CHIMIOMETRIE XVII
January 18-20th, Namur, Belgium

Application of similarity parameters to discriminate and classify herbal medicines

Johan Viaene[1], Goedele Alaerts[1], Bieke Dejaegher[1,2]
Yvan Vander Heyden[1]

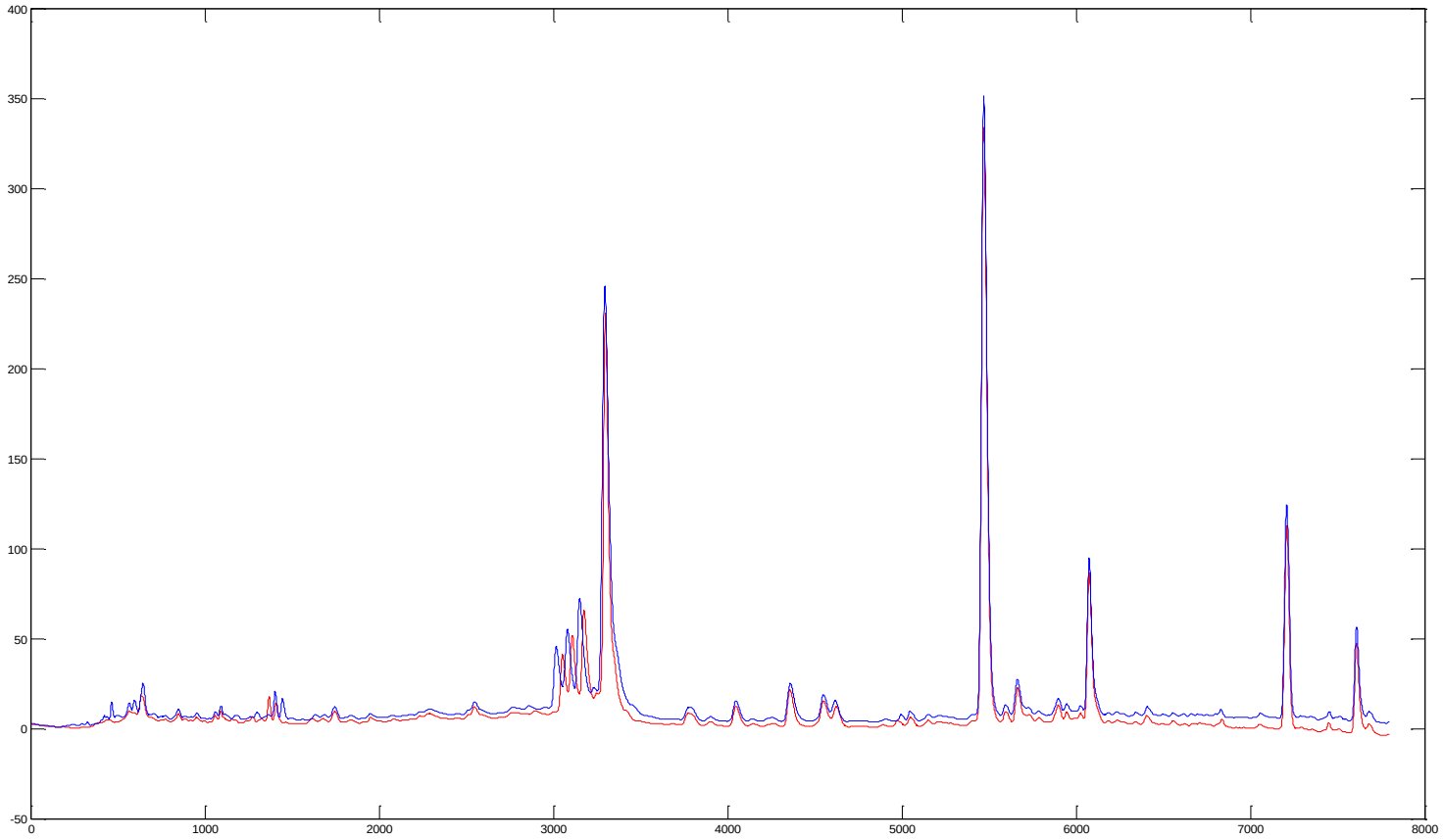
[1]Department of Analytical Chemistry and Pharmaceutical Technology (FABI), Center for Pharmaceutical Research (CePhar), Vrije Universiteit Brussel (VUB), Laarbeeklaan 103, B-1090 Brussels, Belgium

[2]Laboratory of Instrumental Analysis and Bioelectrochemistry, Institute of Pharmacy, Université Libre de Bruxelles (ULB), Campus Plaine CP205/6, Boulevard du Triomphe, B-1050 Brussels, Belgium

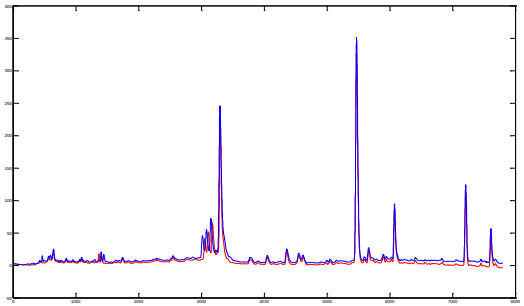
Vrije Universiteit Brussel

Introduction

- **Chromatographic fingerprints**



Introduction



- Chromatographic fingerprints
- Applications
- Similarity analysis
- Distance- and correlation based parameters [a]
  - Euclidean distance

$$de(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{j=1}^{q} (x_{1j} - x_{2j})^2}$$

  - Standardized Euclidean distance

$$ds(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{j=1}^{q} [(x_{1j} - x_{2j})/s_j]^2} \quad \text{with} \quad s_j = \sqrt{\frac{1}{p}\sum_{i=1}^{p} (x_{ij} - \bar{x}_j)^2}$$

  - Mahalanobis distance

$$dm(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{(\mathbf{x}_1 - \mathbf{x}_2)^T \mathbf{C}^{-1} (\mathbf{x}_1 - \mathbf{x}_2)}$$

  - Adapted Similarity score

$$ss*(\mathbf{x}_1, \mathbf{x}_2) = 1 - \frac{\sqrt{\sum_{j=1}^{q} (x_{1j} - x_{2j})^2}}{\max\left(\sum_{j=1}^{q} x_{1j}, \sum_{j=1}^{q} x_{2j}\right)}$$

  - congruence coefficient

$$c(\mathbf{x}_1, \mathbf{x}_2) = \frac{\sum_{j=1}^{q} x_{1j} x_{2j}}{\sqrt{\sum_{j=1}^{q} x_{1j}^2 \sum_{j=1}^{q} x_{2j}^2}} =$$

  - correlation coefficient

$$r(\mathbf{x}_1, \mathbf{x}_2) = \frac{\mathrm{cov}(\mathbf{x}_1 \mathbf{x}_2)}{s_{x1} s_{x2}}$$

[a] Alaerts et al., Journal of Chromatography B, 910 (2012) 61– 70

FABI

Case study

- **3 Curcuma types in Pharmacopoeia of the People's Republic of China:**
  - **32 Rhizoma Curcumae longae (Jianghuang) samples**
    - Species: C. longa
    - Plant part: dried rhizome
  - **29 Rhizoma Curcumae (Ezhu) samples**
    - Species: C. phaeocaulis/C. kwangsiensis/C. wenyujin
    - Plant part: dried rhizome
  - **32 Radix Curcumae (Yujin) samples**
    - Species: C. wenyujin/C. longa/C.kwansiensis/C.phaeocaulis
    - Plant part: dried root tuber

Case study

- **3 Curcuma types in Pharmacopoeia of the People's Republic of China:**
  - **32 Rhizoma Curcumae longae (Jianghuang) samples**
    - Species: C. longa
    - Plant part: dried rhizome
  - **29 Rhizoma Curcumae (Ezhu) samples**
    - Species: C. phaeocaulis/C. kwangsiensis/C. wenyujin
    - Plant part: dried rhizome
  - **32 Radix Curcumae (Yujin) samples**
    - Species: C. wenyujin/C. longa/C.kwansiensis/C.phaeocaulis
    - Plant part: dried root tuber

3 ≠ Traditional Chinese Medicines with **≠ indications**

Aim

Similarity parameters (Euclidean distance) applicable to distinguish the 3 types of Curcuma samples?

Registration of fingerprint data

Preparation of extracts

- 1.5 g powdered and sieved
- 20.0 mL extraction solvent: ethanol/water (80/20 V/V)
- 30 min ultrasonic extraction
- Filter extract + adjust volume to 25.0 mL (extraction solvent)

Registration of fingerprint data

High Pressure Liquid Chromatography

- LaChrom Elite VWR Hitachi (L-2130 pump, L-2200 autosampler, L-2350 column oven, L-2455 diode array detector)
- 4 coupled Chromolith Performance RP-18e C18 columns (100 x 4.6 mm) + guard column at 35°C
- Injection volume: 10 µL
- Mobile phases: water (A) and acetonitrile (B), both + 0.05% trifluoroacetic acid
- Gradient: 0 min: 5% B; 5 min: 33.5%B; 36 min: 60%B; 50-60 min:
- Flow rate: 1mL/min
- Detection wave length: 214 nm
- Each sample analyzed in duplicate
- Daily blank injection in duplicate

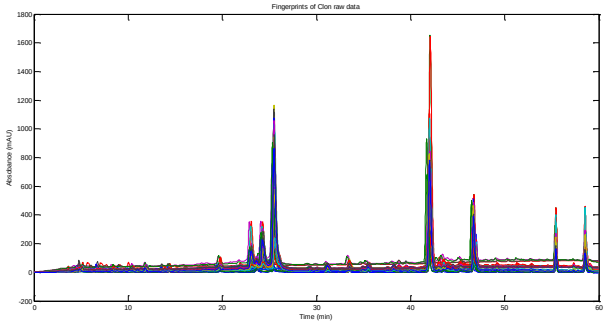
Fingerprint data analysis

Preprocessing

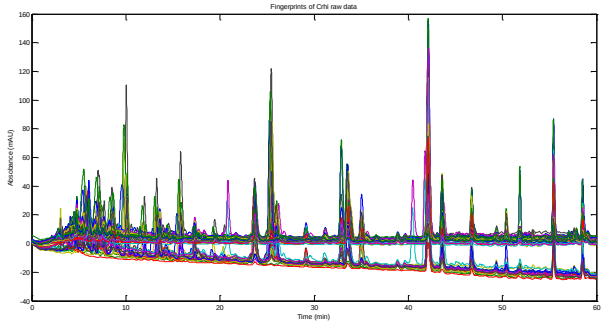
- **Blank correction of the fingerprints**
- **Correlation optimized warping (COW) per Curcuma type**
- **Alignment of the 3 types (COW)**

Fingerprint data analysis

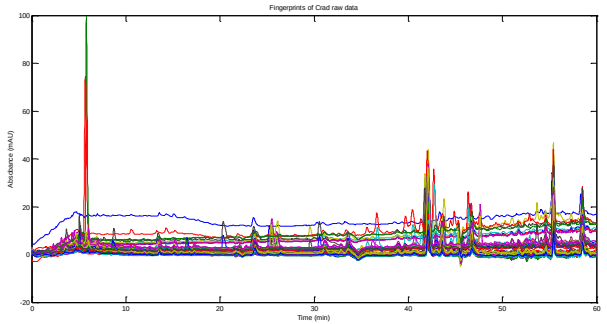
BLANK-CORRECTED DATA



C. longa



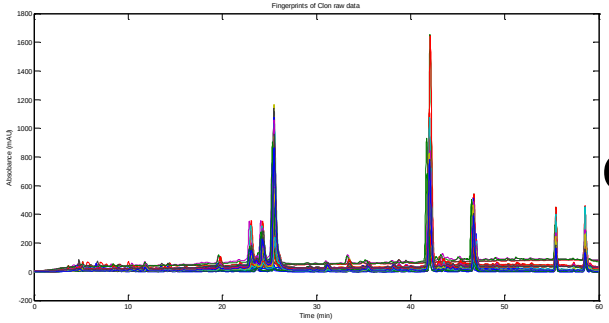
C. rhizoma



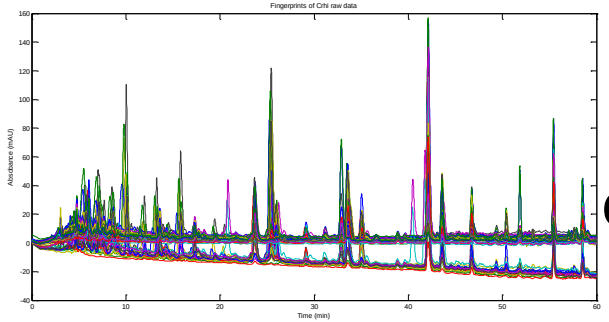
C. radix

Fingerprint data analysis

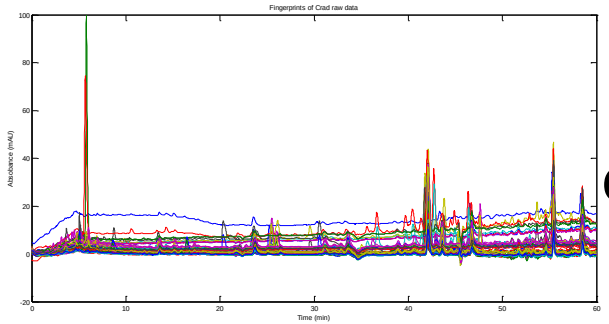
BLANK-CORRECTED DATA



C. longa

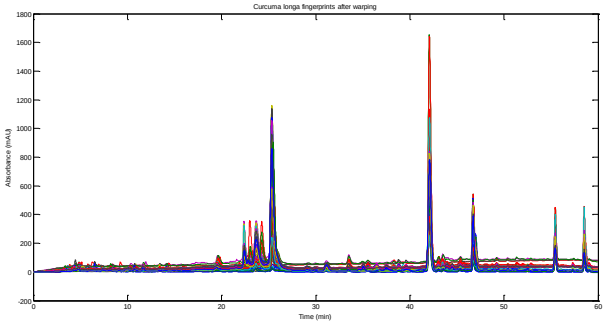


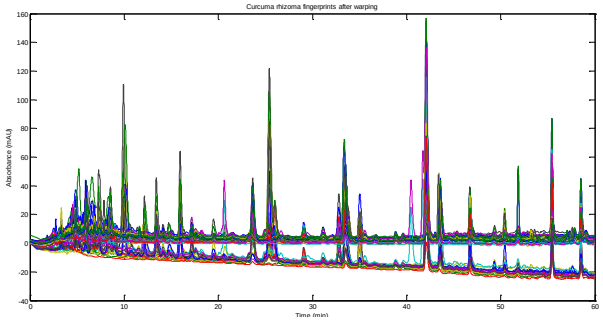
C. rhizoma

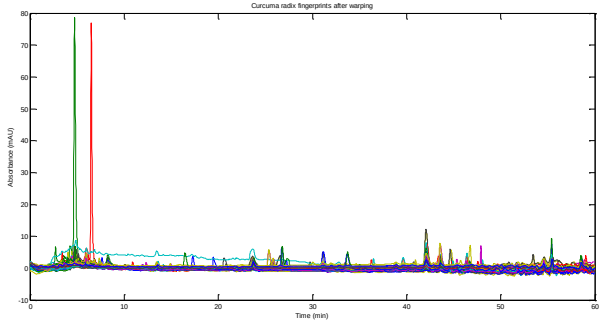


C. radix

DATA AFTER COW







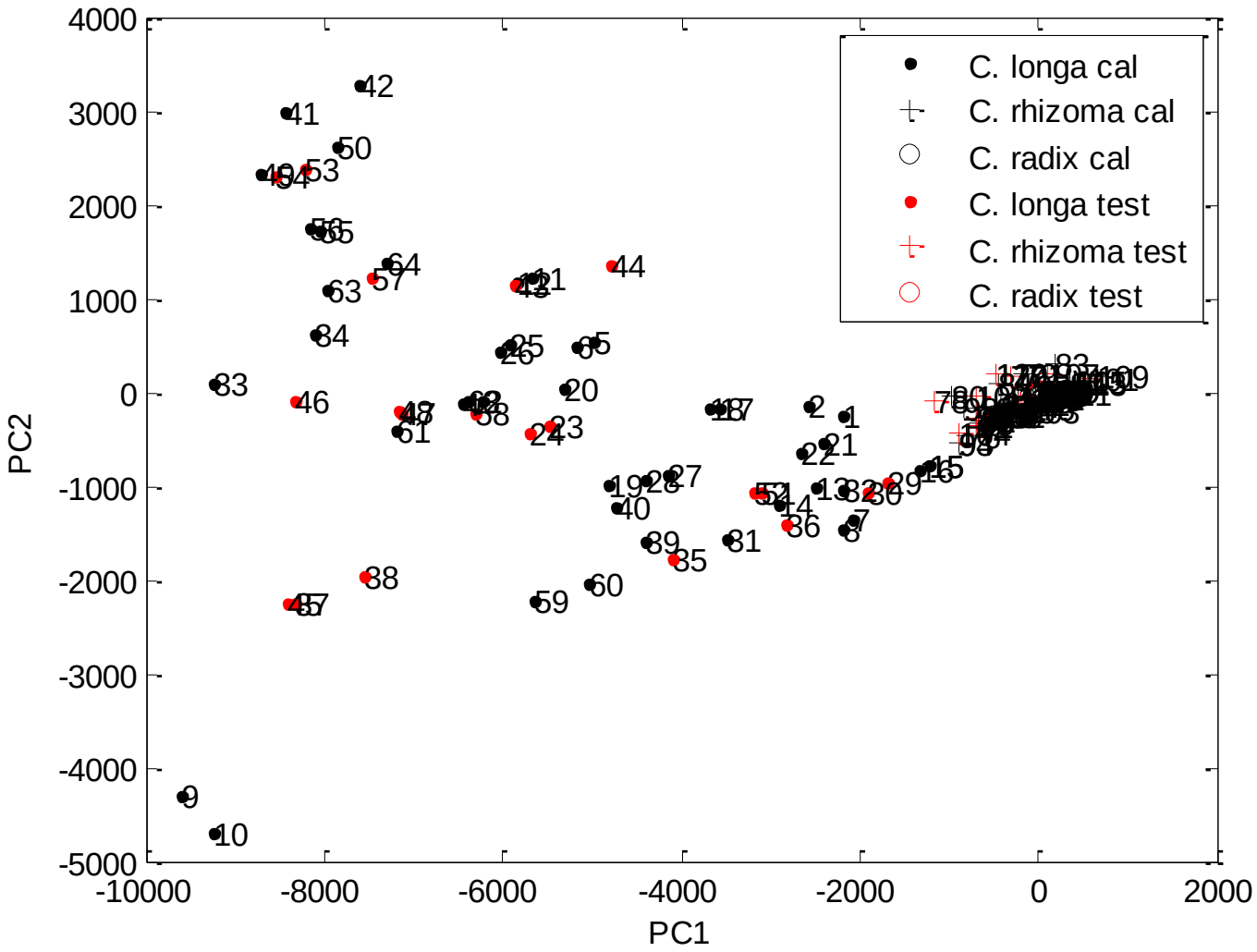
Fingerprint data analysis

Division in calibration and test set

- Per sample: calculation of average (n=2) fingerprint
- Division samples in calibration and test set: duplex algorithm
- Check distribution calibration & test set samples: Principal Component Analysis

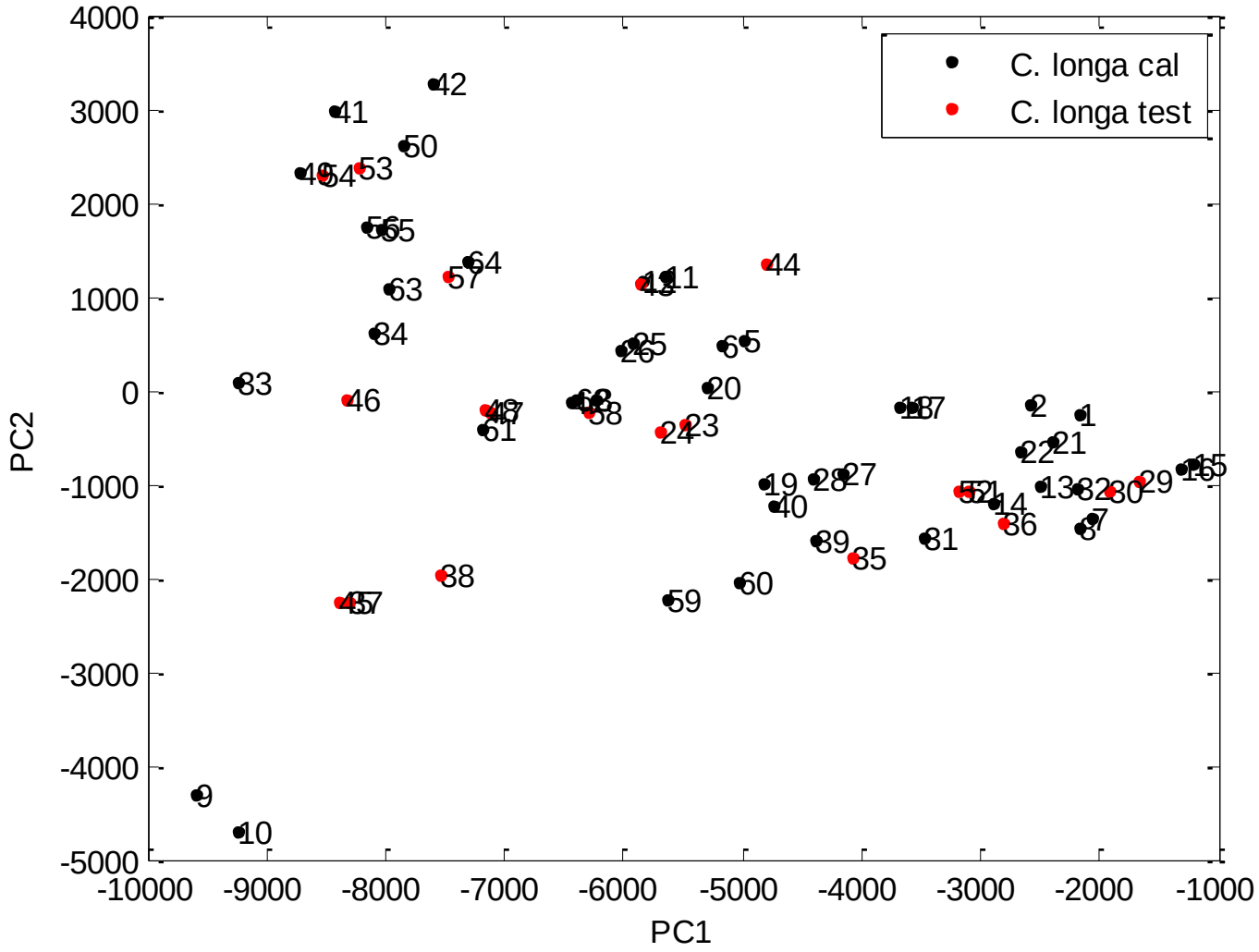
Fingerprint data analysis

DISTRIBUTION CALIBRATION & TEST SET FINGERPRINTS IN PC1-PC2 SPACE



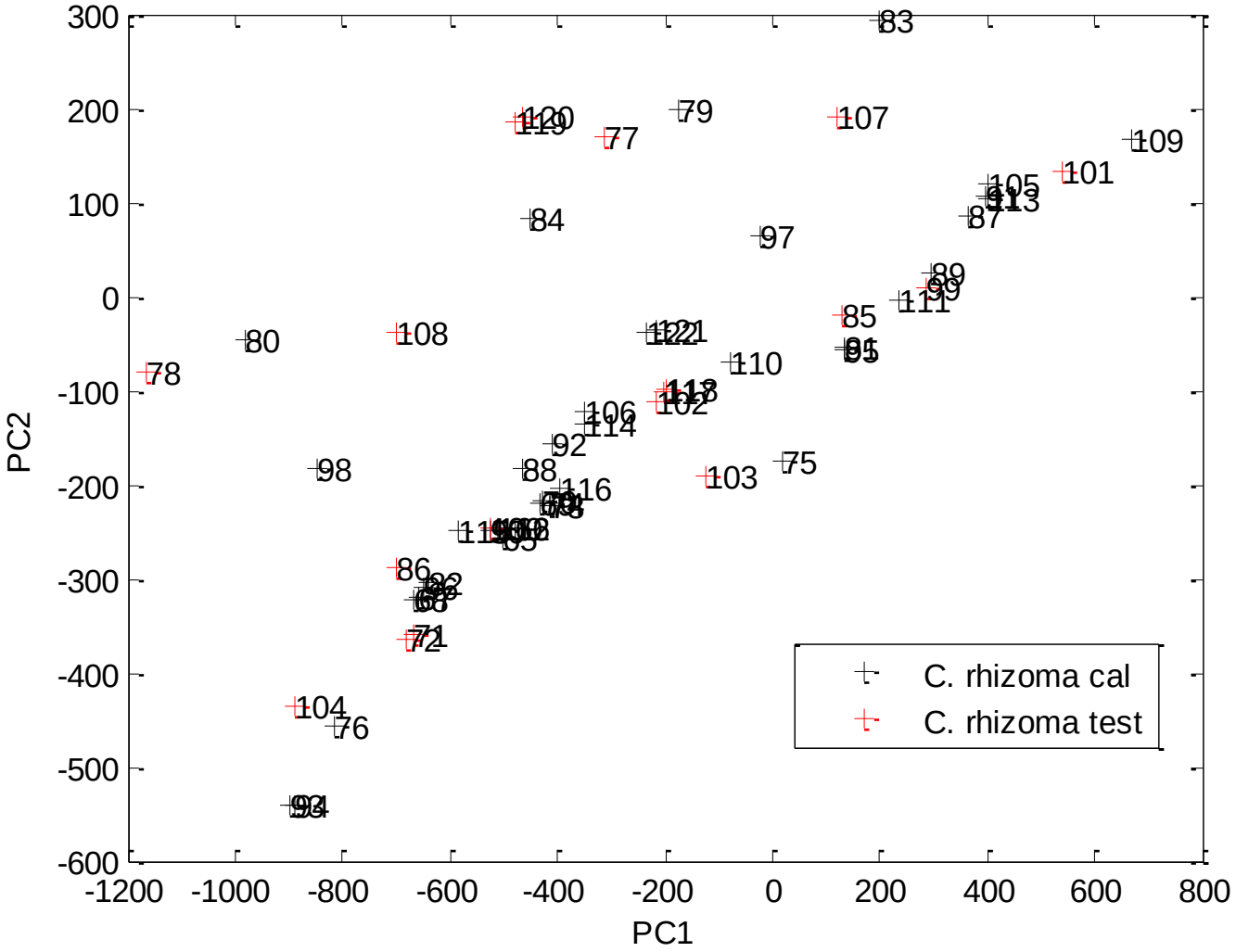
Fingerprint data analysis

DISTRIBUTION CALIBRATION & TEST SET FINGERPRINTS IN PC1-PC2 SPACE



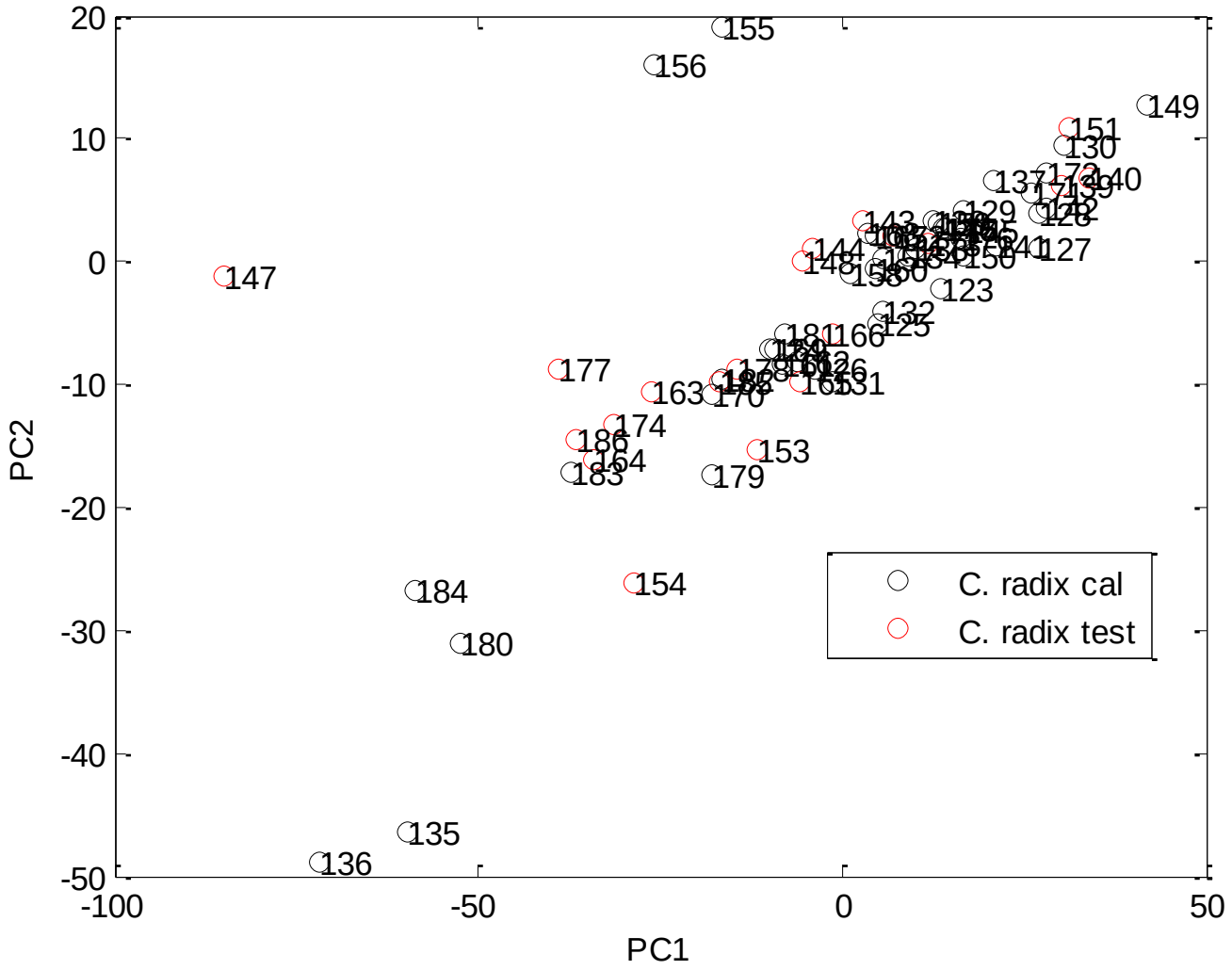
Fingerprint data analysis

DISTRIBUTION CALIBRATION & TEST SET FINGERPRINTS IN PC1-PC2 SPACE



Fingerprint data analysis

DISTRIBUTION CALIBRATION & TEST SET FINGERPRINTS IN PC1-PC2 SPACE

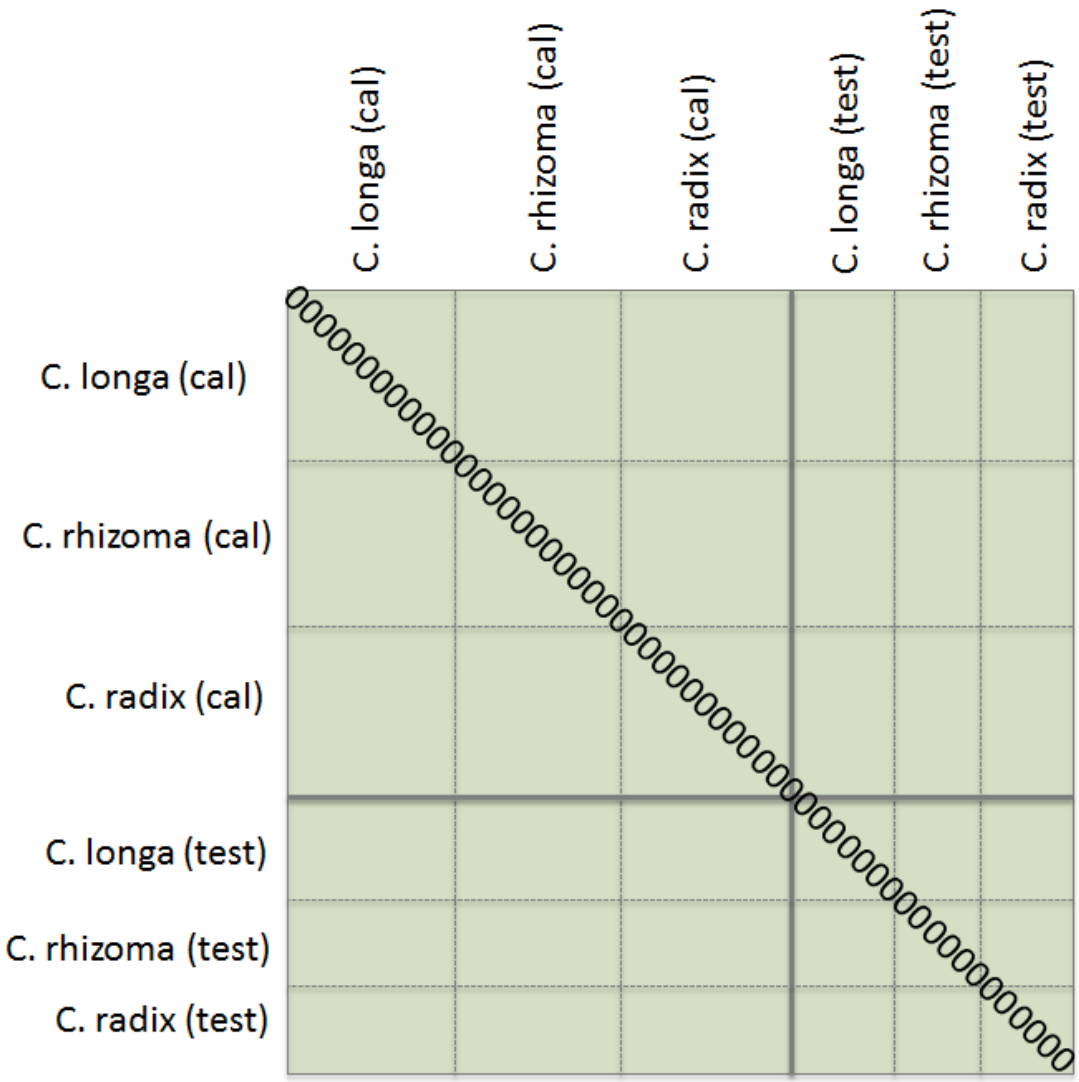


Fingerprint data analysis

Similarity analysis

- **Calculation of Euclidean distance (d) between each pair of fingerprints**

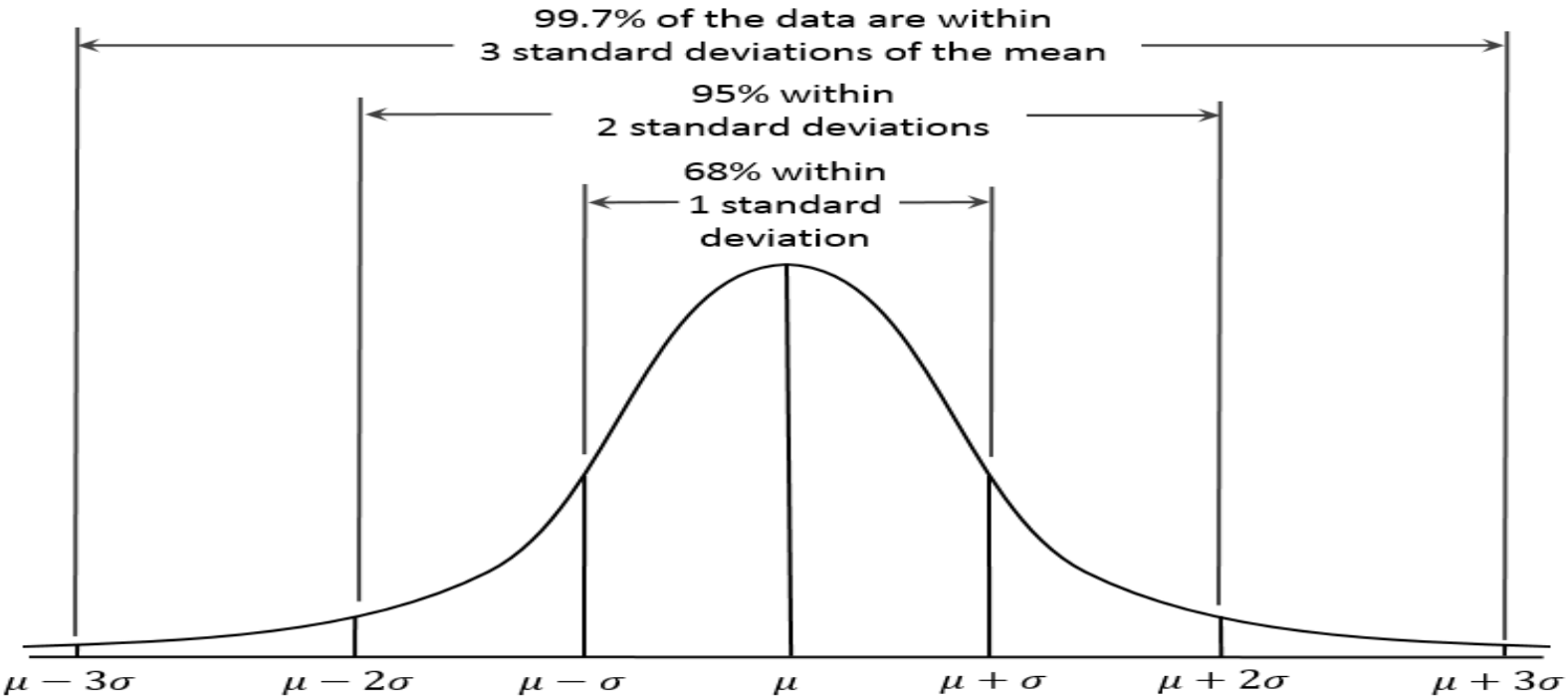
Fingerprint data analysis



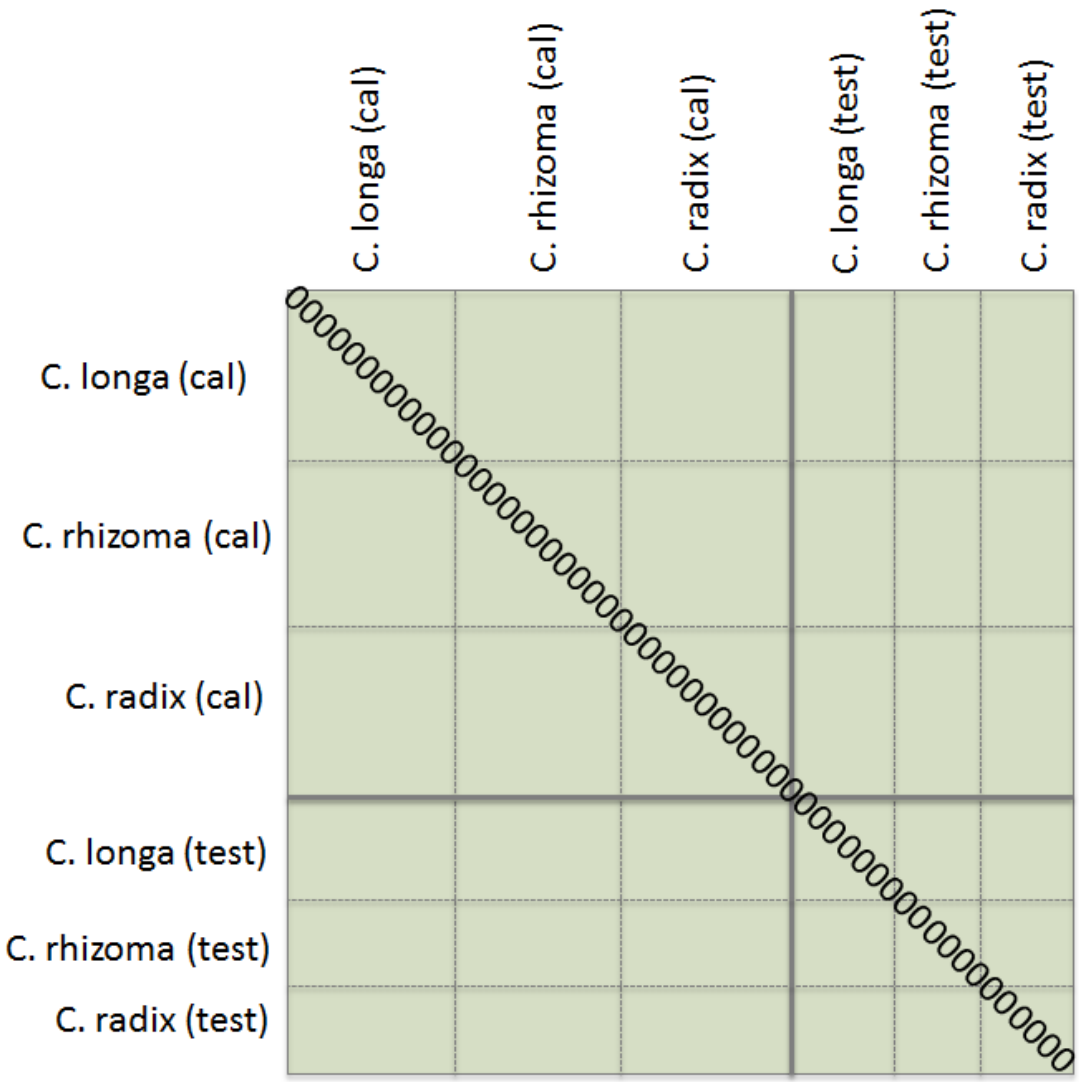
Fingerprint data analysis

Similarity analysis

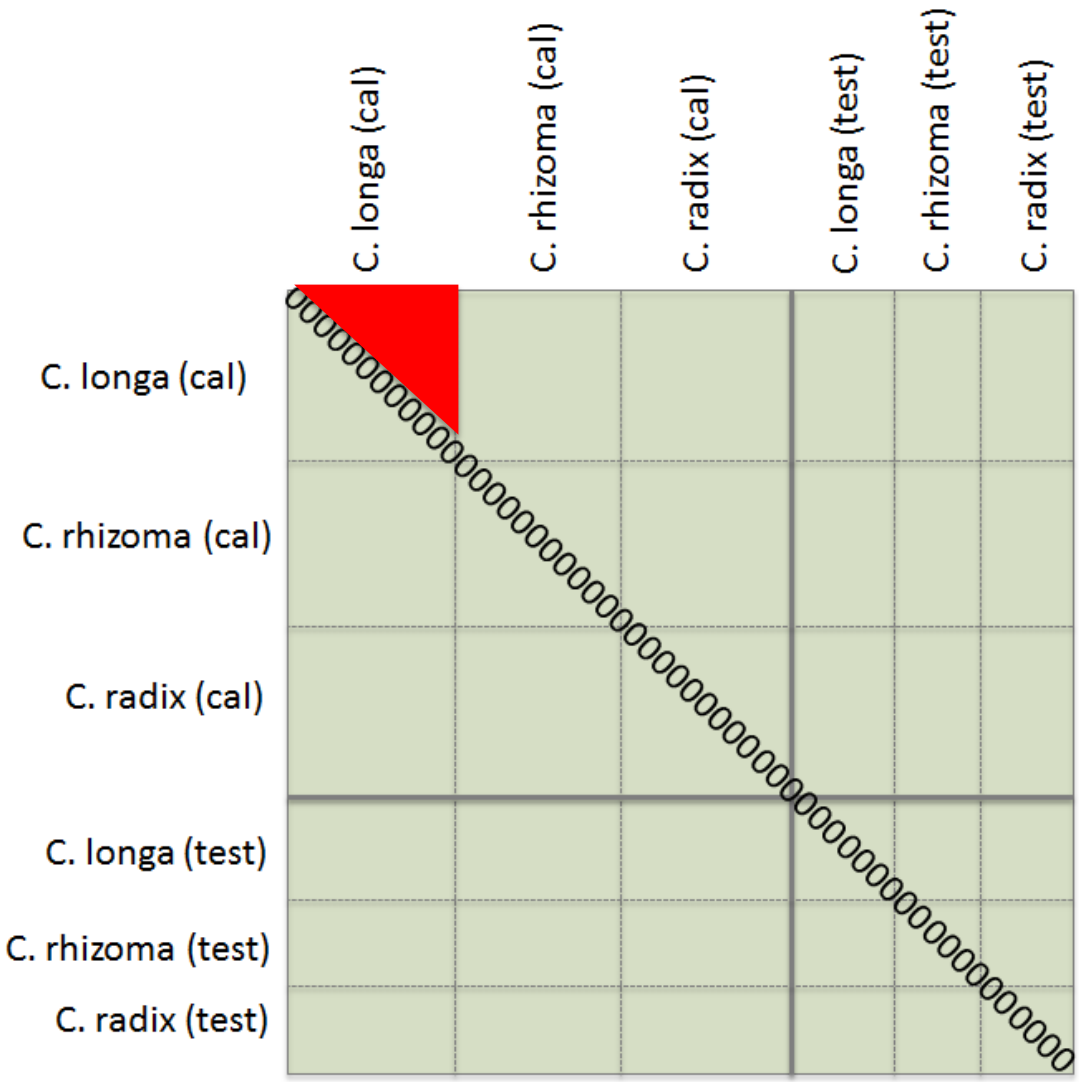
- **Upper Warning and Control limits (UWL & UCL [a]) for d per C. type**
  - UWL = $\bar{d}+\mathbf{1.96}\ \boldsymbol{x}\ s_d$ → 97.5% of d values per C. type < $UWL_{C.type}$
  - UCL = $\bar{d}+\mathbf{3.09}\ \boldsymbol{x}\ s_d$ → 99.85% of d values per C. type < $UCL_{C.type}$



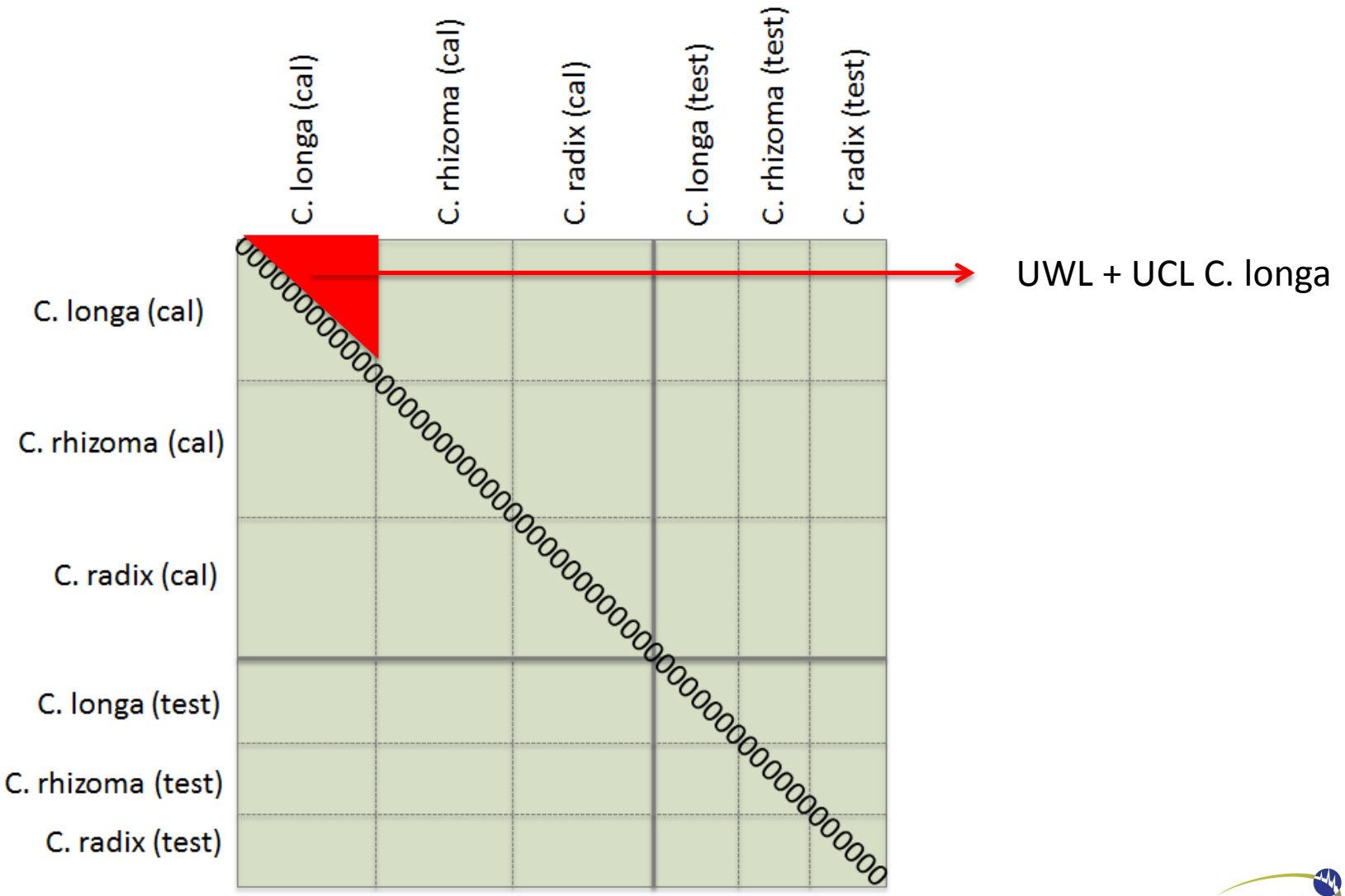
Fingerprint data analysis



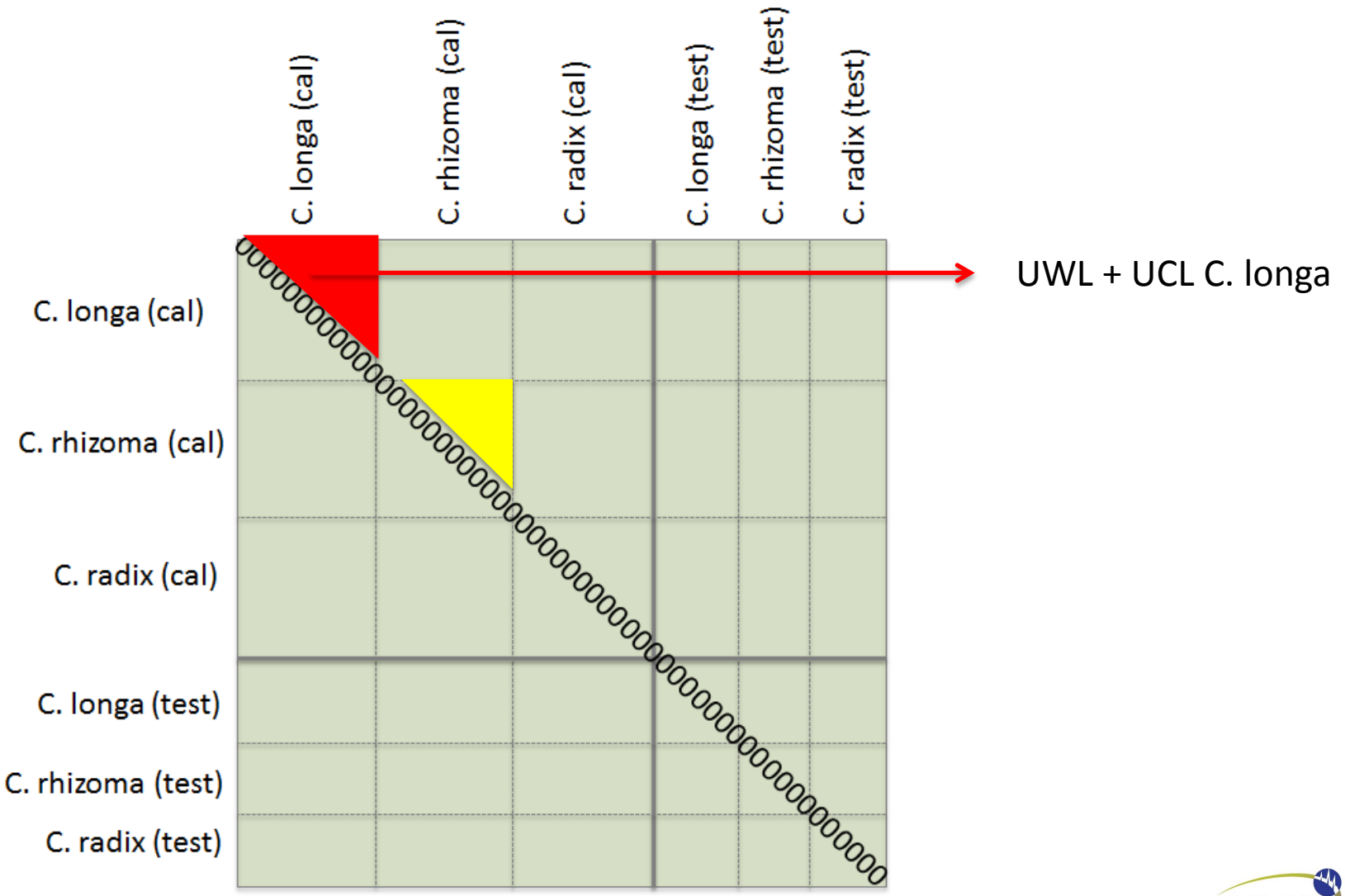
Fingerprint data analysis



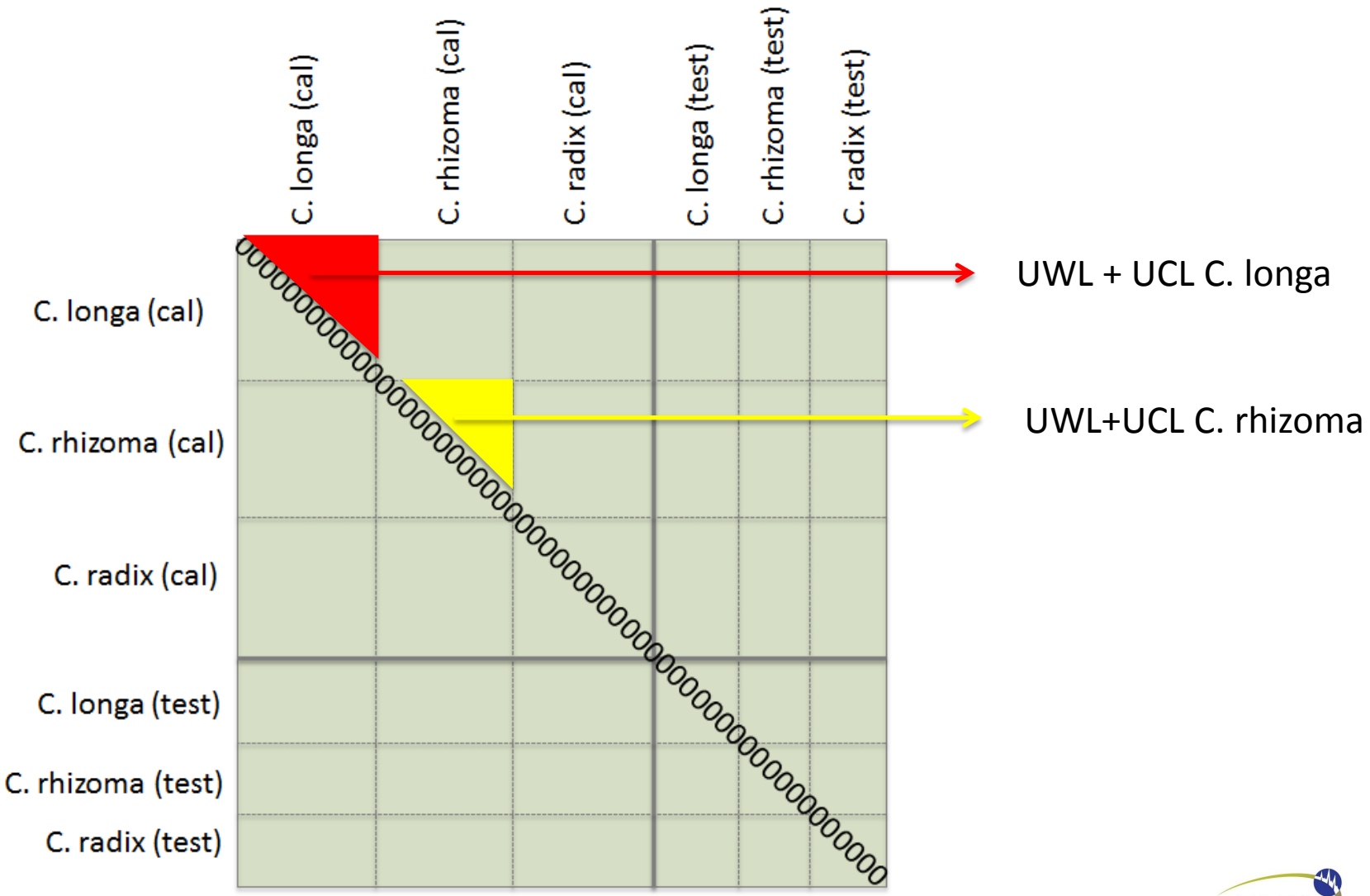
Fingerprint data analysis



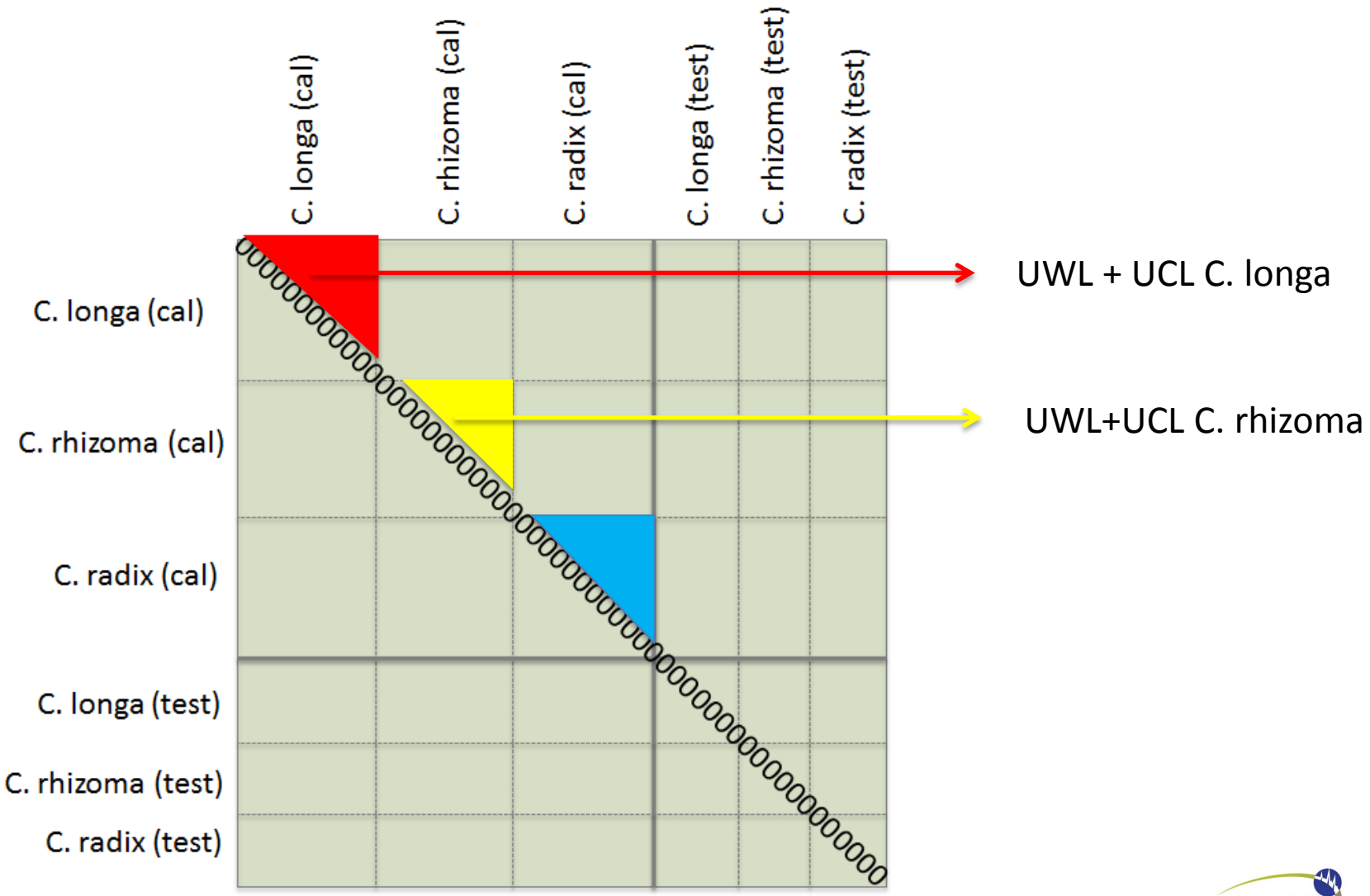
Fingerprint data analysis



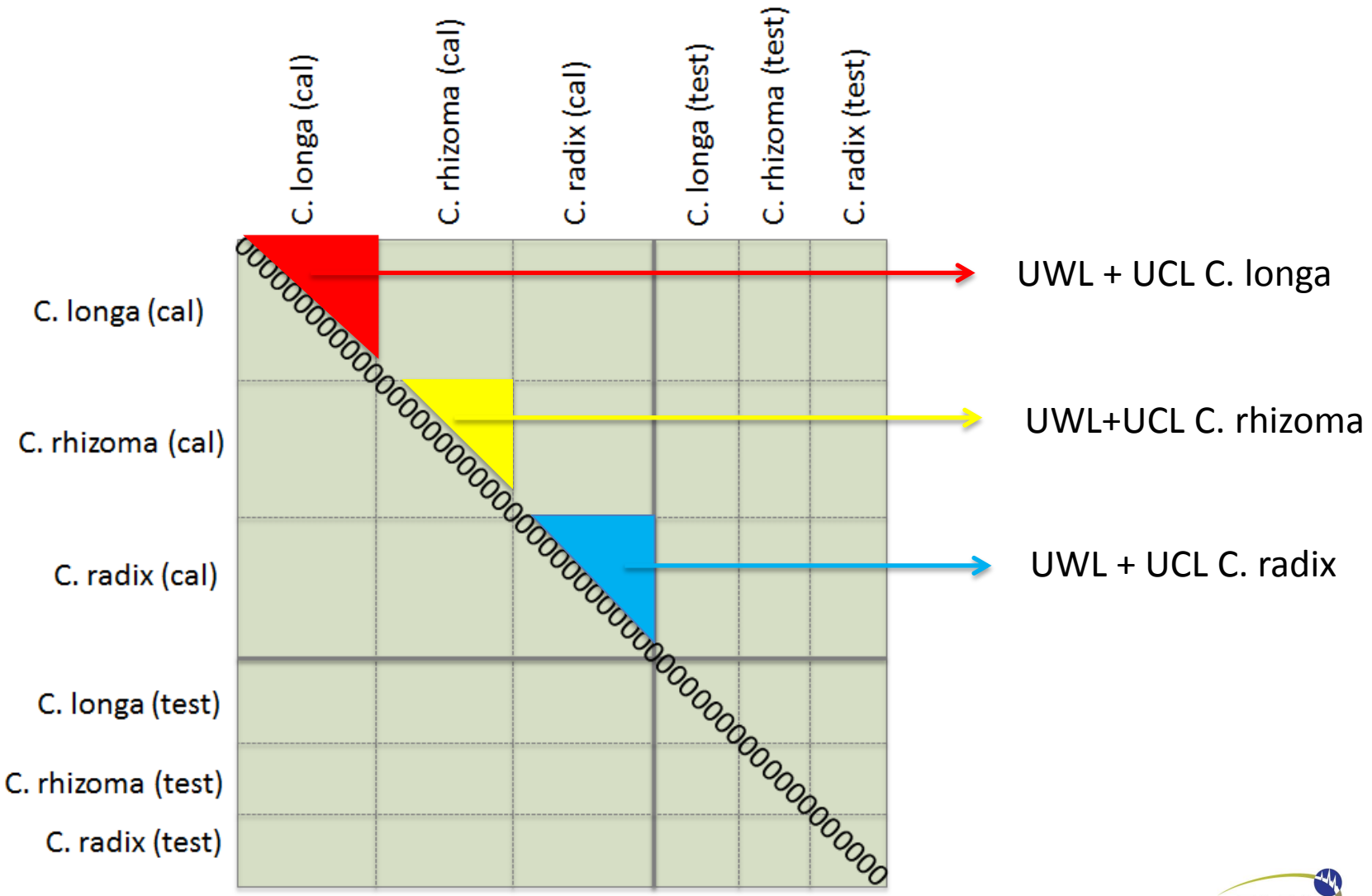
Fingerprint data analysis



Fingerprint data analysis



Fingerprint data analysis



Fingerprint data analysis

Similarity analysis

- **How to use these limits?**

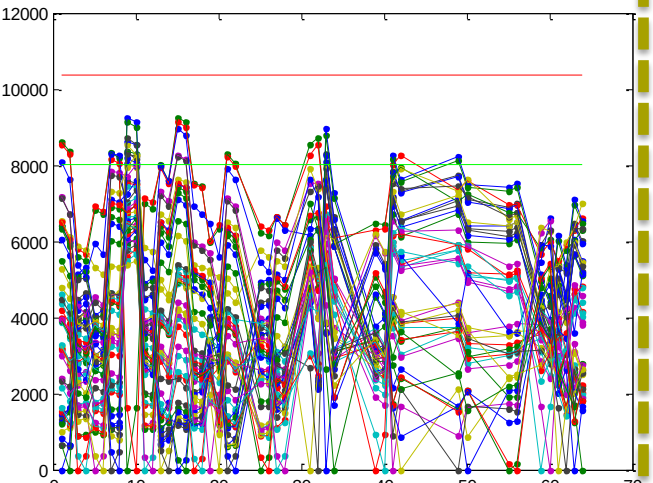
Fingerprint data analysis

Similarity analysis

- **How to use these limits?**

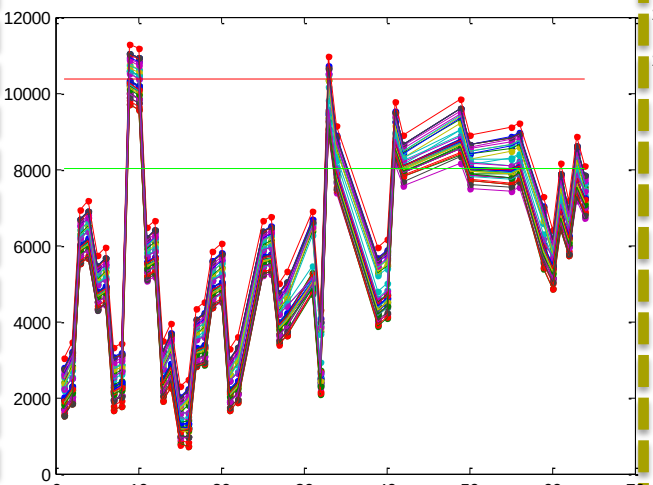
Green horizontals: $UWL_{C.longa}$
Red horizontals: $UCL_{C.longa}$

d from C. longa calibration samples



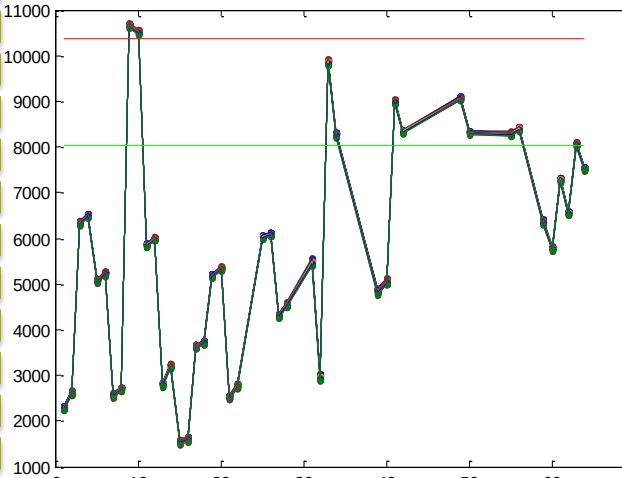
to C. longa calibration samples

d from C. rhizoma calibration samples



to C. longa calibration samples

d from C. radix calibration samples



to C. longa calibration samples

Global behavior relative to $UWL_{C.\,longa}$ and $UCL_{C.\,longa}$ differs per C. type!!!

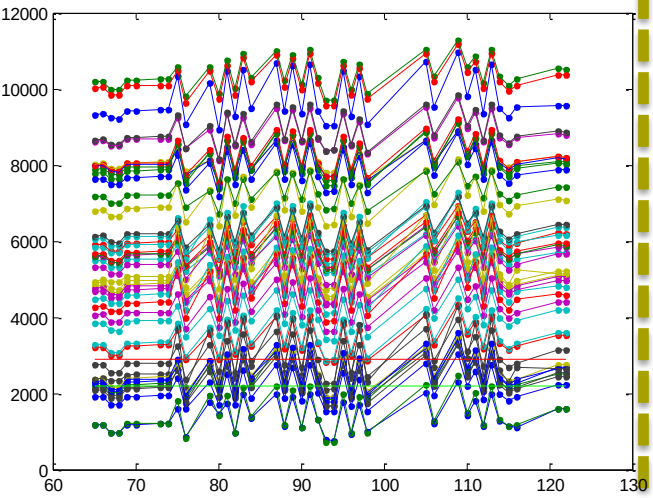
Fingerprint data analysis

Similarity analysis

- **How to use these limits?**

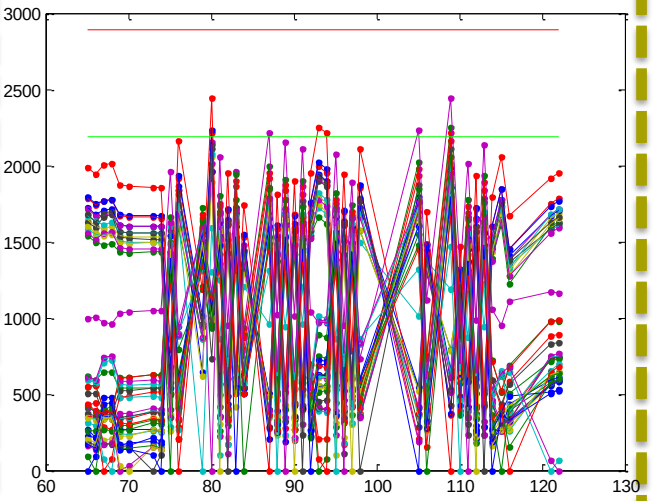
Green horizontals: $\mathrm{UWL}_{\mathrm{C.rhizoma}}$
Red horizontals: $\mathrm{UCL}_{\mathrm{C.rhizoma}}$

d from C. longa calibration samples



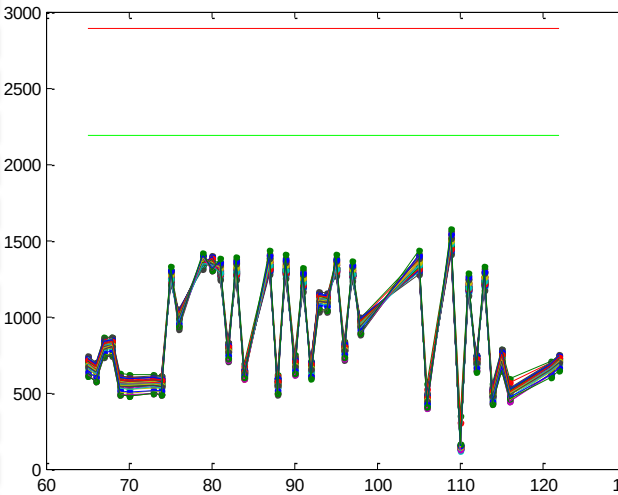
to C. rhizoma calibration samples

d from C. rhizoma calibration samples



to C. rhizoma calibration samples

d from C. radix calibration samples



to C. rhizoma calibration samples

Global behavior relative to $UWL_{C.\,rhizoma}$ and $UCL_{C.\,rhizoma}$ differs per C. type!!!

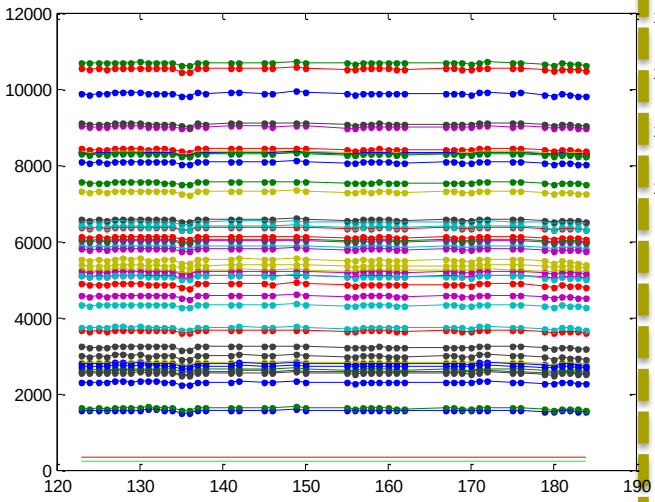
Fingerprint data analysis

Similarity analysis

- **How to use these limits?**

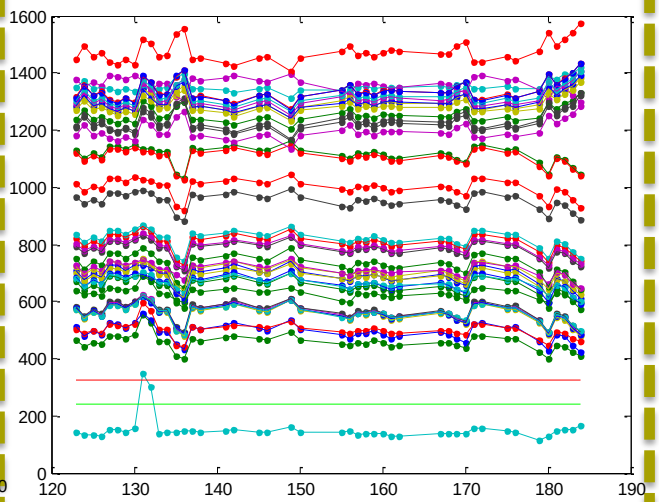
Green horizontals: $\mathrm{UWL}_{\mathrm{C.radix}}$
Red horizontals: $\mathrm{UCL}_{\mathrm{C.radix}}$

d from C. longa calibration samples



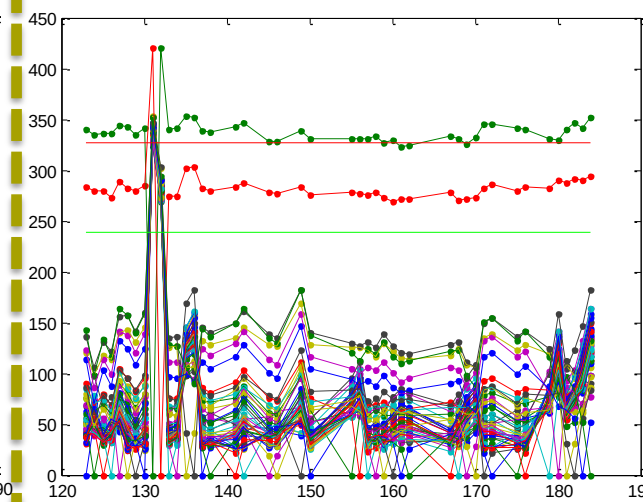
to C. radix calibration samples

d from C. rhizoma calibration samples



to C. radix calibration samples

d from C. radix calibration samples



to C. radix calibration samples

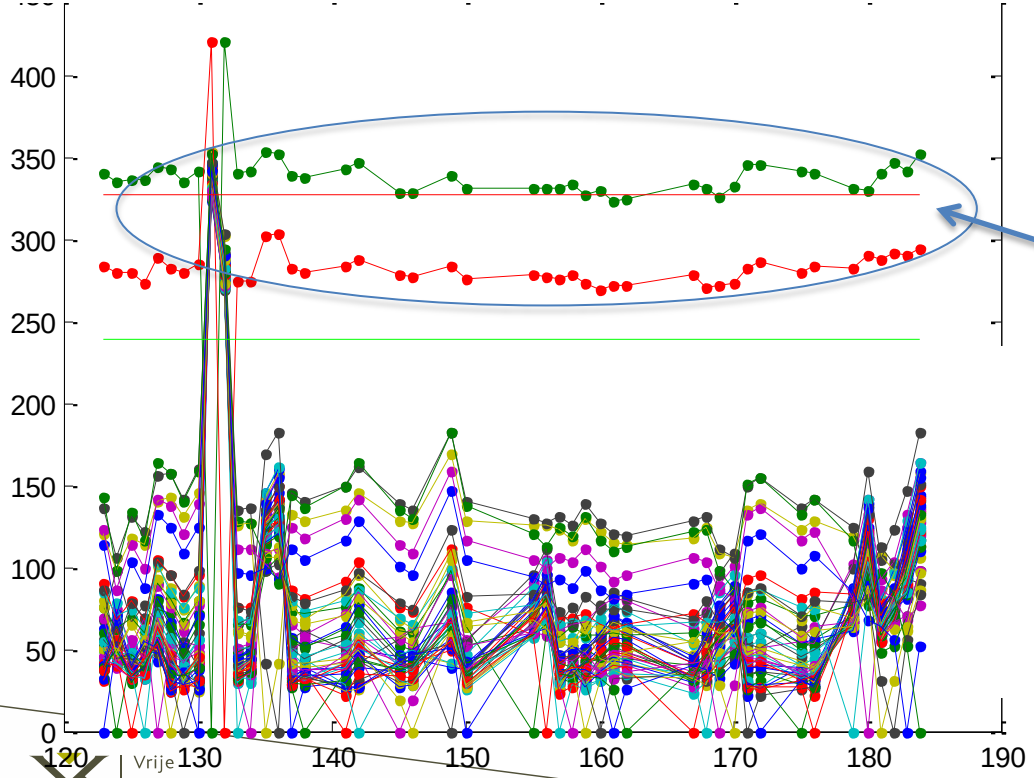
Global behavior relative to $UWL_{C.\,radix}$ and $UCL_{C.\,radix}$ differs per C. type!!!

Fingerprint data analysis

Similarity analysis

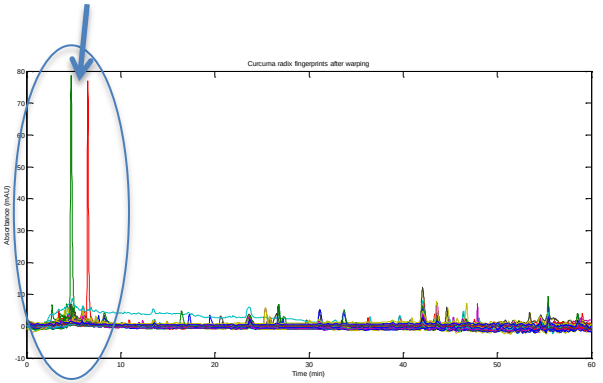
- **Identification of outliers**

d of C. radix calibration samples to C. radix calibration samples



Green horizontals: $\mathrm{WL}_{\mathrm{C.\,radix}}$
Red horizontals: $\mathrm{CL}_{\mathrm{C.}\,radix}$

OUTLIERS!



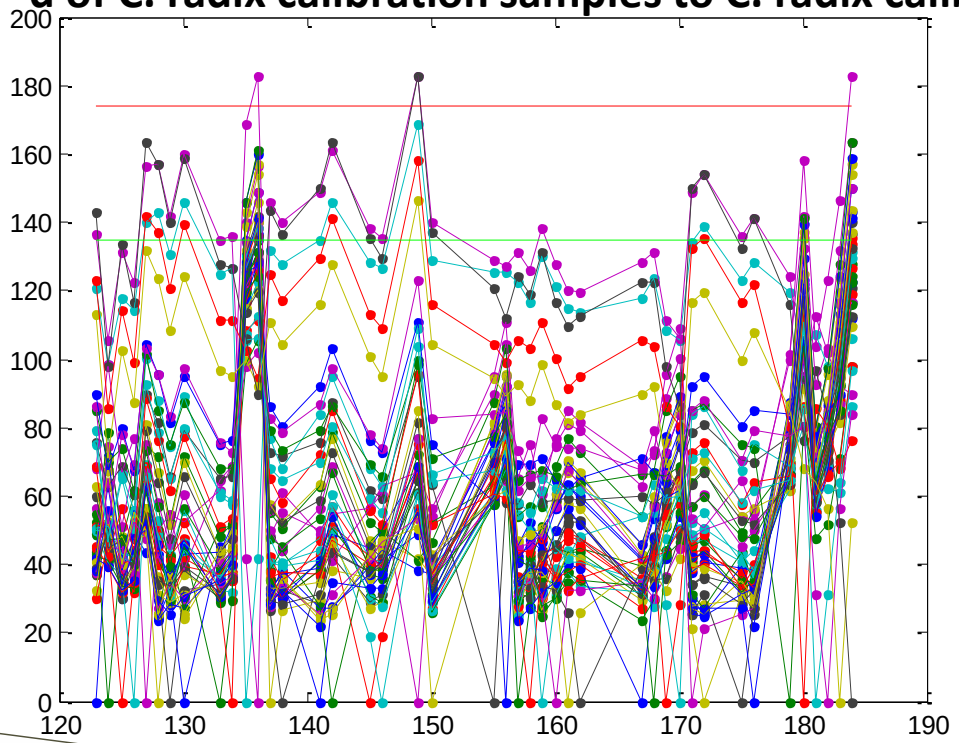
Vrije Universiteit Brussel

FABI

Fingerprint data analysis

Similarity analysis

- **Identification of outliers**



Green horizontals: $\mathrm{WL}_{\mathrm{C.\,radix}}$
Red horizontals: $\mathrm{CL}_{\mathrm{C.}\,radix}$

Fingerprint data analysis

Similarity analysis

- **Deriving discrimination/classification rules**

Sample	Known Class	% UCL C. longa exceedings	% UWL C. longa exceedings	% UCL C. rhizoma exceedings	% UWL C. rhizoma exceedings	% UCL C. radix exceedings	% UWL C. radix exceedings
1	C. longa	0	7	3	40	100	100
2	C. longa	0	5	25	65	100	100
3	C. longa	0	0	100	100	100	100
4	C. longa	0	0	100	100	100	100
5	C. longa	0	0	100	100	100	100
6	C. longa	0	0	100	100	100	100
7	C. longa	0	7	18	50	100	100
8	C. longa	0	7	20	80	100	100

Fingerprint data analysis

Similarity analysis

- **Deriving discrimination/classification rules**
  - To distinguish C. samples only UCL values necessary
  - C. longa sample if:
    - d to ≤ 1% of C. longa calibration samples exceeds $UCL_{C.longa}$
    - d to > 1% of C. rhizoma calibration samples exceeds $UCL_{C.rhizoma}$
    - d to > 1% of C. radix calibration samples exceeds $UCL_{C.radix}$
  - C. rhizoma sample if:
    - d to ≤ 1% of C. rhizoma calibration samples exceeds $UCL_{C.rhizoma}$
    - d to > 1% of C. radix calibration samples exceeds $UCL_{C.radix}$
  - C. radix sample if:
    - d to > 1% of C. longa calibration samples exceeds $UCL_{C.longa}$
    - d to ≤ 1% of C. rhizoma calibration samples exceeds $UCL_{C.rhizoma}$
    - d to ≤ 1% of C. radix calibration samples exceeds $UCL_{C.radix}$
  - Samples not compliant to these rules: outliers

Fingerprint data analysis

Application of Euclidean distance based classification/ discrimination rules and comparison with classical discrimination/classification methods

- Linear and Quadratic Discriminant Analysis
- Soft Independent Modelling by Class Analogy (nearest class)

DISCRIMINATION/ CLASSIFICATION TECHNIQUE	CCR CALIBRATION SET	CCR TEST SET	DISCRIMINATION/ CLASSIFICATION TECHNIQUE	NUMBER OF RETAINED LV/PC
EUCLIDEAN DISTANCE	0.9516	0.9667	LDA	16
LDA	0.9921	0.8167	QDA	1
QDA	0.9841	0.9833	SIMCA	1 (C. longa), 2 (C. rhizoma), 3 (C. radix)
SIMCA	0.3492	0.3333		

Euclidean distance based classification ≈ best discriminant analysis technique (QDA)

Conclusions

- **Euclidean distance was succesfully used to eliminate outliers in the Radix Curcumae fingerprints**
- **Rizoma Curcumae longae, Rhizoma Curcumae and Radix Curcumae samples can be distinguished based on Euclidean distances**
- **Euclidean distance based discrimination/classification performs almost equal to best discrimination technique**

Acknowledgements

- To the co-authors for their contributions!
- To you for your kind attention