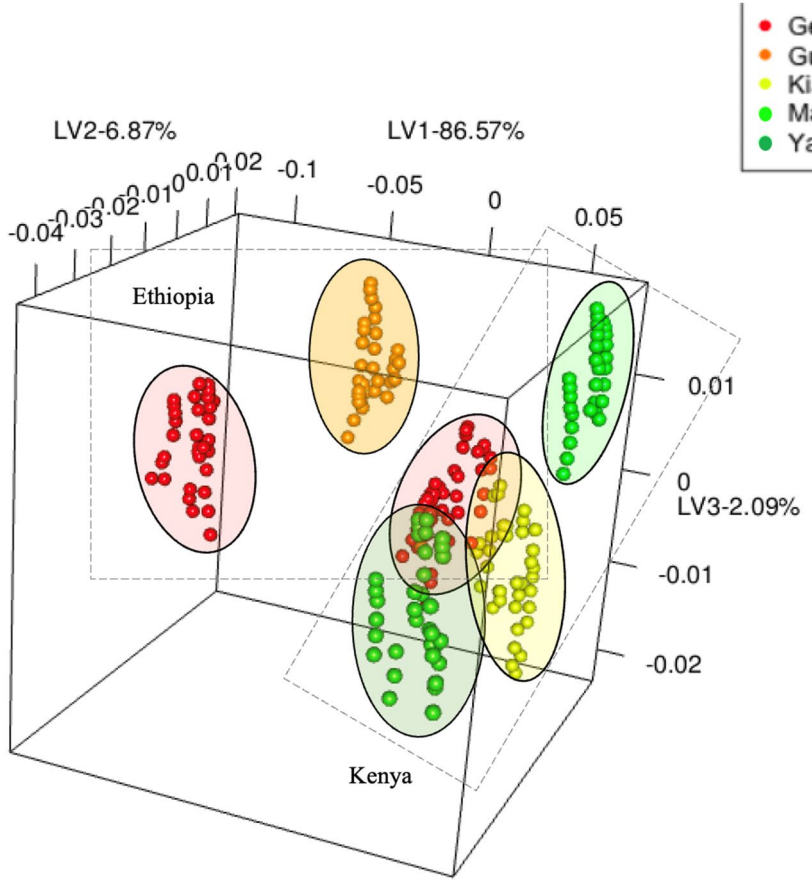

Africa
log 1/R
0.004
-0.004
1100
1400
1700
2000
2300
Wavelength (nm)

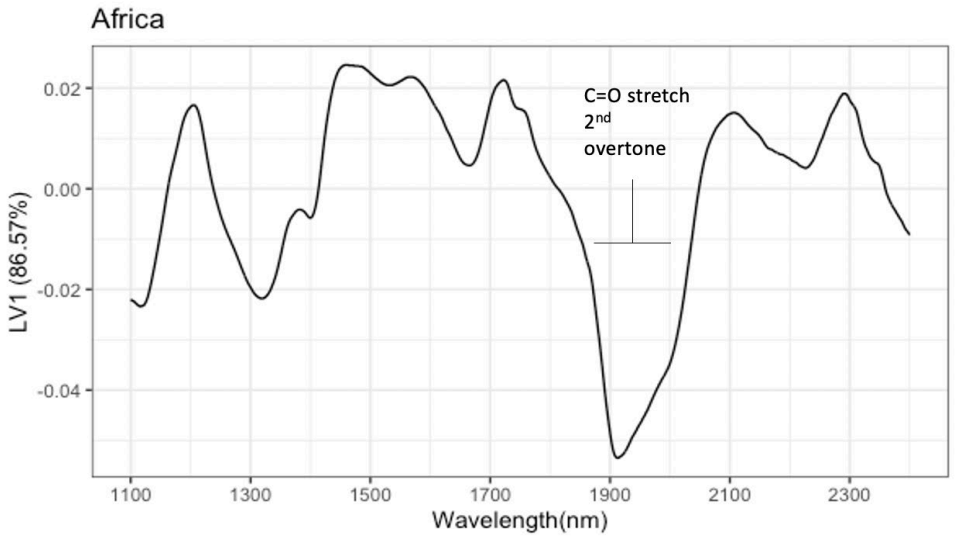

Africa
log 1/R
0.006
0.002
-0.004
1100
1400
1700
2000
2300
Wavelength (nm)

Africa


Gedio
Guji
Kiajibbi
Maywal
Yara
LV2-6.87%
LV1-86.57%
LV3-2.09%
Ethiopia
Kenya

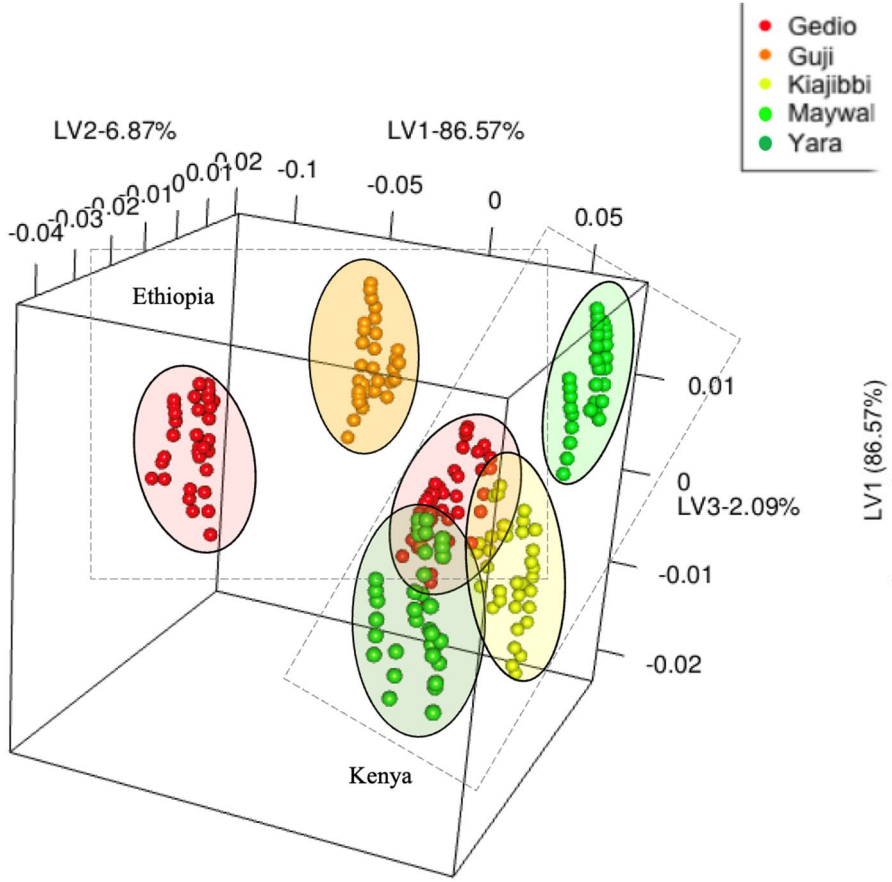
Africa

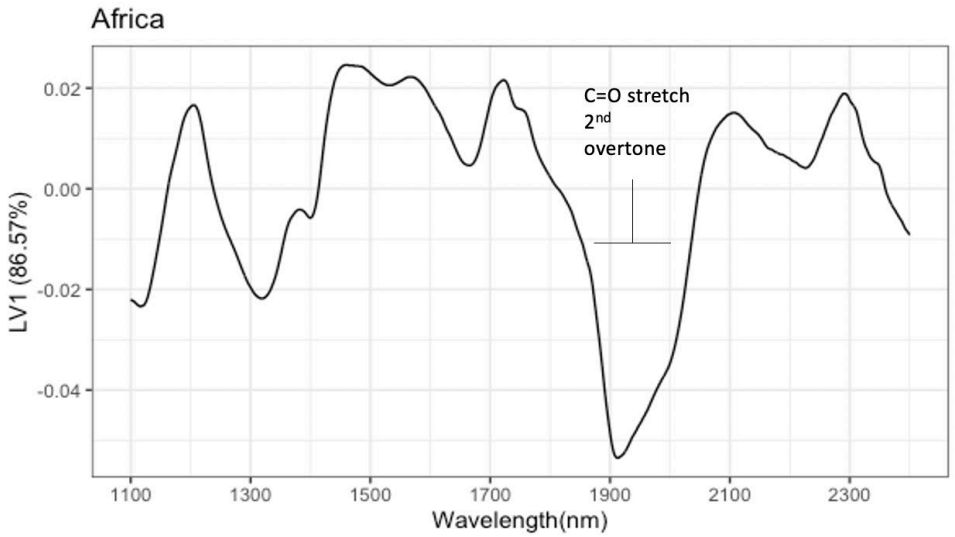

LV2-6.87%
LV1-86.57%
LV3-2.09%
Ethiopia
Kenya
Gedio
Guji
Kiajibbi
Maywal
Yara


Africa
C=O stretch
2nd
overtone
LV1 (86.57%)
Wavelength(nm)

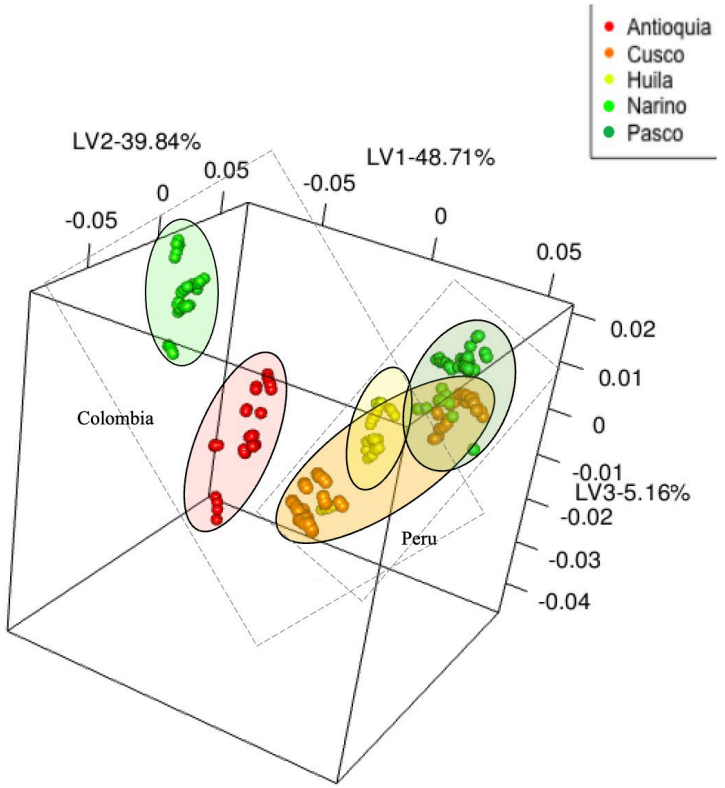
RMSEC 0.2485
RMSEP 0.2430

Accuracy, CV: 0.76
Accuracy, pred: 0.79

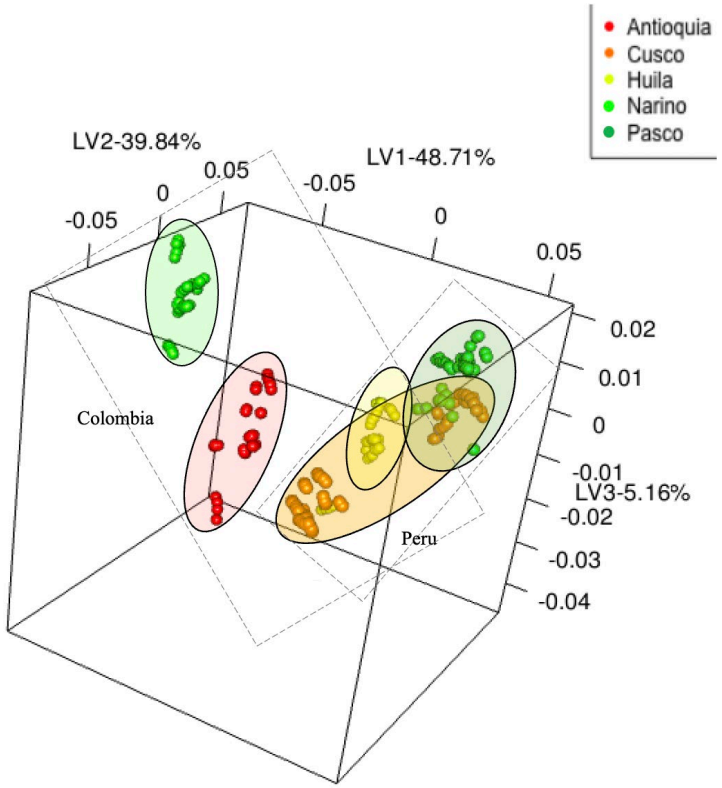


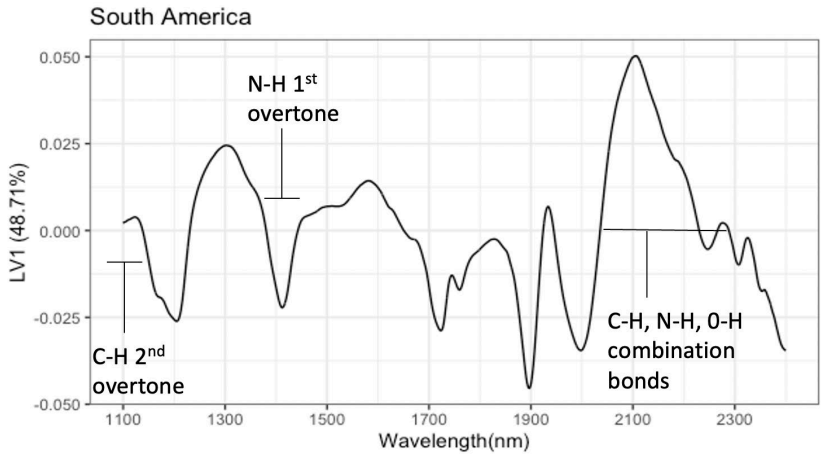


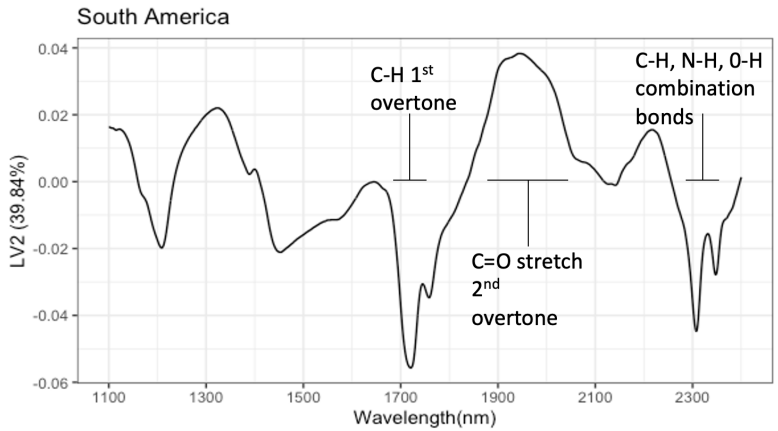
South America

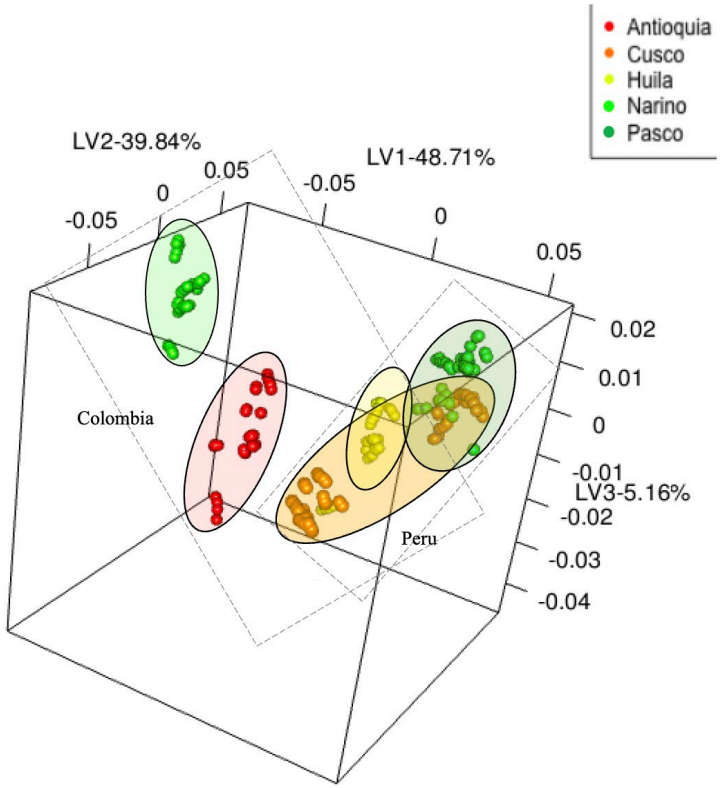

Antioquia
Cusco
Huila
Narino
Pasco
LV2-39.84%
0.05
0
-0.05
LV1-48.71%
-0.05
0
0.05
0.02
0.01
0
-0.01
LV3-5.16%
-0.02
-0.03
-0.04
Colombia
Peru

South America


Antioquia
Cusco
Huila
Narino
Pasco
LV2-39.84%
LV1-48.71%
LV3-5.16%
Colombia
Peru

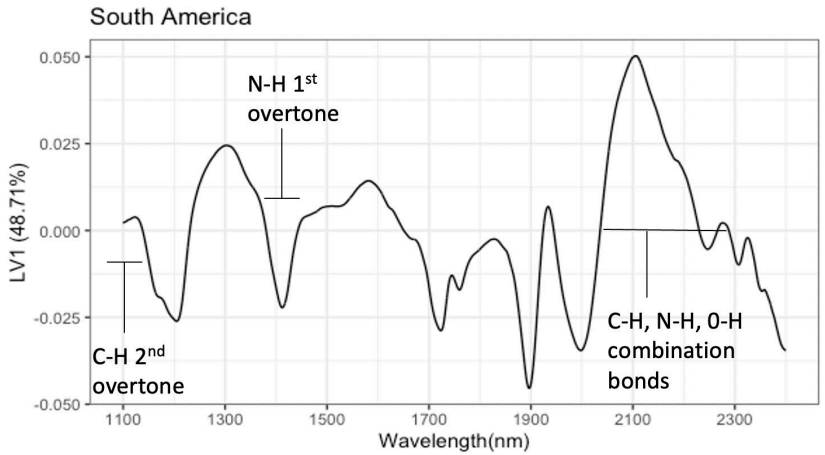

South America
N-H 1st overtone
C-H 2nd overtone
C-H, N-H, 0-H combination bonds
LV1 (48.71%)
Wavelength(nm)

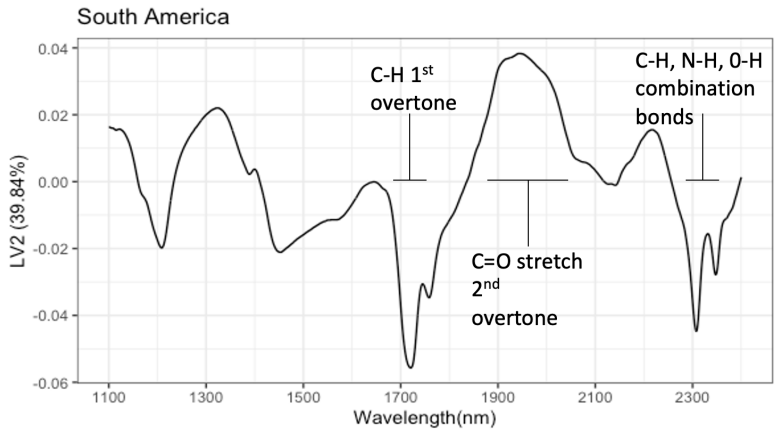

South America
C-H 1st overtone
C=O stretch 2nd overtone
C-H, N-H, 0-H combination bonds
LV2 (39.84%)
Wavelength(nm)



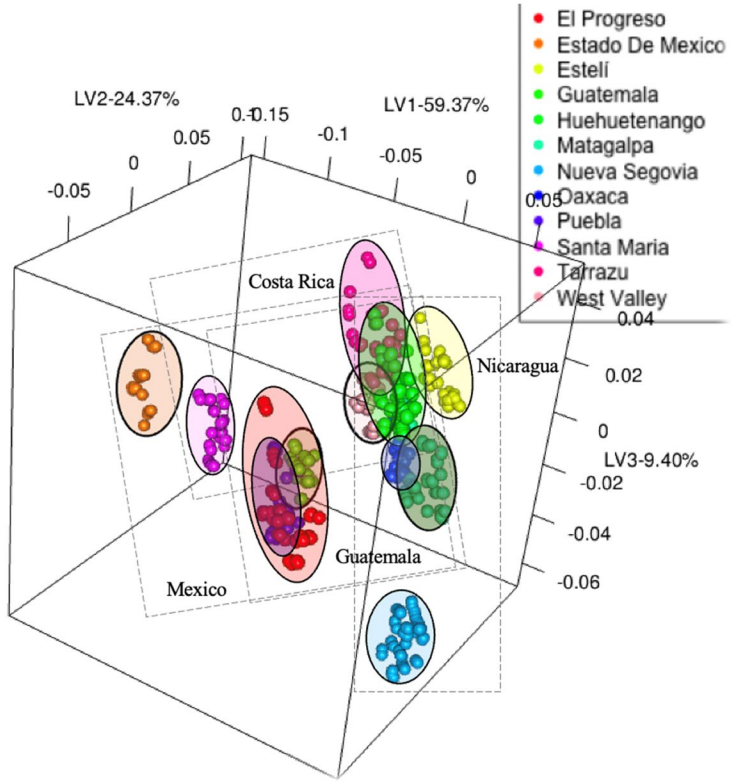
RMSEC: 0.2207
RMSEP: 0.2462

Accuracy, CV: 0.79
Accuracy. Pred: 0.70

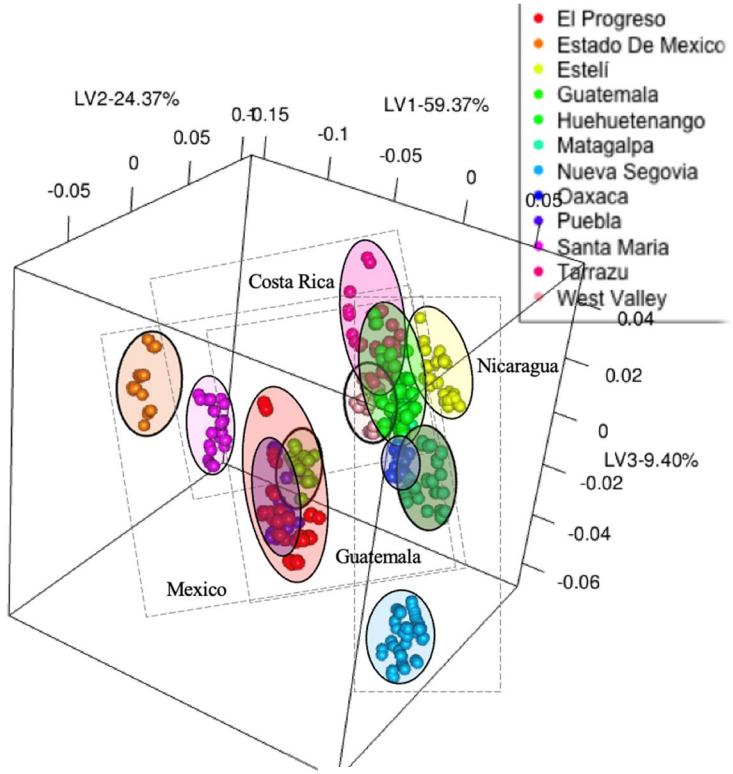


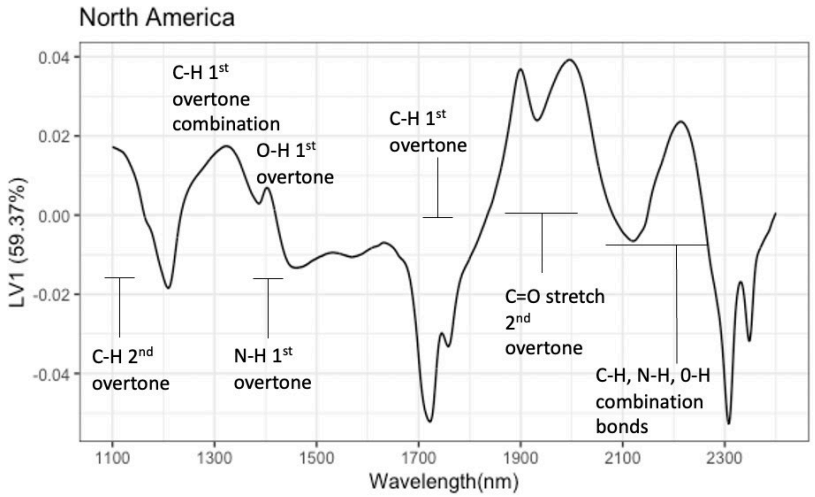


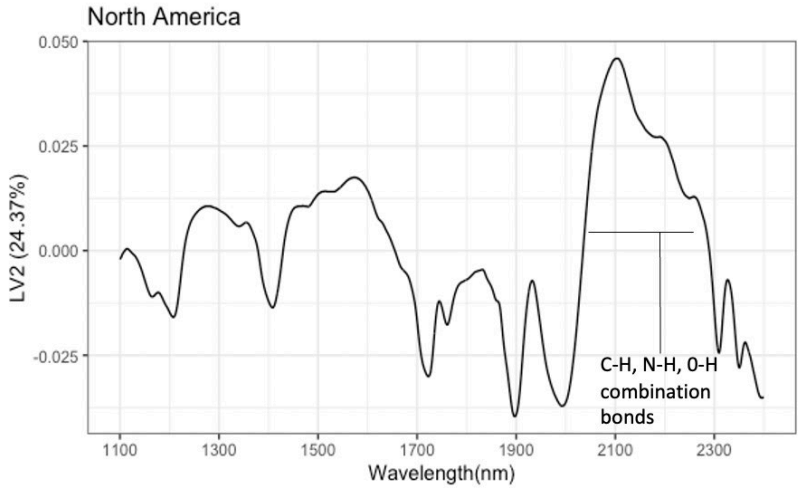
Central America



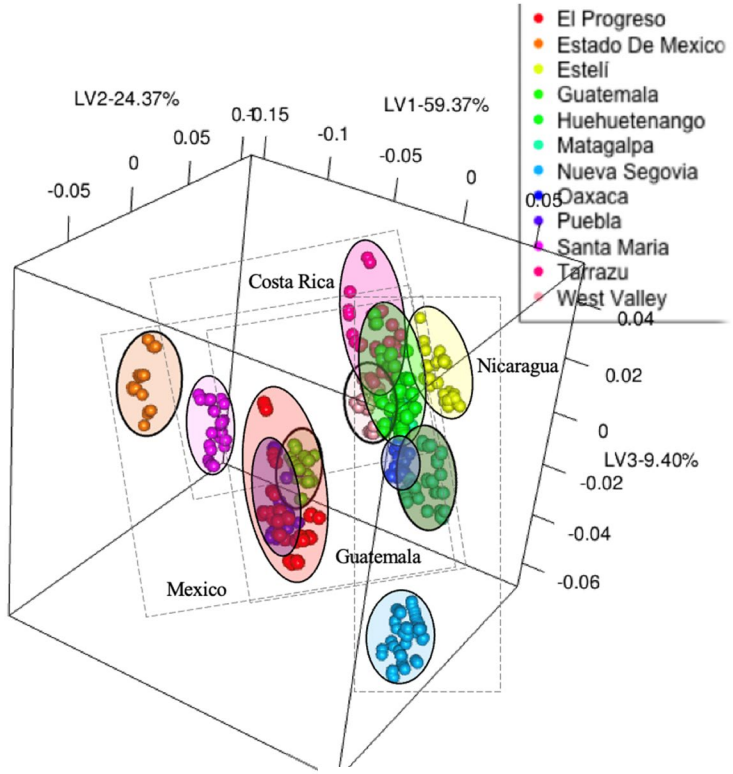
Central America



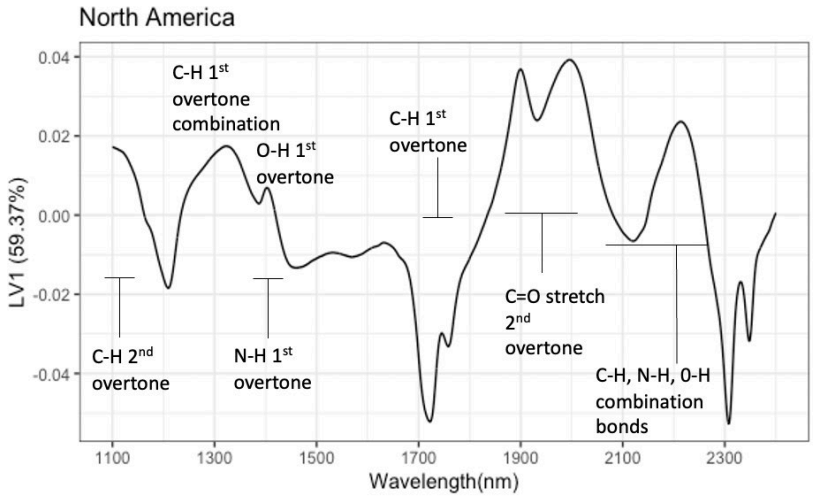


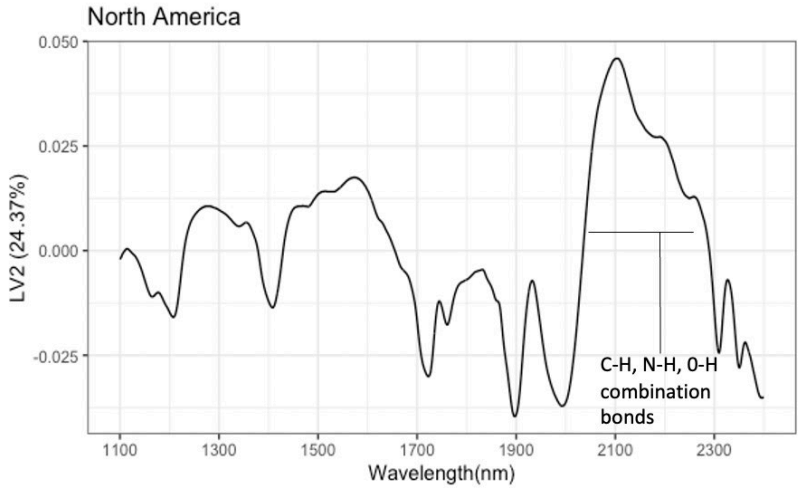


Central America



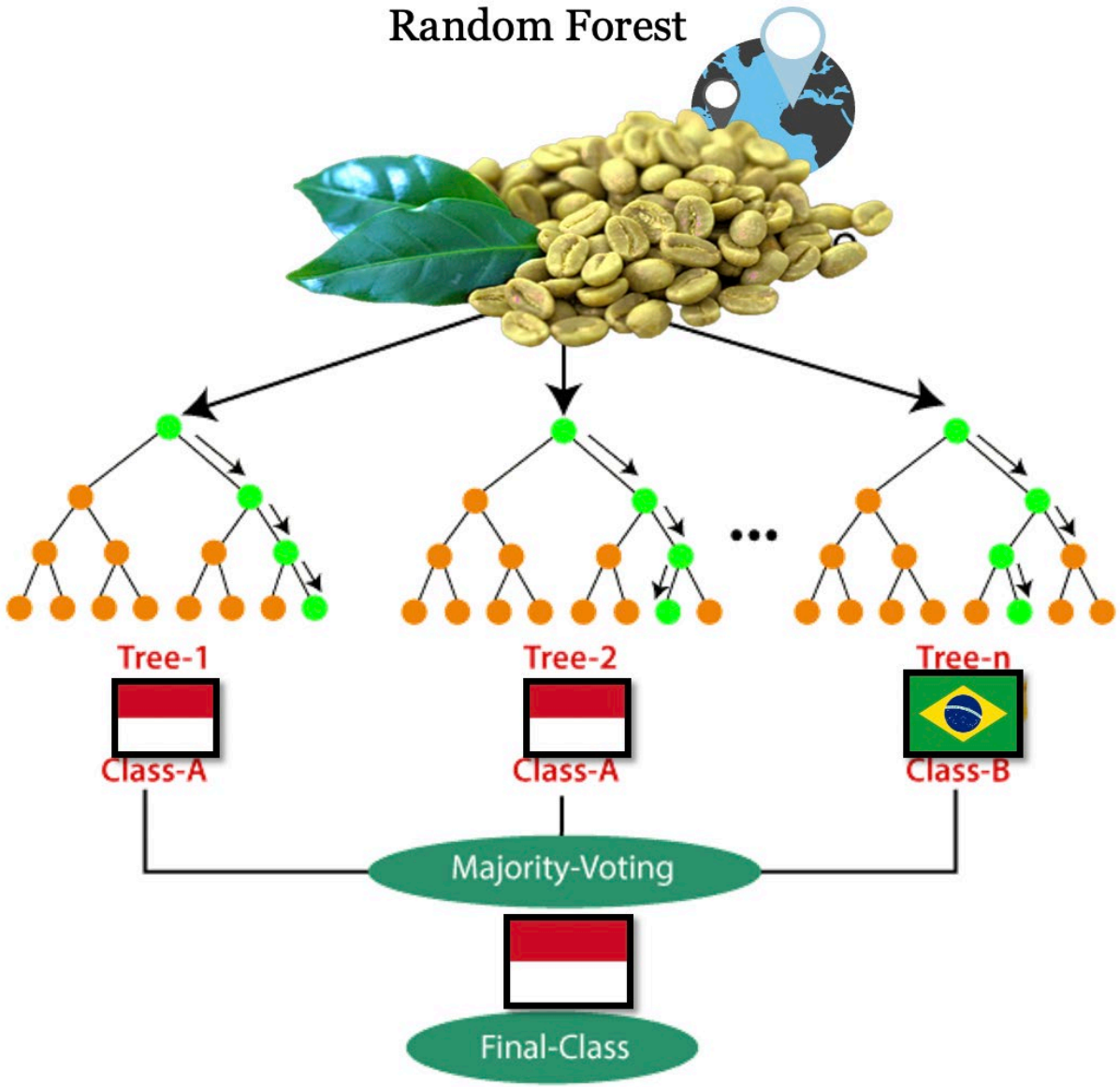
RMSEC: 0.2214
RMSEP: 0.2257
Accuracy, CV: 0.65
Accuracy, Pred: 0.64



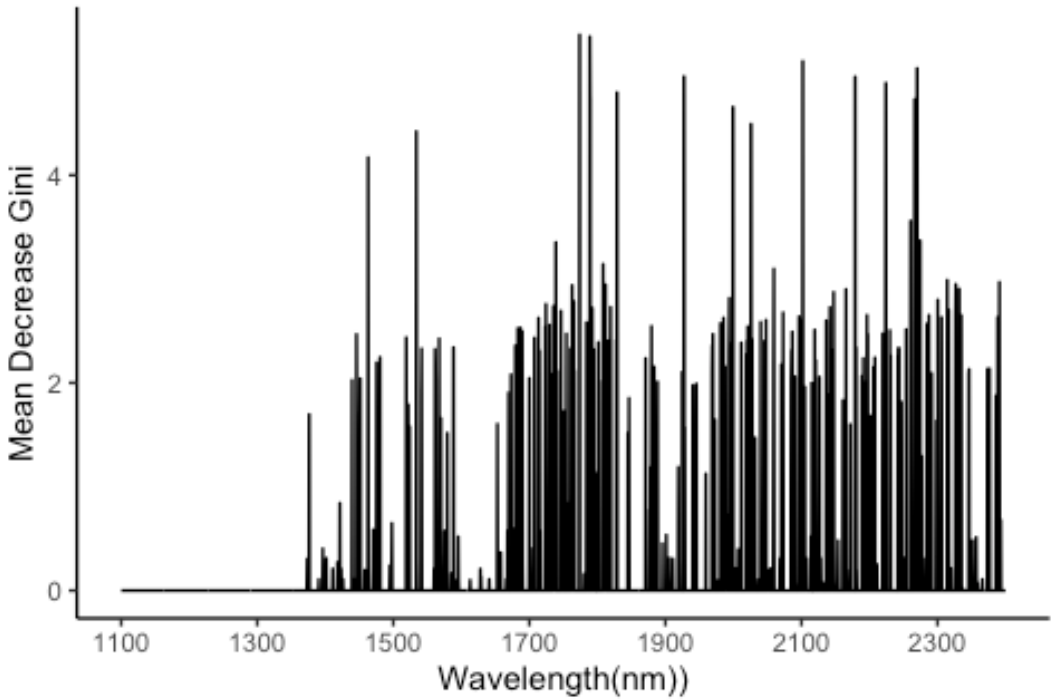


Central America

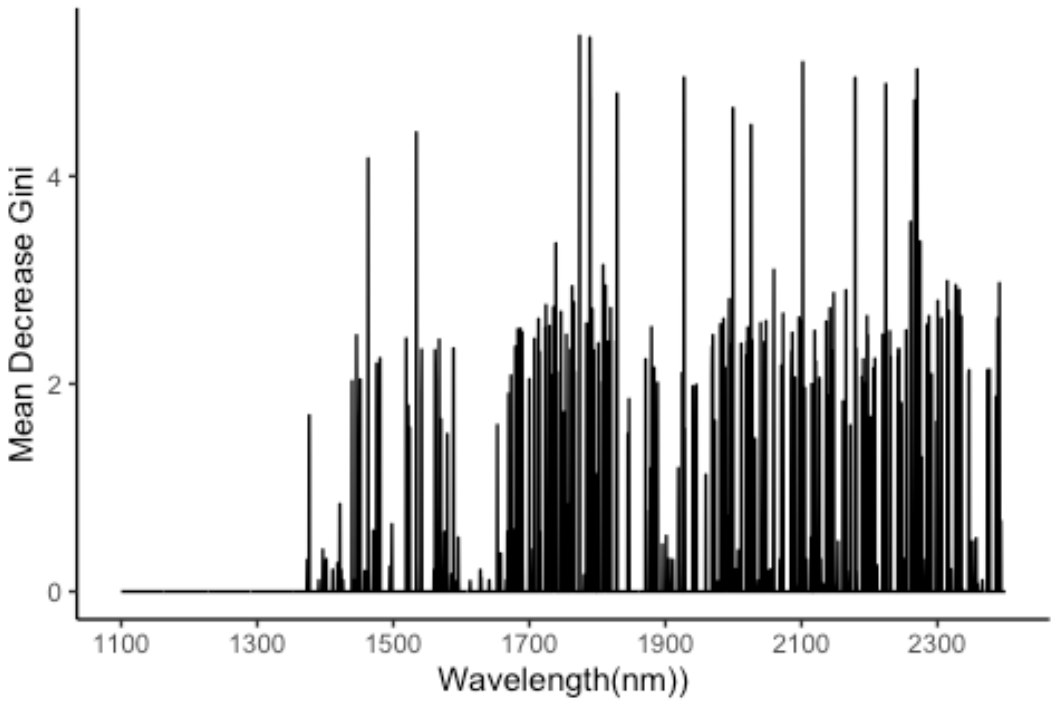
Central America

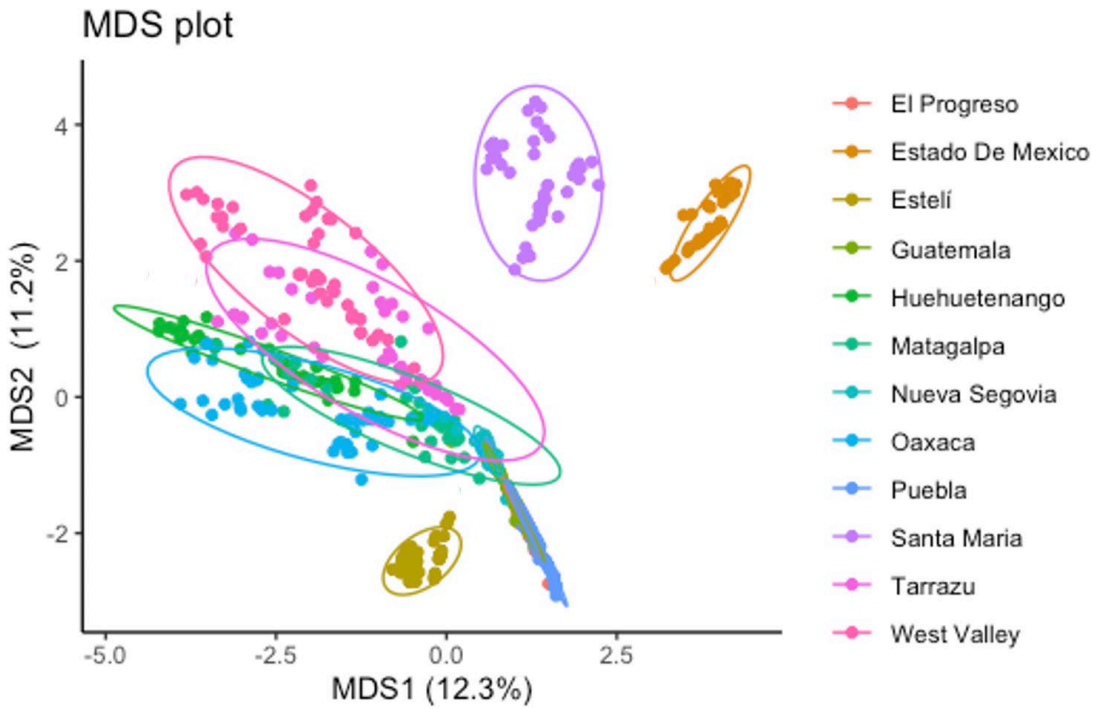


Central America



Central America

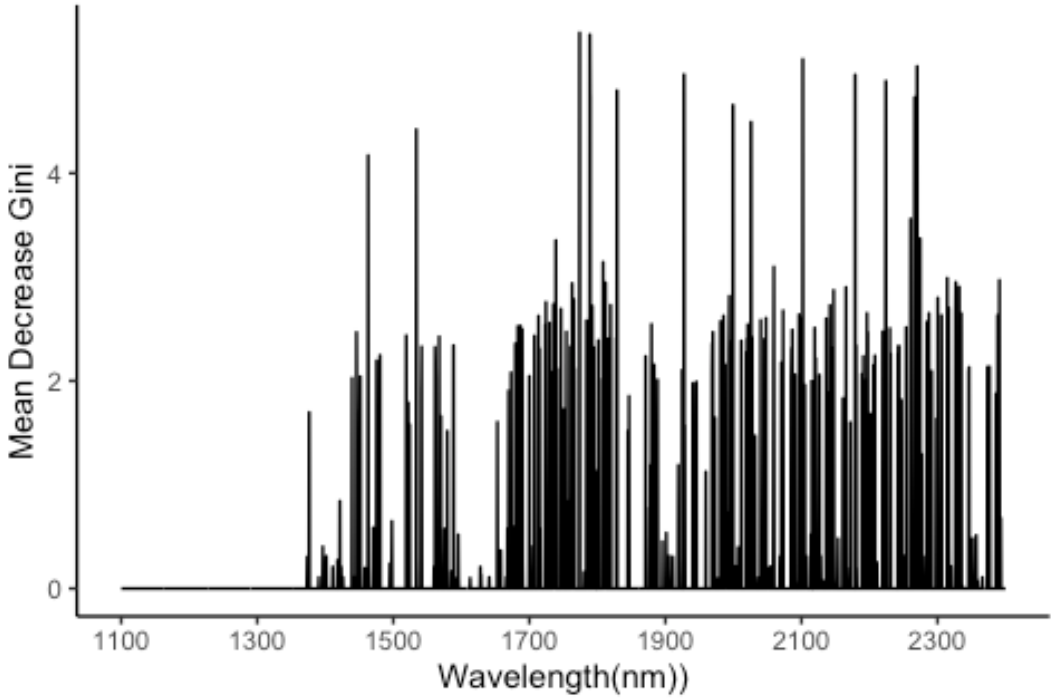


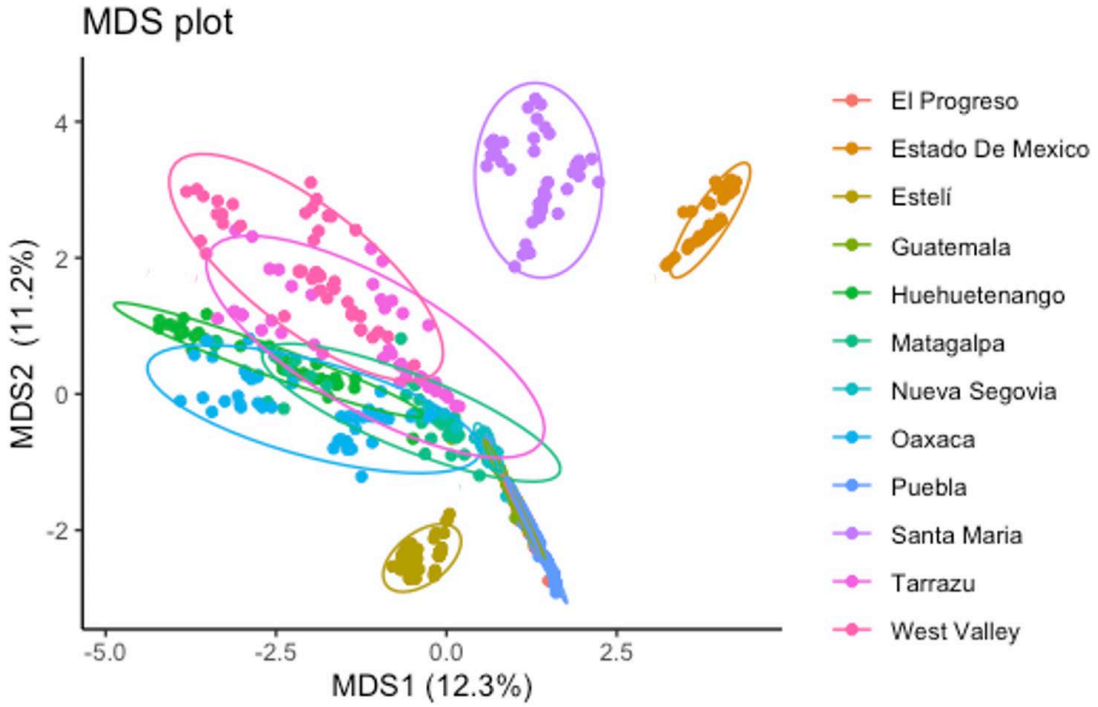


Central America

Mtry: 4
nTrees: 18

Accuracy, Pred: 0.92





Conclusion & Future directions

Conclusion & Future directions

Research Objective: To build an effective and rapid toolbox for coffee origin traceability

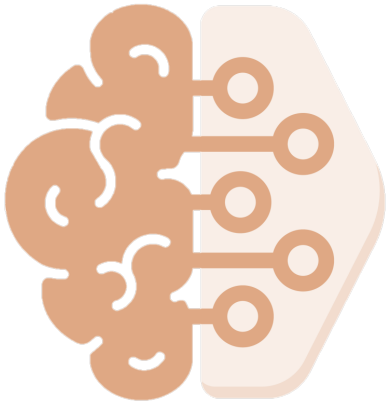
Rapid
Sustainable
Solvent-less
Cost-effective

Conclusion & Future directions

Research Objective: To build an effective and rapid toolbox for coffee origin traceability

Rapid
Sustainable
Solvent-less
Cost-effective

Non-linear
machine
learning

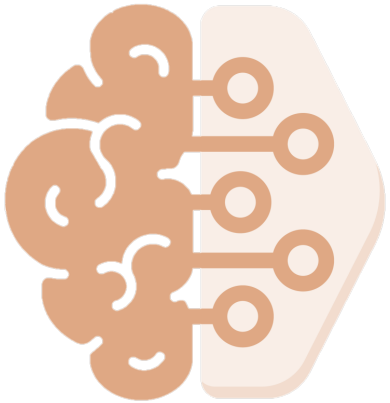


Conclusion & Future directions

Research Objective: To build an effective and rapid toolbox for coffee origin traceability

Rapid
Sustainable
Solvent-less
Cost-effective

Non-linear
machine
learning



Non-destructive

Acknowledgements

Dr Biniam Kebede

Professor Indrawati Oey

Professor Russell Frew

Dr Cushla Mcgoverin