UNIVERSITY OF CHEMISTRY AND TECHNOLOGY, PRAGUE
Faculty of Food and Biochemical Technology
Department of Food Analysis and Nutrition

Metabolomic fingerprinting for food authentication
The Whys and Hows of metabolomics

A general introduction

MultiCoop

Jana Hajslova, Vít Kosek

Training school

Prague, November 27, 2018

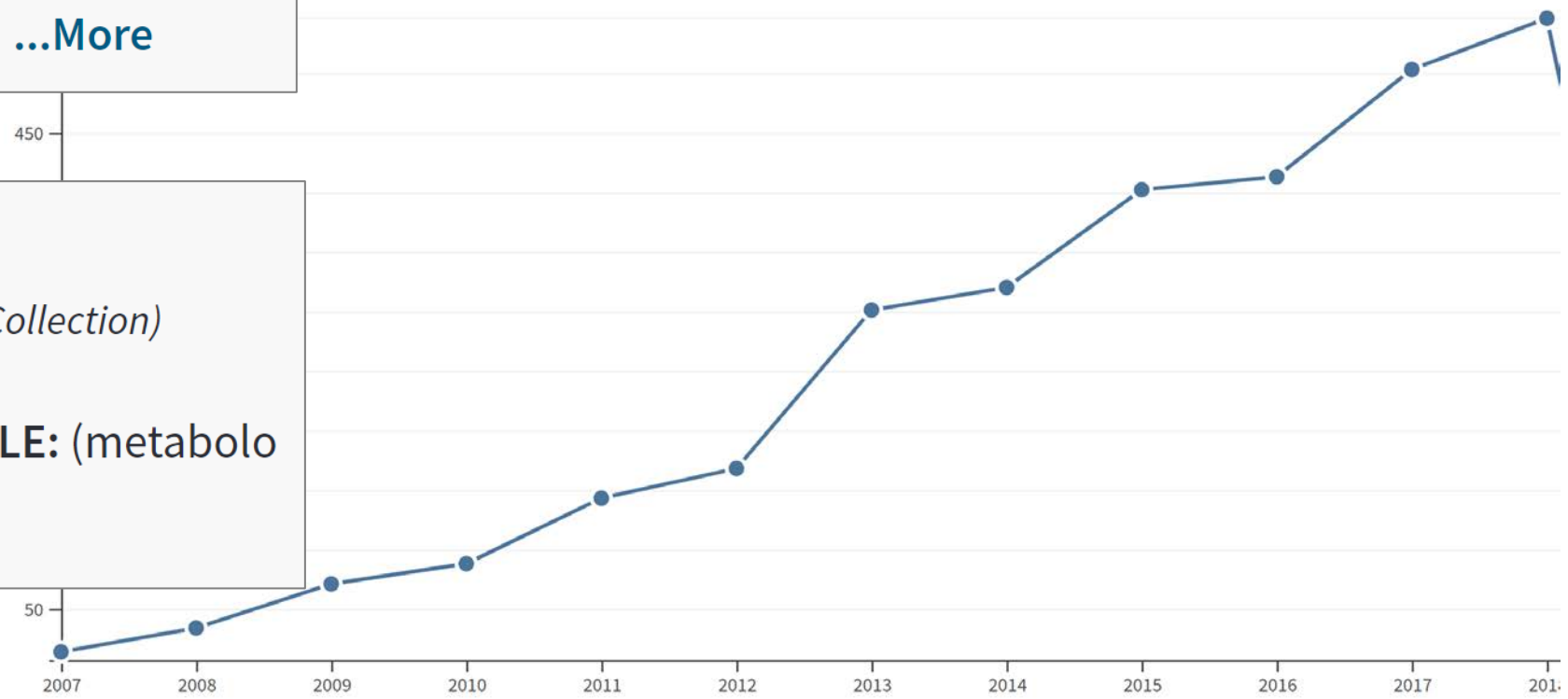
Growing interest in metabolomics….

Results: 194
(from Web of Science Core Collection)

You searched for: TITLE: (metabolomic) *AND* **TOPIC:** (food) **…More**

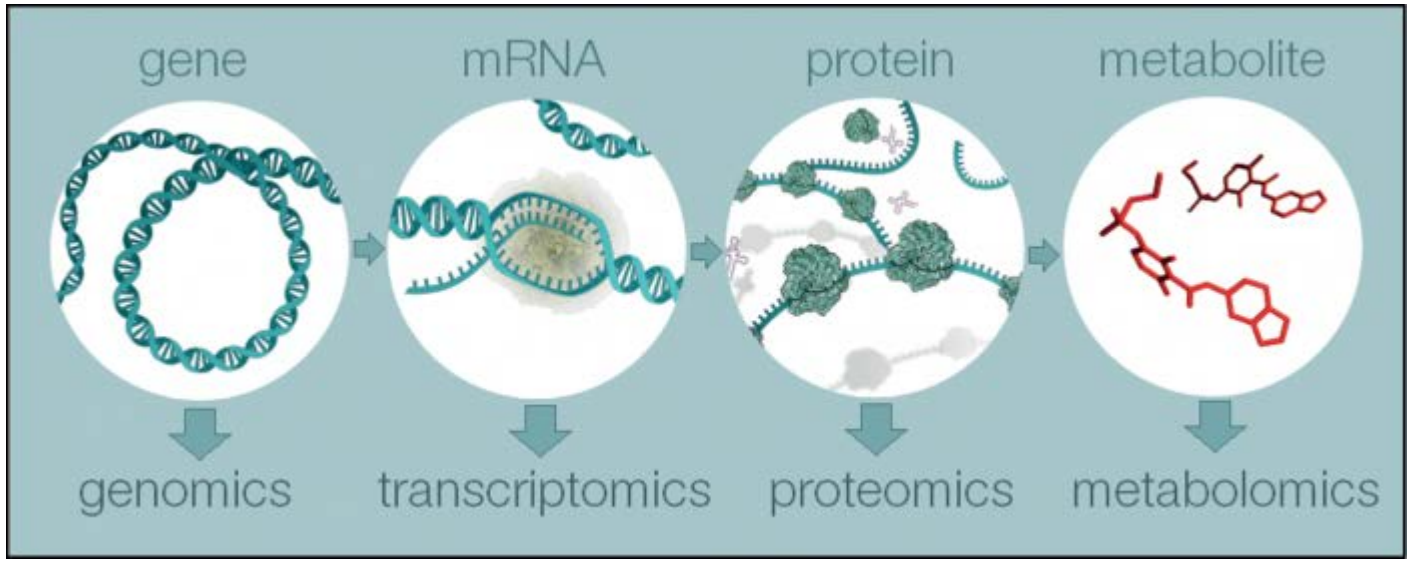
Results: 4,294
(from Web of Science Core Collection)

You searched for: TITLE: (metabolomic) **…More**



Central dogma of systems biology

- **-omics:** Unbiased **large-scale analysis** of molecules present in the biological system



https://www.ebi.ac.uk/training/online/sites/ebi.ac.uk.training.online/files/user/2760/images/Metabolomics/central_dogma_figure_1_.png

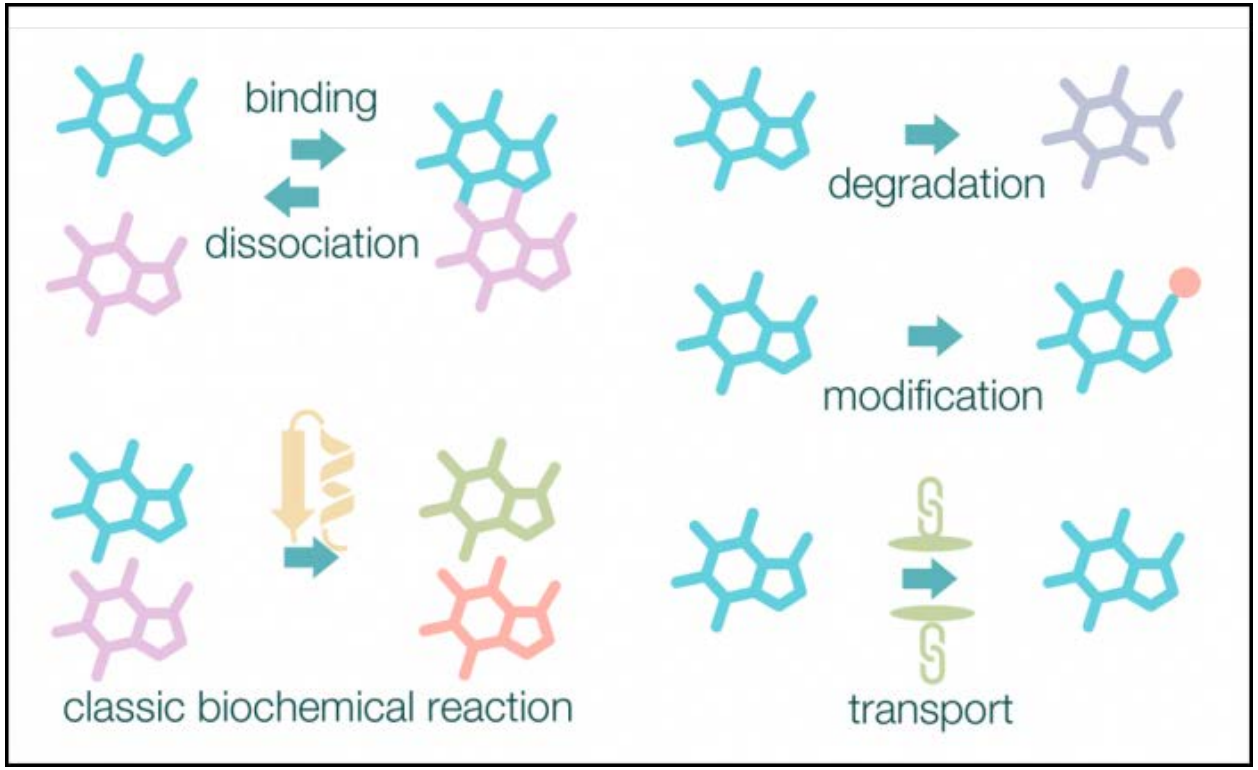
- **Metabolomics:** directly related to biological activity *i.e.* **phenotype** and **biological state**

Metabolic reactions

- Set of reactions usually catalysed by enzymes
- More dynamic than other –omes

Metabolomics fingerprint

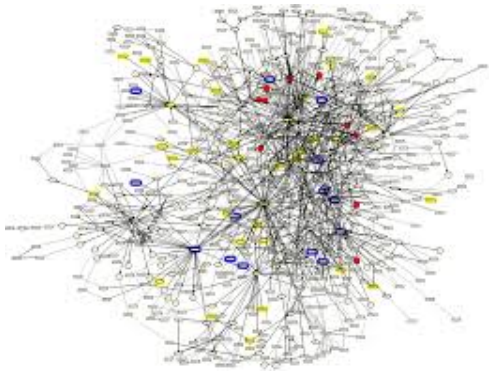
- A „snapshot“ of the metabolome can be taken



https://www.ebi.ac.uk/training/online/sites/ebi.ac.uk.training.online/files/user/2760/images/Metabolomics/metabolic_reactions.png

Metabolome

- A set of small molecules inside an organism (matrix)



METABOLOME IS INHERENTLY VERY DYNAMIC

interaction both within and between biological systems, and with the

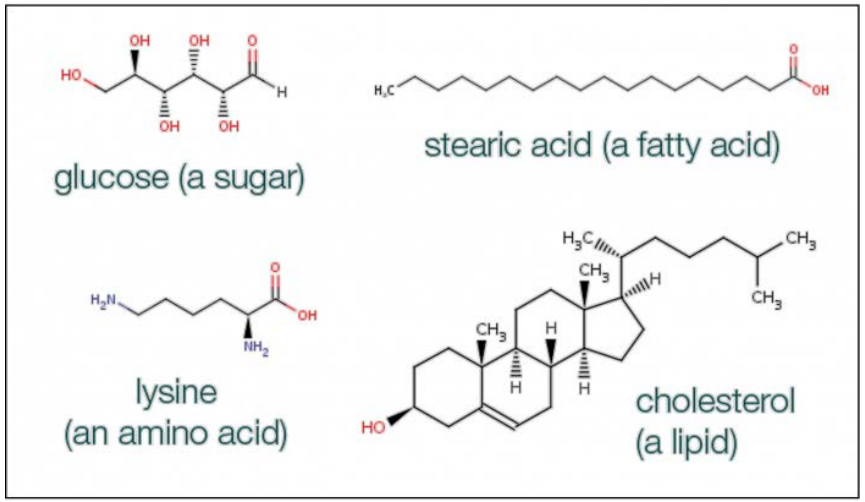
EXTERNAL ENVIRONMENT

climatic conditions ◆ soil ◆ pests ◆ agrochemicals

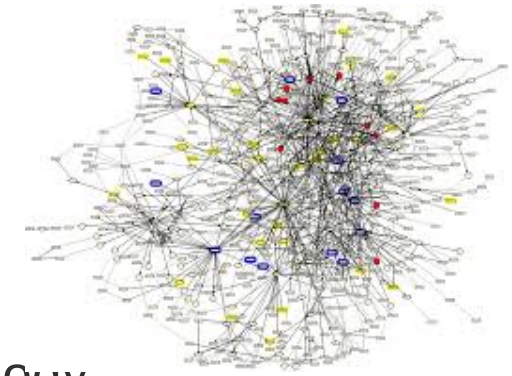
CULTIVAR

Metabolomics

- Oliver, 1998
- -omics with small molecules usually < 1200 Da
- Animals and humans < 3000 metabolites at the single moment
- Plants: > 5000 metabolites
- High variability in concentrations
- High variability of structures



Source:
https://www.ebi.ac.uk/training/online/sites/ebi.ac.uk.training.online/files/user/2760/images/Metabolomics/small_molecules_0.png



- 200,000 metabolites occur in living organisms
- 7,000 - 15,000 within individual plant species
- ´only´ 3,000 endogenous or common metabolites in human body

METABOLOME IS INHERENTLY VERY DYNAMIC

interaction both within and between biological systems, and with the

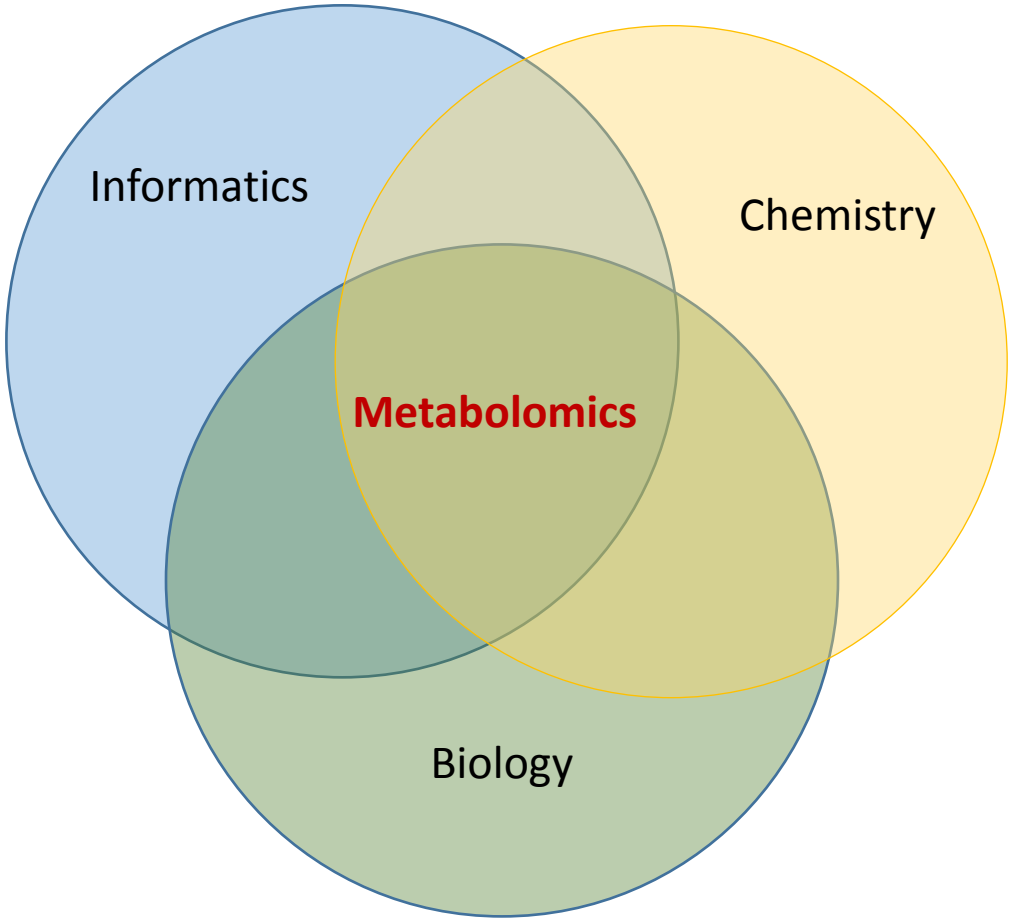
EXTERNAL ENVIRONMENT

climatic conditions ◆ soil ◆ pests ◆ agrochemicals

CULTIVAR

Intersection of fields

- Metabolomics is at the intersection of three fields



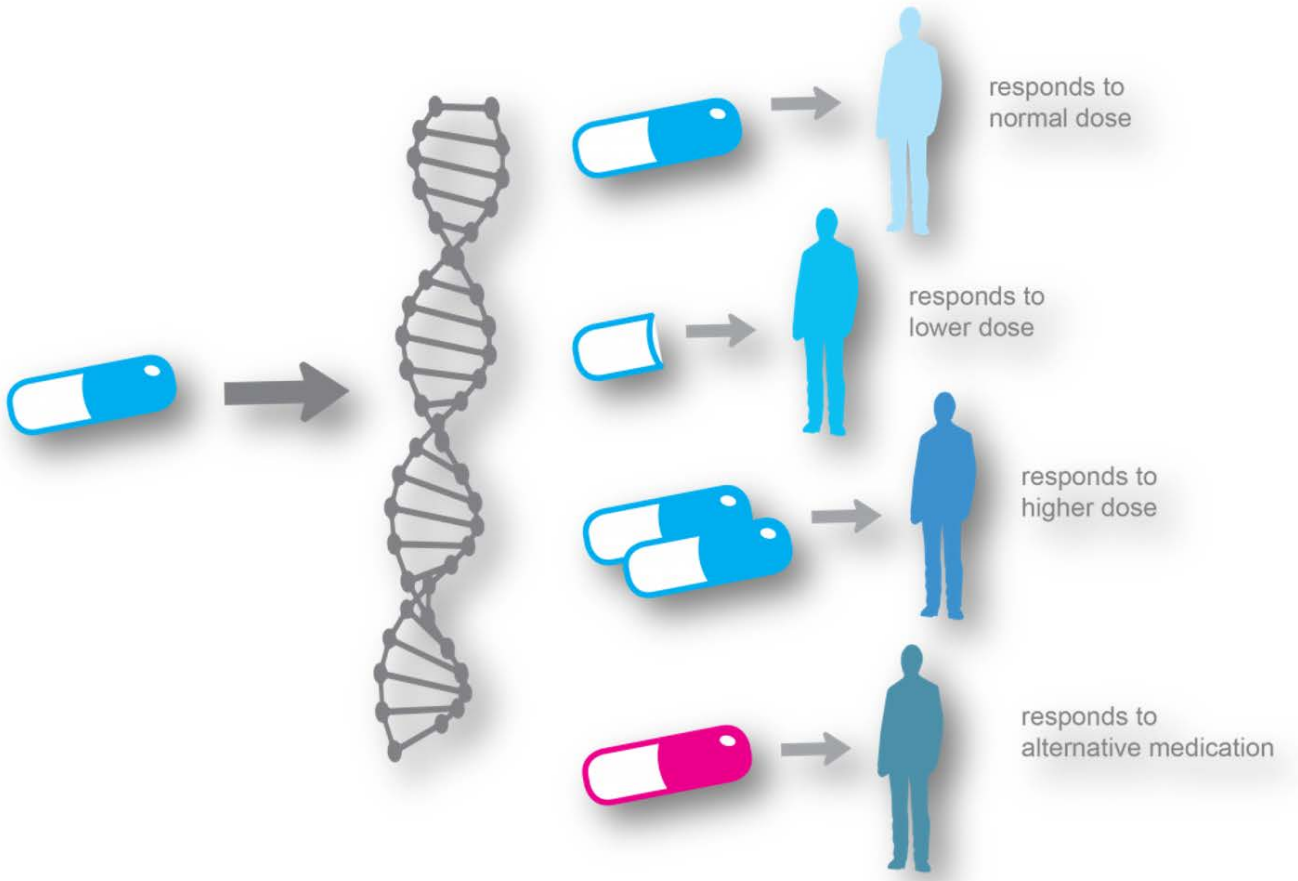
In which fields can metabolomics be used?

Almost everywhere!

- Medicine
- Pharmaceutical science
- Environmental science
- Petroleum chemistry
- Food production
- Food authentication
- …….

Possible applications

- Biomarker discovery
- Personalised medicine
- Pharmaceutics evaluation
- Food safety
- Food authenticity
- Pesticide action

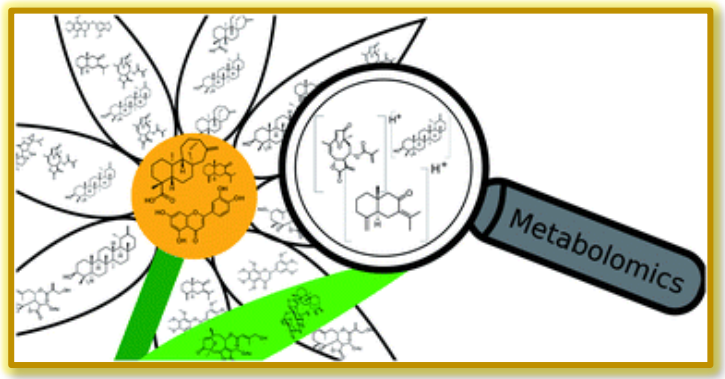


Instrumental platforms for metabolomic fingerprinting

Ambient/DI-MS

SFC/LC/CE–MS

METABOLOME



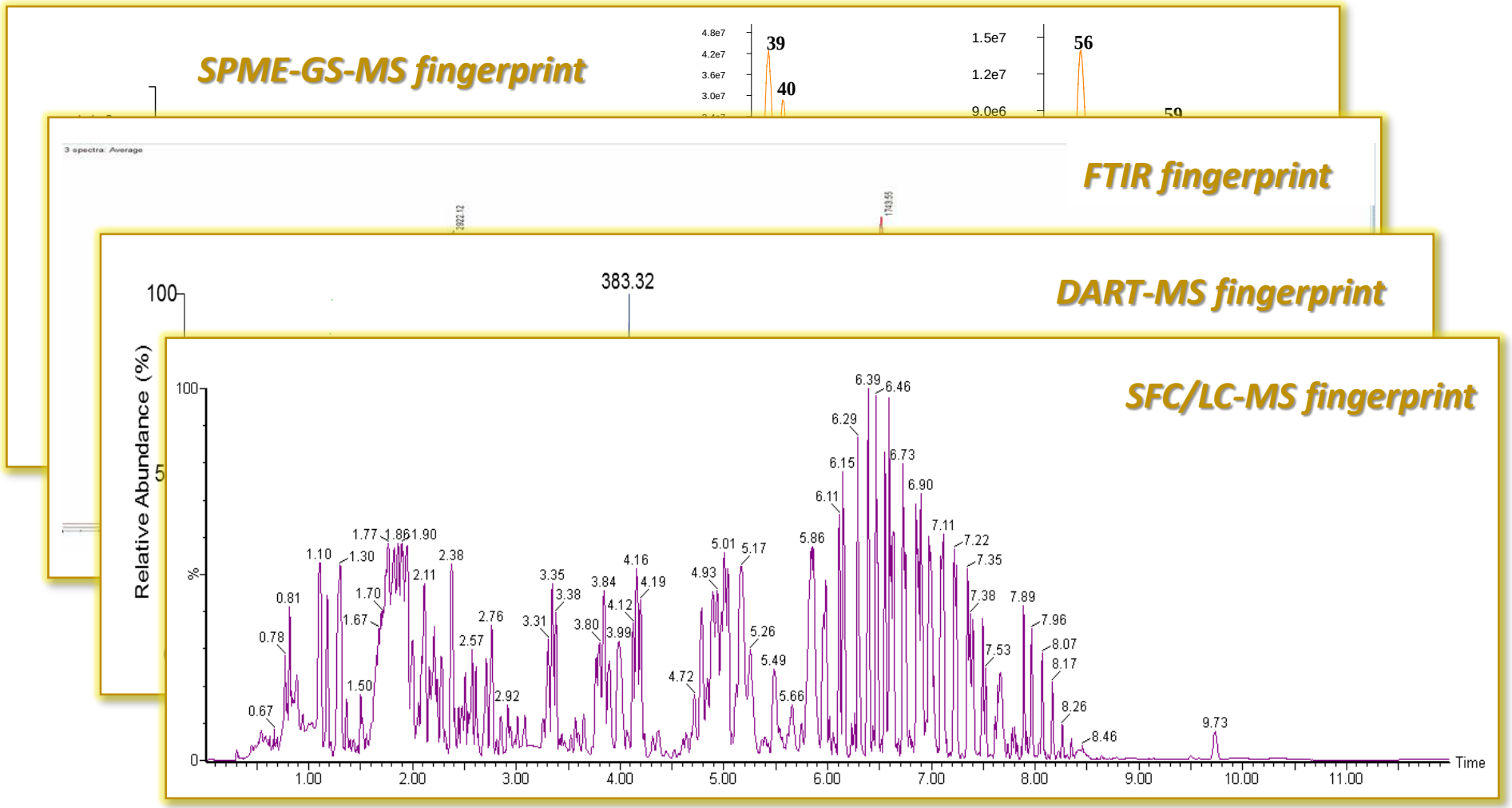
ICP–MS

NMR

GC-MS
SPME-GC-MS

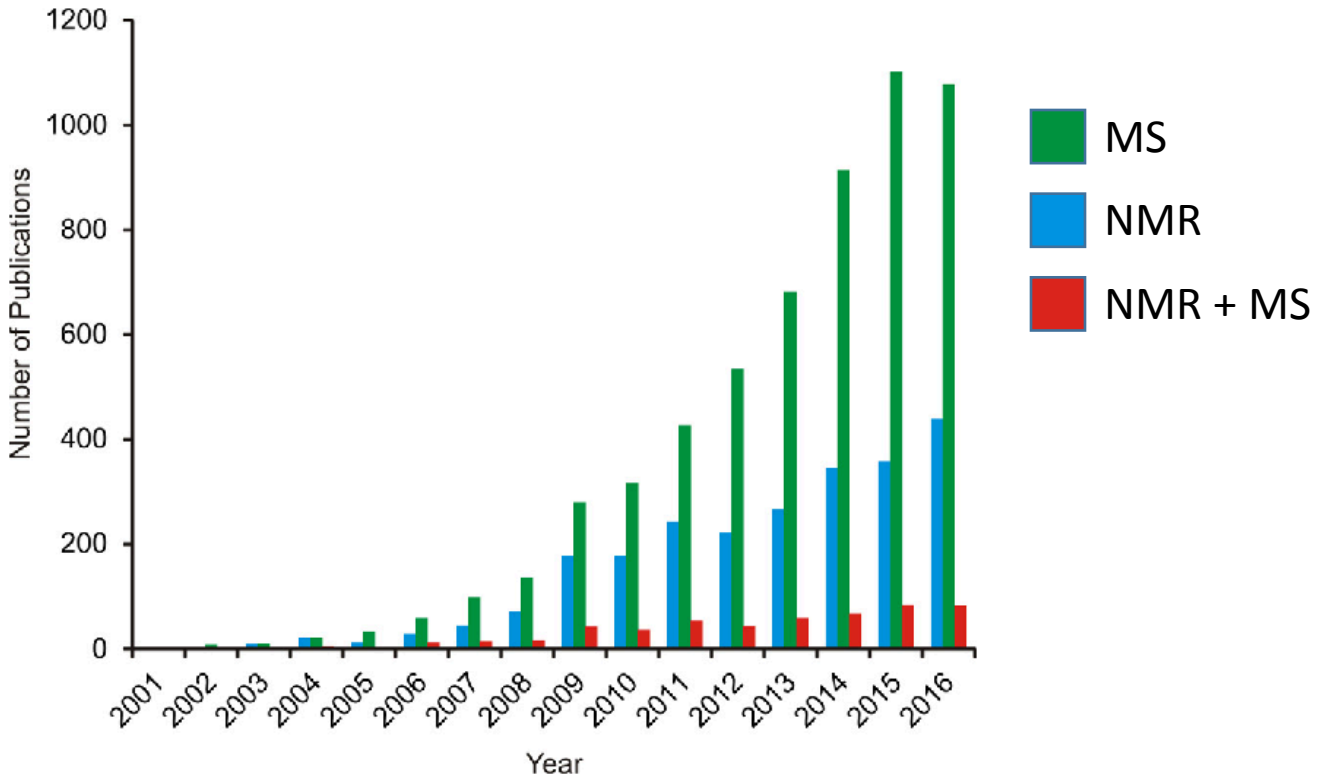
IR / Raman

Metabolomics fingerprint (snapshot)



The expanding field of metabolomics

- While relatively new approach, the field is rapidly emerging
- Two leading techniques NMR and MS



Beyond the paradigm: Combining mass spectrometry and nuclear magnetic resonance for metabolomics.

Darrell D. Marshall, Robert Powers • Published 2017 in Progress in nuclear magnetic resonance... •
DOI: 10.1016/j.pnmrs.2017.01.001

Metabolic unknowns

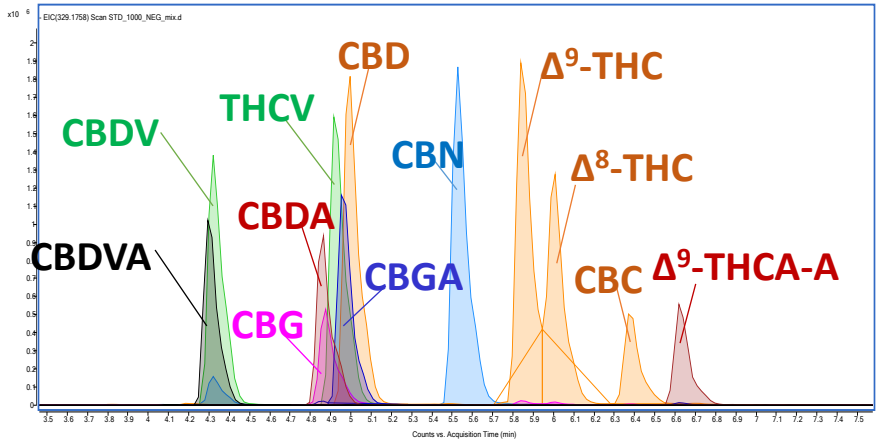
Reports that say that something hasn't happened are always interesting to me, because as we know, there are **known knowns**; there are things we know we know. We also know there are **known unknowns**; that is to say we know there are some things we do not know. But there are also **unknown unknowns** – the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones

Donald Rumsfeld, United States Secretary of Defense, 2002

Metabolomic strategies

Metabonomics

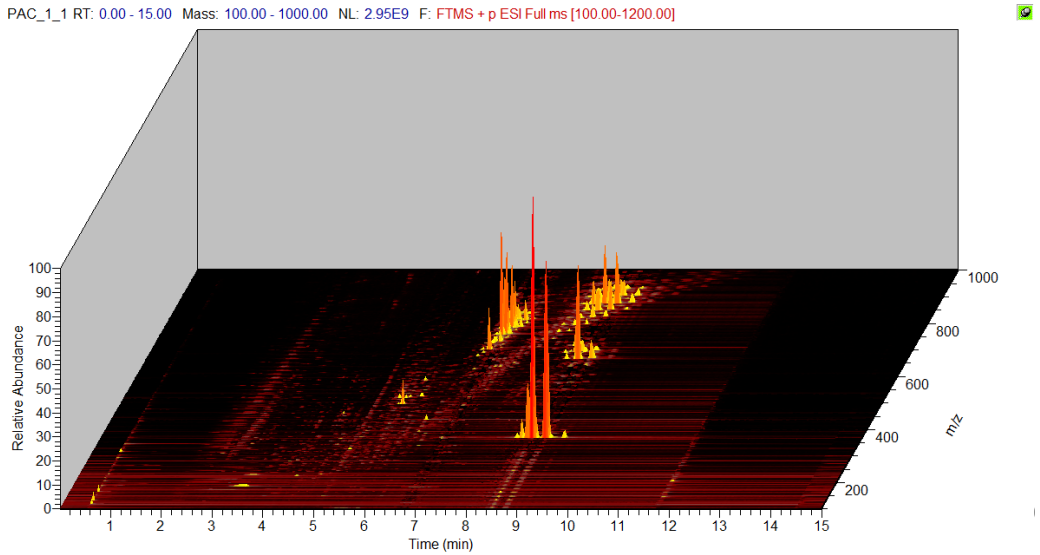
- Targeted
- Quantitative
- Coverage of small set of metabolites
- Biased – compounds from certain biosynthetic pathways, signalling molecules
- Hypothesis driven *e.g.* Cannabinoid metabolism



Metabolomic strategies

Metabolomic profiling

- Targeted
- Non-quantitative
- Coverage of smaller set of metabolites
- Biased – commonly a group of compounds related by structure or function
- Hypothesis driven *e.g.* phenolics for wine authentication



Metabolomic strategies

Metabolomic fingerprinting

- Non-targeted
- qualitative
- Coverage of large set of metabolites
- Least biased
- Hypothesis generating

Metabolomics workflow

- All the parts must be carefully planned
- Quality control must be maintaned during the whole process

- *Sample size*
- *Variability*
- *Confounding factors*
- *...*

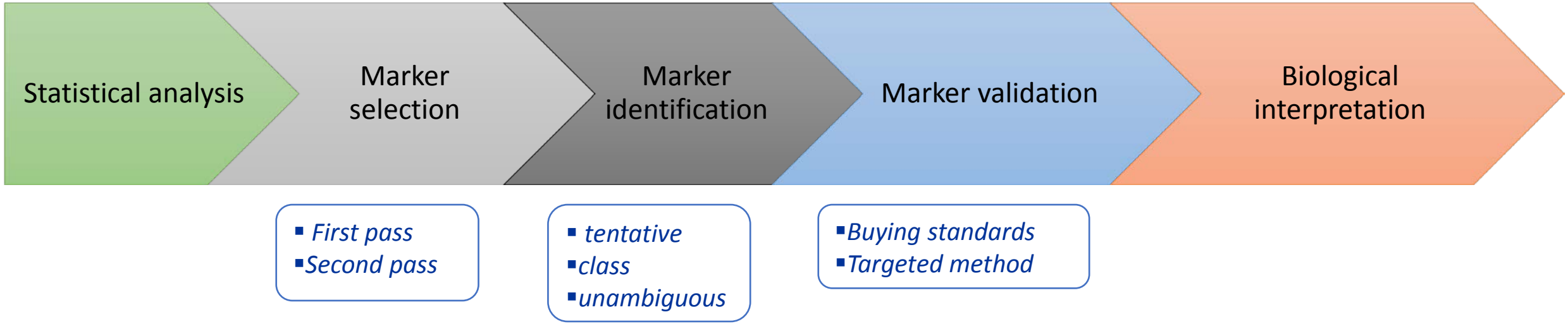
- *Normalisation*
- *Extraction*
- *Repeatability*
- *QC preparation*
- *...*

- *Normalisation*
- *Internal standards*
- *Batch effect*
- *Criteria for peaks selection*
- *Repeatability*
- *...*

- *Overfitting*
- *Models validation*
- *...*

Metabolomics workflow

- Statistical analysis and interpretation is usually the most time consuming step



Sample preparation

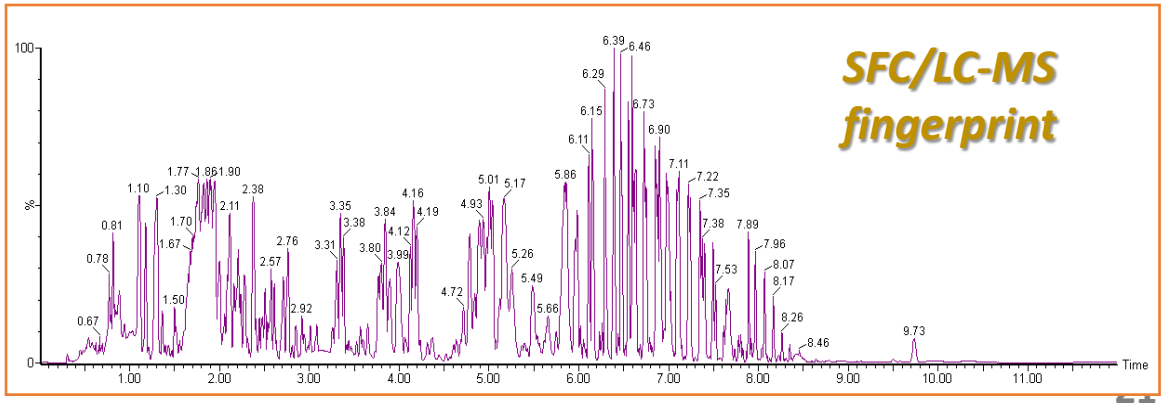
- Generally very simple -> reduction of error
- Generic methods are used -> increasing coverage
- Derivatization may be neded for certain analytes

Data acquisition

Requirements:

- Stable conditions throughout the measurement
- Quality control procedures

- *Incorrectly acquired data -> useless results*

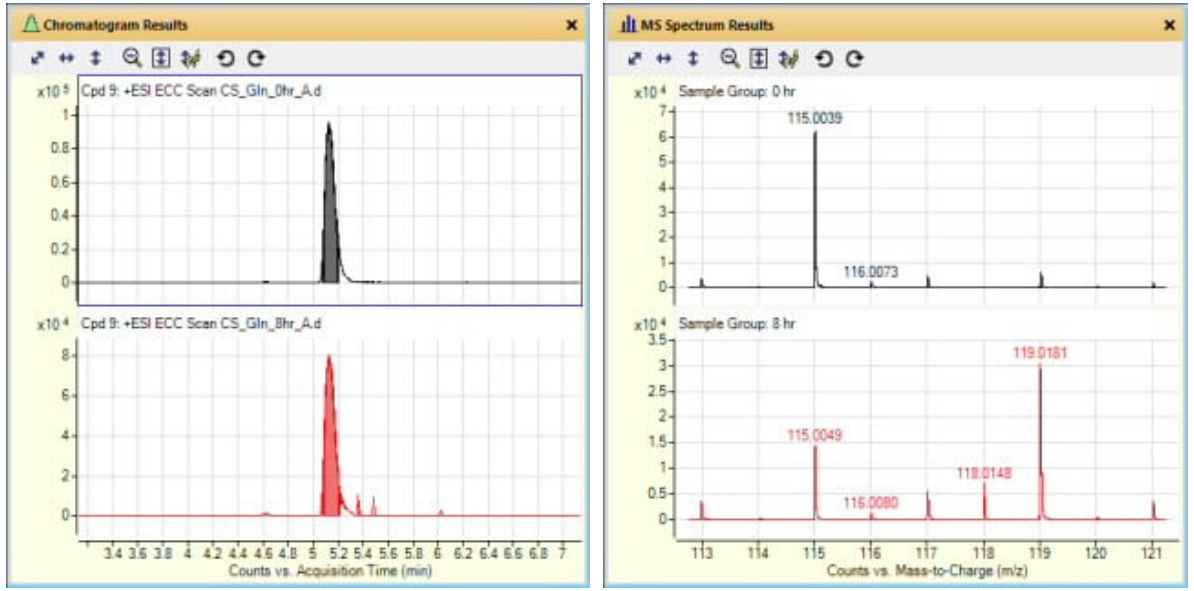


Data processing

- One of the most crucial steps in the workflow

Bad data + good processing = bad results
Good data + bad processing = bad results
Good data + good processing= GOOD RESULTS

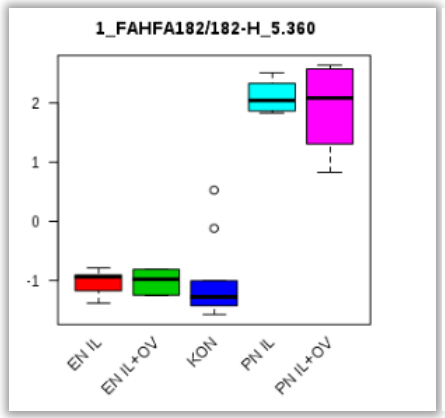
- Peak picking
- Peak alignment
- Data normalization
- Transformation
- Scaling

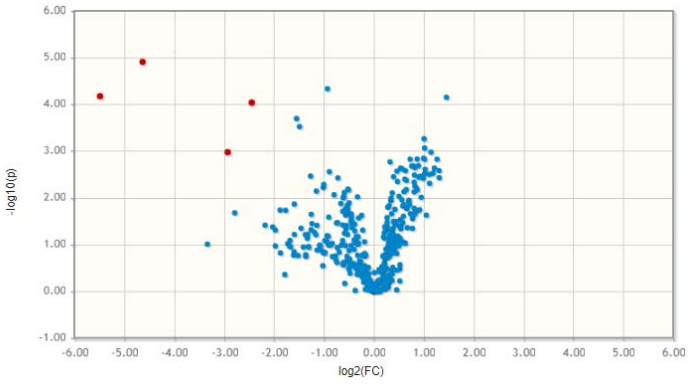


Statistical analysis

Univariate statistics

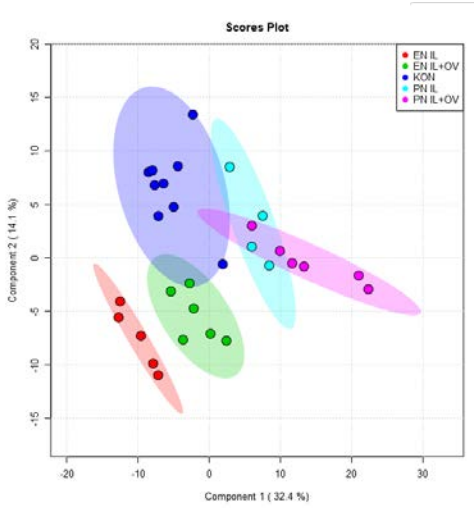
- More traditional approach
- Useful for individual variable evaluation and filtering

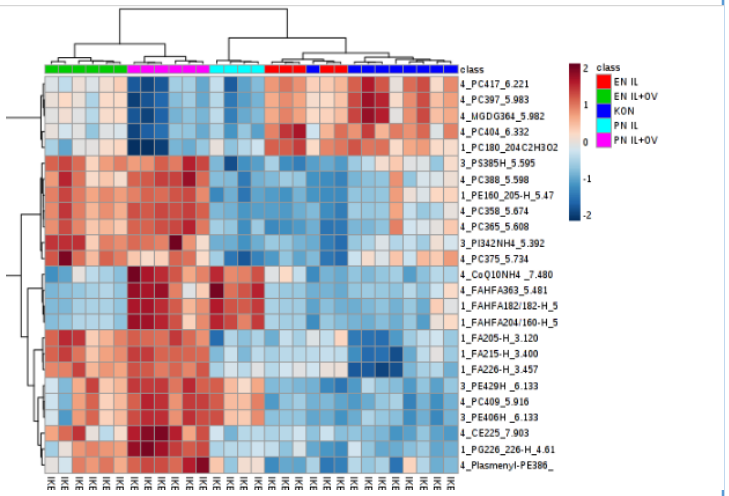




Multivariate statistics

- Simplification of multidimensional data
- Useful for evaluation of patterns in the data





Challenges - biological

Challenge 1: be able to characterize discrete signature

Limitation 1: it may be hidden by other sources of variability

Challenge 2: be able to characterize one system through the generation of a unique metabolic profile

Limitation 2: the metabolome is a dynamic system (for example diurnal and seasonal variation in human studies…)

Challenge 3: be able to connect genome and metabolome (systems biology)

Limitation 3: difficulties to collect both informations

… these biological challenges correspond to future directions in research

Challenges - analytical

- **Challenge 1:** be able to characterize the whole metabolome

  Limitation 1: at the moment, there is no such versatile instrument allowing to analyze such chemical diversity

- **Challenge 2:** long term repeatability of analytical sequences (when 100's to 1000's samples are analyzed)

  Limitation 2: still insufficient stability of the MS-instrument acknowledged, need for efficient way of normalization with Quality Controls

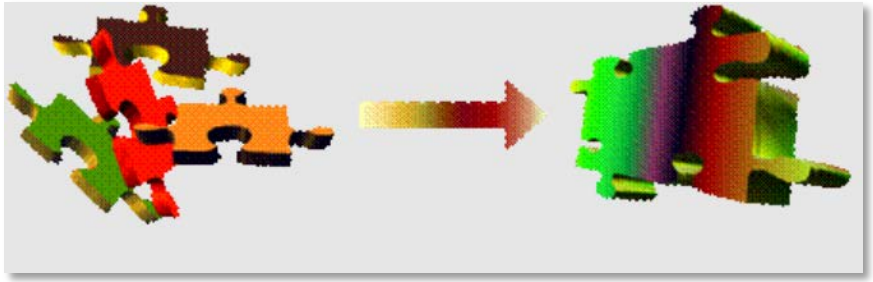
- **Challenge 3:** be reproducible between analytical platforms to allow comparison

  Limitation 3: Used protocols are different, need for standardization procedures

… these analytical challenges correspond to future directions in research

Summary

- Metabolomics is a large-scale study of small molecules (<1200 Da)
- Various applications: clinical, pharmaceutical, agri-food
- Untargeted and targeted approaches
- Planning and design of the study is crucial for succesful employment of metabolomics
- Two major analytical aproaches: NMR and MS (hyphenated or not)



UNIVERSITY OF CHEMISTRY AND TECHNOLOGY, PRAGUE
Faculty of Food and Biochemical Technology
Department of Food Analysis and Nutrition

Case study: WHAT IS THE ORIGIN OF THIS GARLIC?

METABOLOMIC FINGERPRINTING EMPLOYING HRMS MAY GIVE A RAPID ANSWER

Jana Hajslova, Vojtek Hrbek,
Misa Rektorisova, Monika Tomaniova

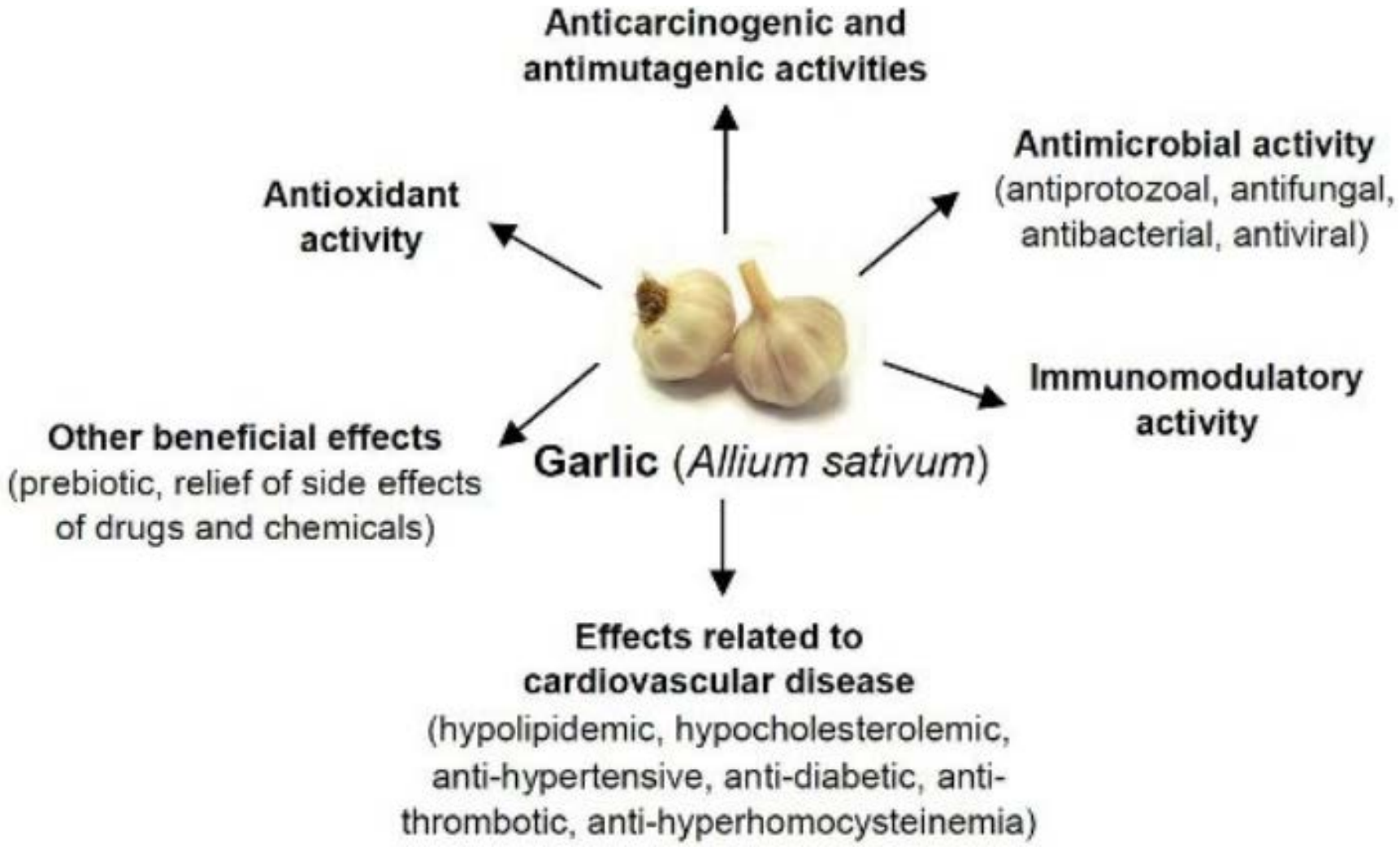
Garlic (*Allium sativum* L.)

Top garlic producers in 2014	
Country	Production (millions of tonnes)
China	20.0
India	1.25
South Korea	0.35
Egypt	0.26
Russia	0.26
World	**25.0**
May include official, semi-official or estimated data Source: UN Food & Agriculture Organisation[18]	

One of the most important vegetables throughout the world, with a total annual production of 24 mil tonnes of dry bulbs

Planting of garlic, 15th century

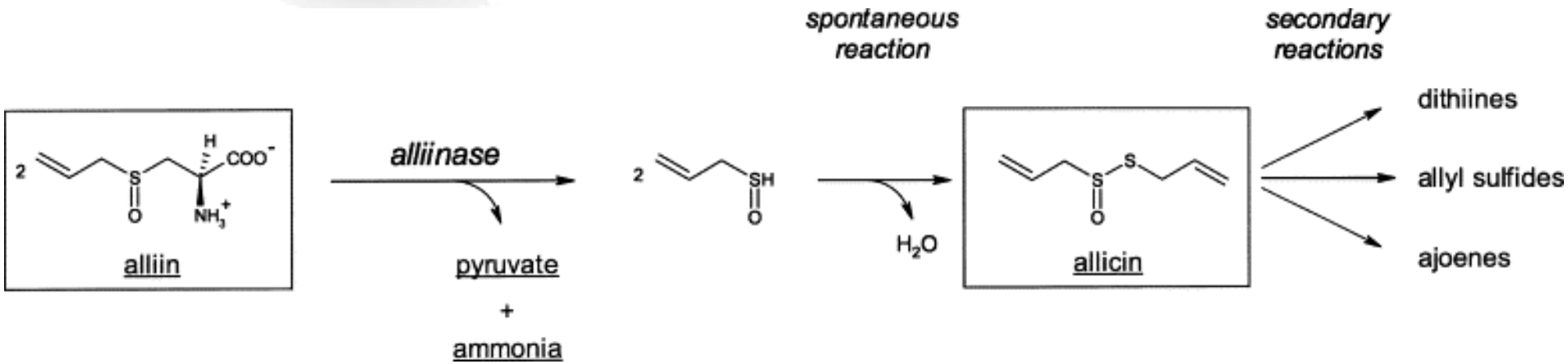
Health benefits associated with garlic



Bioactive compounds in garlic

- Vitamins
- Flavonoids
- Minerals
- …

UNIQUE COMPONENTS:
Sulfur Amino Acids *S*-alk(en)yl-L-cysteinsulfoxides (ACSO):
alliin, isoalliin, methiin, propiin…



AROMA of freshly chopped garlic

AROMA of culinary modified garlic

Garlic – a very popular vegetable in the Czech republic

Alike other highly valued food commodities, garlic may become subject of fraudulent practices, mislabelling of its origin being one of the most common one.

Fraud on garlic?

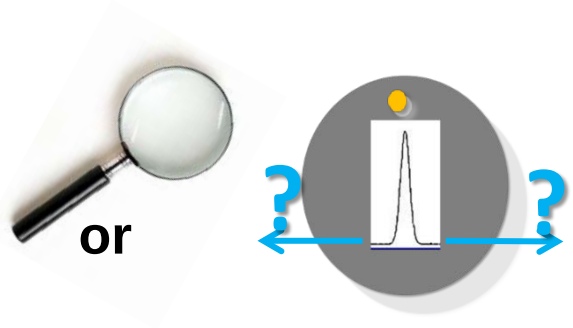
China Has Been Secretly Bleaching Garlic And Shipping It To The U.S. Here's How To Spot It.

Food, Health Like 2.3M

Let´s search solution: FINGERPRINTS (?)

´CLASSIC´ APPROACH

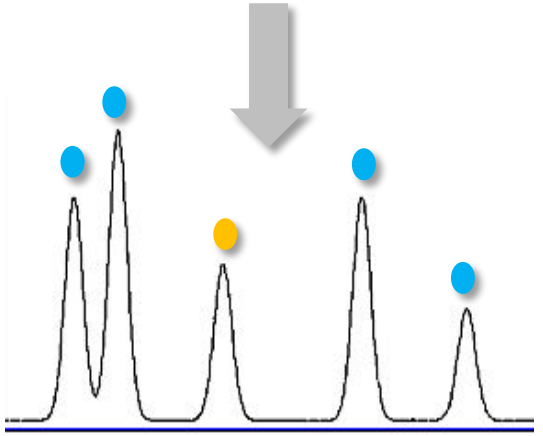
A set of physico-chemical and/or biochemical measurements → TARGET ANALYSIS of one few markers



NOVEL STRATEGY

Metabolomic fingerprinting

→ NON TARGET SCREENING



- characterization of components pattern
- detection / identification of ´unknown´ components (even retrospective)
- identification of a set of composition markers

Assessment of HRMS based platforms employed for fingerprinting purpose

Ambient mass spectrometry DART - HRMS

- ↑ Fast, no simple preparatio
- ↓ Limited scope, isomers unresolved

Direct infusion mass spectrometry HRMS

- ↑ Entire metabolome fingrrprinted
- ↓ Ion source contamination

Chromatography mass spektrometry UHPLC – HRMS/MS

- ↑ Two separation dimensions
- ↓ Time demands, Rt reproducibility

Garlic samples – experimental set

The samples e delivered from Crop Research Institute in Prague

Variables

- varieties
- morphotype
- origin
- growing parctices

Number of samples

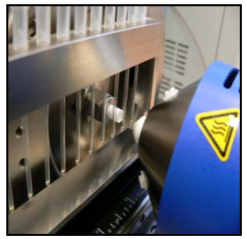
19 - Czech Rep

17 – Spain

11 – China

METHANOLIC EXTRACTS PREPARED

1. DART–HRMS

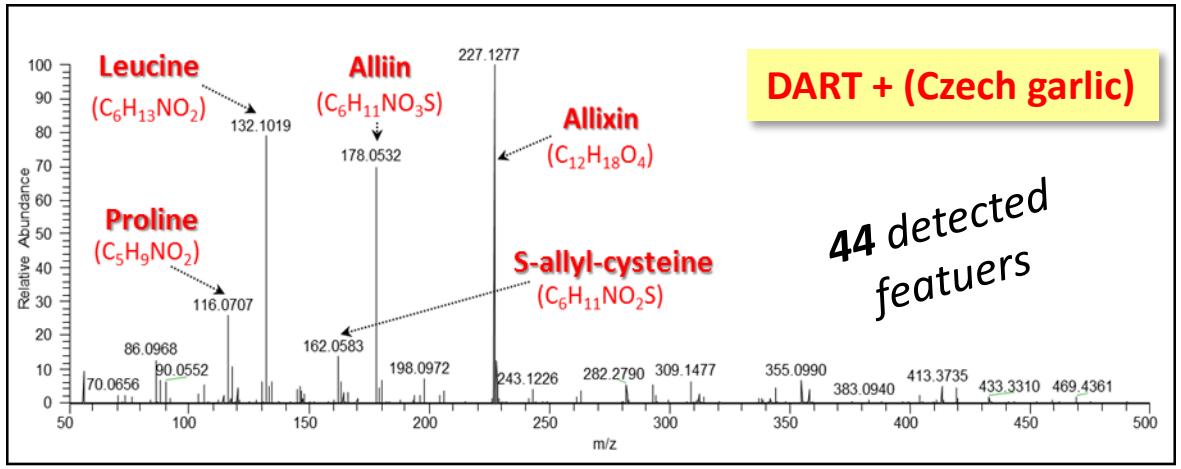


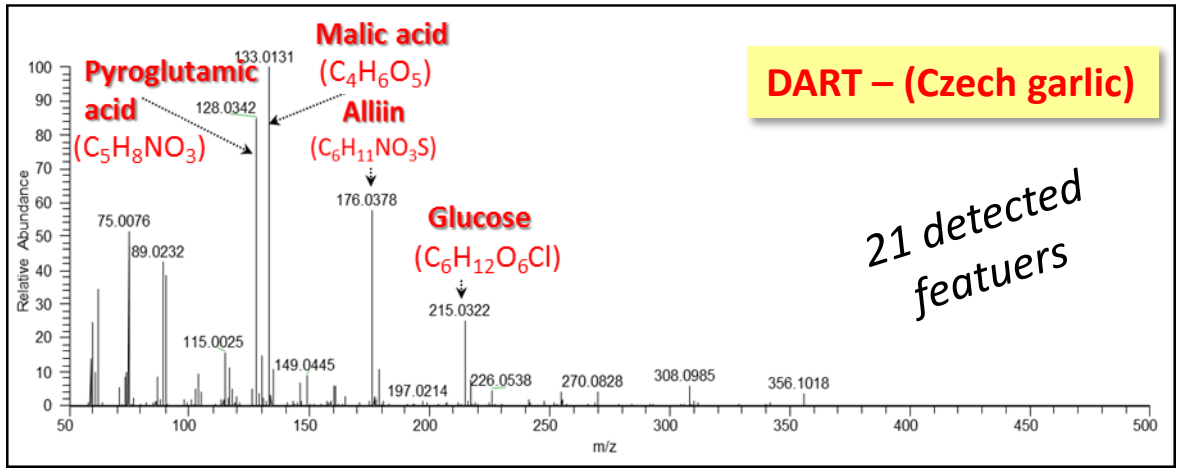
Characteristic fingerprint of the garlic samples

characterization, authentication of garlic samples

Instrumentation:

- Exactive OrbitrapMS (Thermo)
- DART – **D**irect **A**nalysis in **R**eal **T**ime (IonSence)





2. DI–ESI-HRTOFMS (direct infusion)

Instrumentation:

Synapt G2 –LC-QTOFMS (Waters)

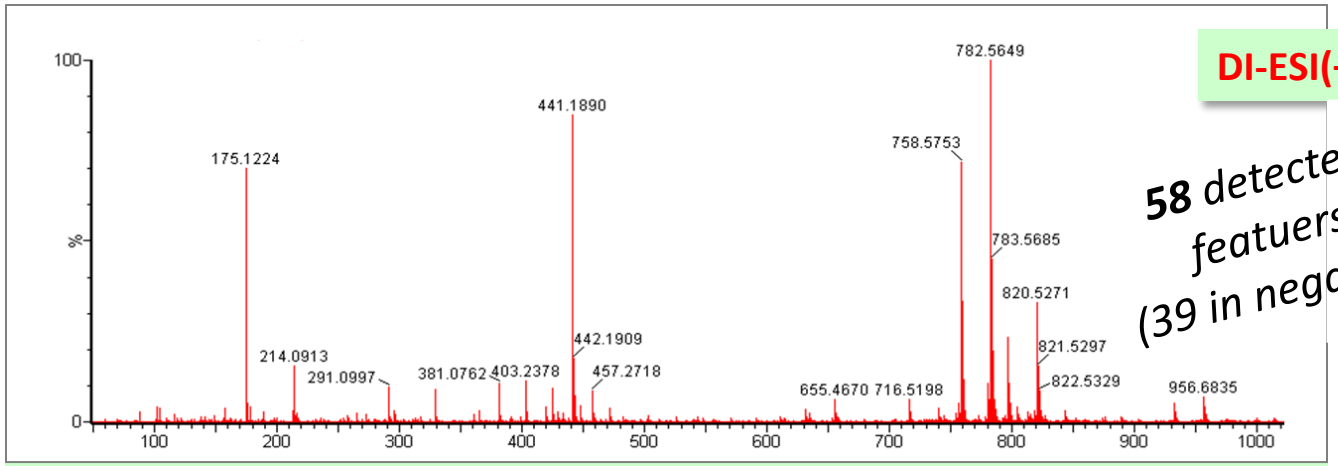
- No separation involved!
- Mobile phase: MeOH:H_2O, 50:50

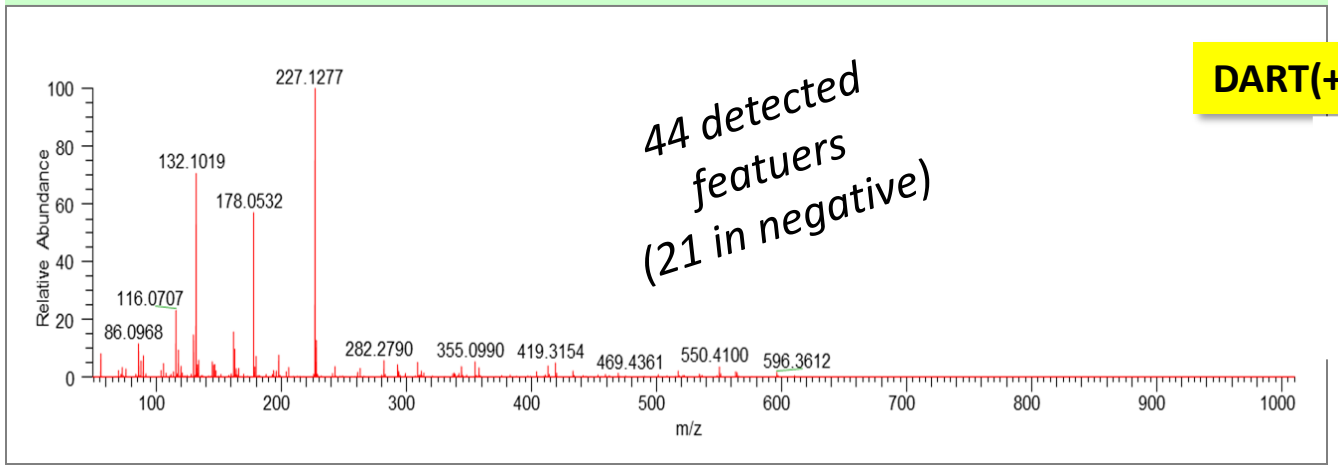
Alternative to DART-MS

- **DI: different way of ionization (ESI)**
- **DI: better automation due to autosampler**

Characteristic fingerprint of the garlic samples

characterization, authentication of garlic samples





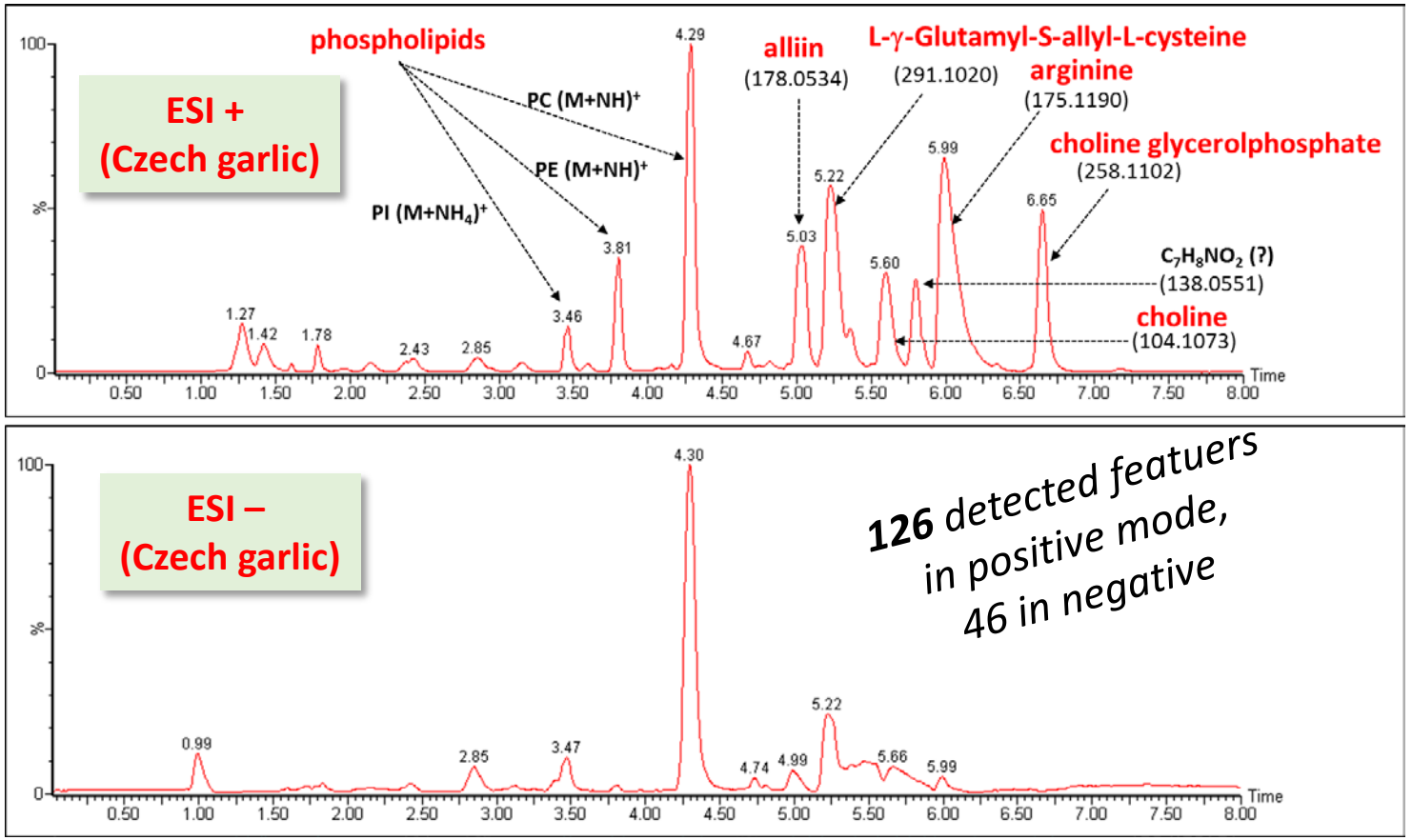
3. LC–ESI-QTOFMS

Instrumentation: Synapt G2 – LC-ESI-QTOFMS (Waters)

- Column: HILIC (2.1x150 mm, 3 μm)
- Mobile phase: Acetonitrile; H_2O (50 mM $HCOONH_4$, 0.2% HCOOH)

Characteristic fingerprint of the garlic samples

characterization, authentication of garlic samples



Validation of fingerprinting analysis

Repeatability of measurements expressed as relative standard deviation (RSD, %), for each of the tested MS techniques

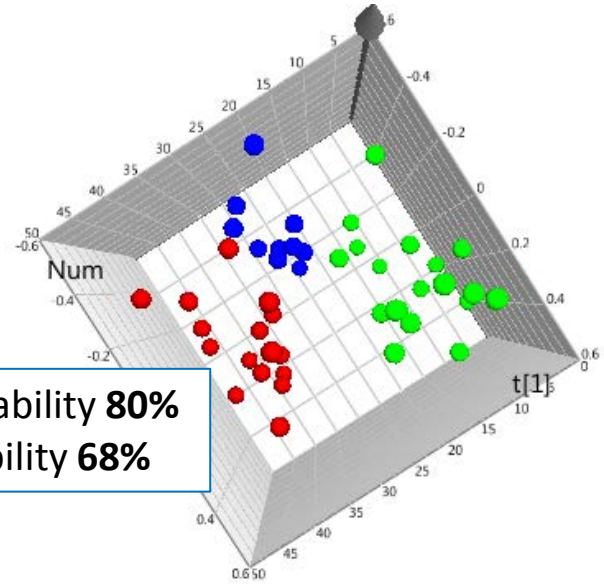
technique	**HPLC-HRMS**		**DI-HRMS**		**DART-HRMS**	
resolving power (FWHM)	20 000		20 000		50 000	
ionization mode	positive	negative	positive	negative	positive	negative
RSD (%)	**3.4-6.8**	**3.7-7.2**	**11.4-18.6**	**12.5-18.4**	**16.5-22.6**	**19.3-28.4**

Statistical analysis – positive ions

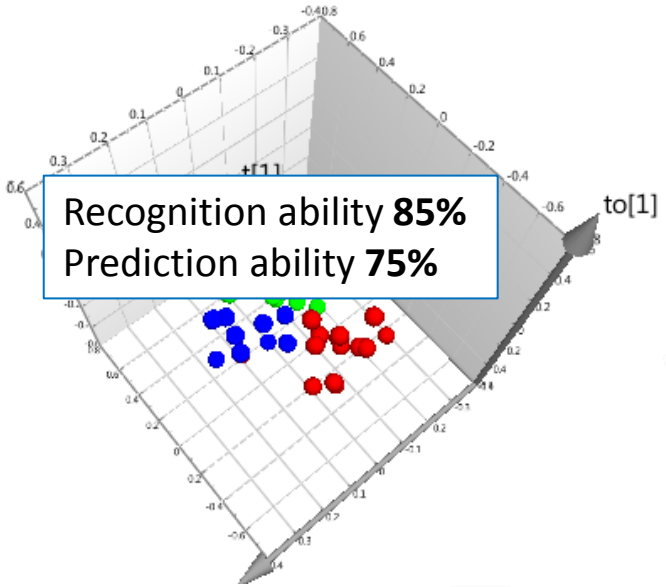
1. step = PCA analysis

2. step = PLS-DA / OPLS-DA analysis

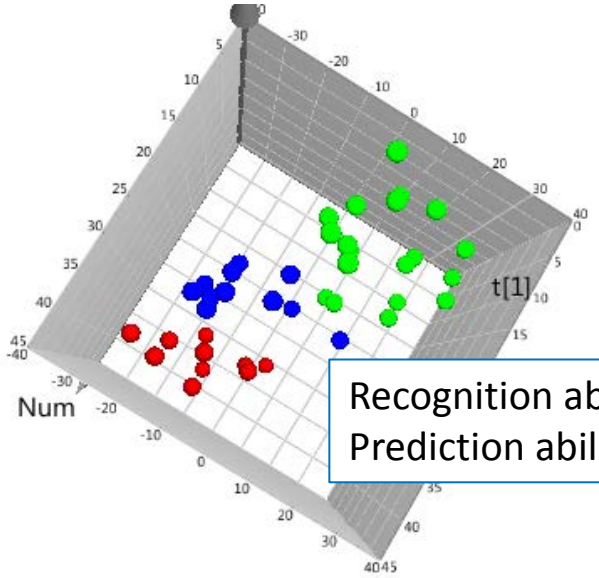
PLS-DA models from data of all 3 groups of samples, all 3 country of origin …



Recognition ability **80%**
Prediction ability **68%**



Recognition ability **85%**
Prediction ability **75%**



Recognition ability **96%**
Prediction ability **85%**

Czech Republic **China** **Spain**

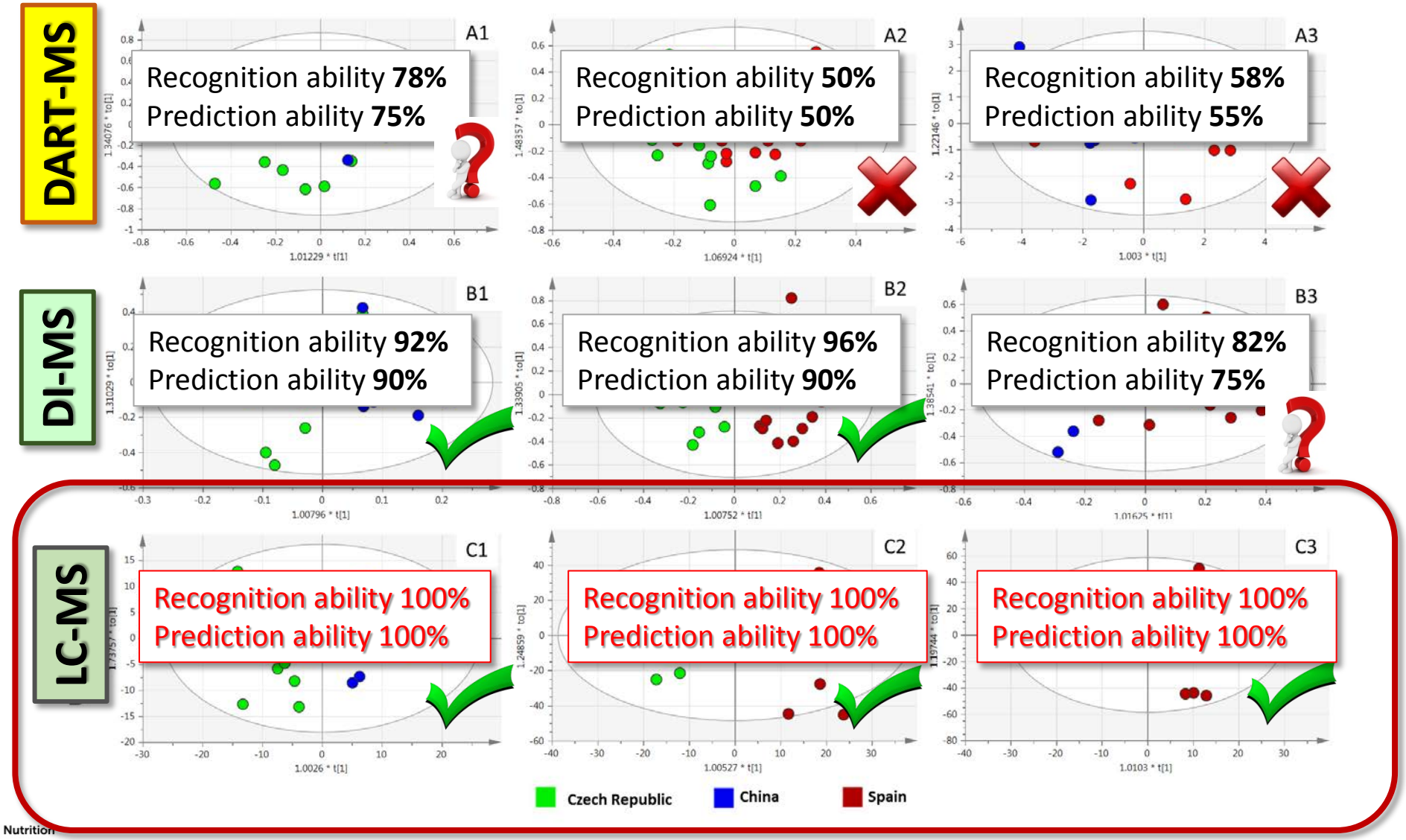
DART-MS
44 featuers

DI-MS
58 featuers

LC-MS
126 featuers

Statistical analysis

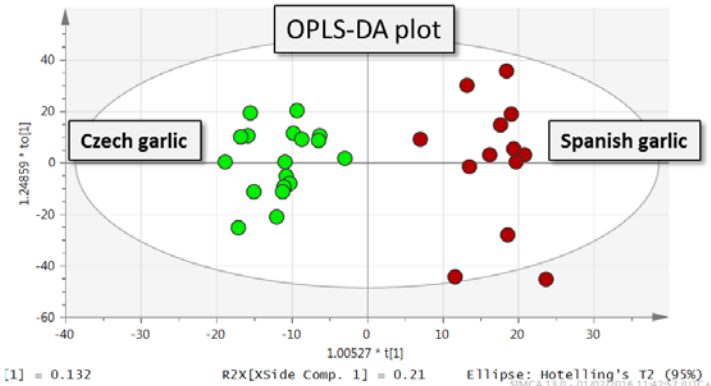
Models from data of only 2 groups of samples, only 2 country of origin

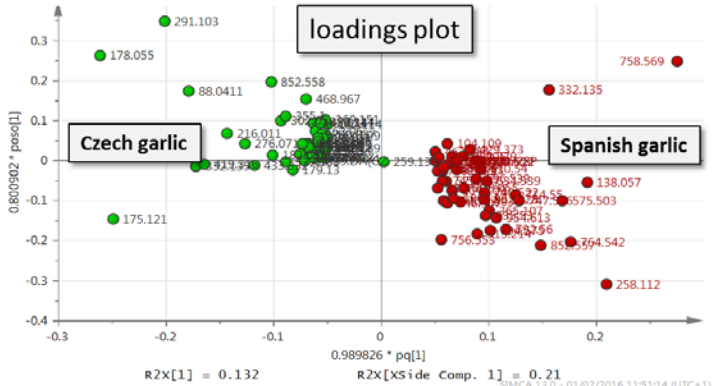


Examples of markers identification

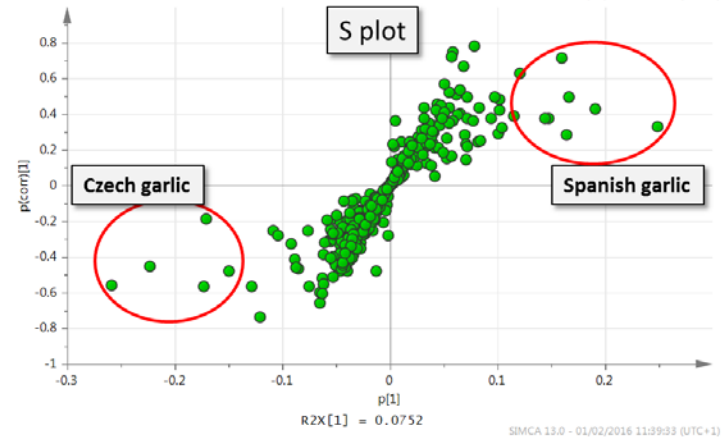
SIMCA

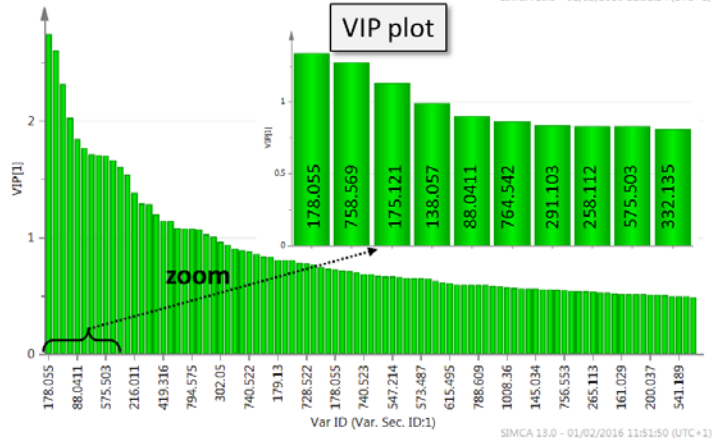
1)



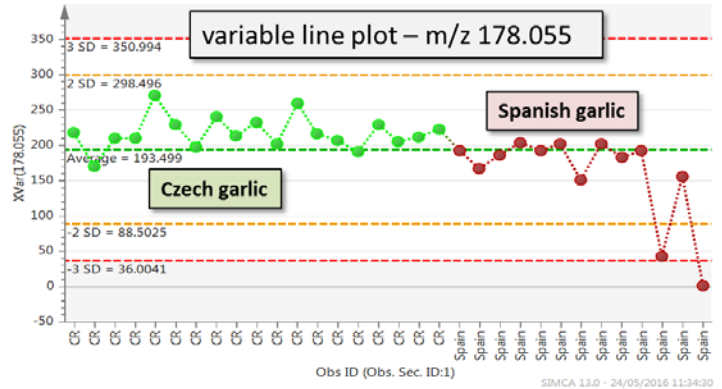


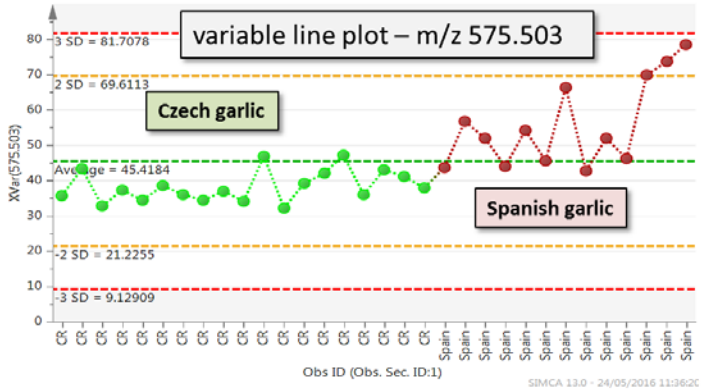
2)





3)





Tentatively identified markers

ion *m/z* ⇒ exact mass, Δppm, isotopic profile, calculated formula ⇒ database ⇒ **TENTATIVE IDENTIFICATION**

Ten the most important ions for the separation of garlic samples, sorted by descending importance:

retention time	m/z	elemental formula	mass error Δppm	tentative identification
5.03	178.0551	$C_6H_{11}NO_3S$	1.7	alliin
4.29	758.5690	$C_{42}H_{80}NO_8P$	0.7	PC (16:0/18:2)
5.99	175.1208	$C_6H_{14}N_4O_2$	0.6	Arginine
5.8	138.0570	$C_7H_7NO_2$	3.6	---
5.04	88.0411	$C_3H_5NO_2$	3.5	dehydroalanine/oxazolidinone
2.84	764.5425	$C_{43}H_{74}O_8NP$	2.9	PE (16:0/22:6)
5.24	291.1026	$C_{11}H_{18}N_2O_5S$	1.0	L-γ-Glutamyl-S-allyl-L-cysteine
6.65	258.1118	$C_8H_{20}NO_6P$	2.3	choline glycerolphosphate
3.45	575.5031	$C_{37}H_{66}O_4$	0.0	fragment of PI (16:0/18:2), lost of phosphatydylinositol part
4.36	332.1354	$C_8H_{22}N_5O_7P$	0.9	---

CONCLUSIONS

- Metabolomic fingerprinting of garlic (MeOH extracts) performed by HRMS-based techniques has been demonstrated to have a potential to identify garlic origin
- Regarding the reporting value of generated data: UHPLC- HRMS > > DI-HRMS >DART-HRMS
- Excellent prediction ability, up to 100%, of OPLS-DA models could be achieved in particular case
- Marker compounds applicable for garlic screening
- **CONTINUOUS BUILDING UP THE DATABASE BASED ON THE DATA OBTANED ON AUTHENTIC SAMPLES IS NEEDE TO GET ROBUST CLASSIFICATION MODELS AND VALIDATED MARKERS**

Journal of Food Composition and Analysis 67 (2018) 19–28

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

journal homepage: www.elsevier.com/locate/jfca

Authenticity assessment of garlic using a metabolomic approach based on high resolution mass spectrometry

Vojtech Hrbek[a], Michaela Rektorisova[a], Hana Chmelarova[a], Jaroslava Ovesna[b], Jana Hajslova[a,*]

[a] University of Chemistry and Technology, Prague, Faculty of Food and Biochemical Technology, Department of Food Analysis and Nutrition, Technicka 3, 166 28 Prague 6, Czech Republic

[b] Crop Research Institute, Prague, Drnovska 507/73, 161 06 Prague 6, Czech Republic

ARTICLE INFO

Keywords:
Garlic
Authenticity
Geographical origin
Metabolomics
High resolution mass spectrometry
Opls-Da
Classification
Food composition
Food analysis

ABSTRACT

Depending on conditions in a growing locality and several other factors, marketed garlics (*Allium sativum* L.) may largely differ in content of flavour significant compounds and other biologically active components. To enable verification of traders declarations on the geographic origin, a new analytical, metabolomic fingerprinting, was employed for analysis of 47 samples of garlic with the designated country of origin Czech Republic, Spain and China. Non-target screening of metabolome components occurring in garlic extracts was performed employing following three instrumental platforms based on high resolution mass spectrometry (HRMS): (i) ambient mass spectrometry utilizing direct analysis in real time ionization (DART) ion source coupled to HRMS; (ii) direct infusion (DI) of sample into electrospray ion source (ESI) coupled to HRMS; (iii) high performance liquid chromatography (HPLC) – ESI – HRMS. Statistical models (Orthogonal Partial Least Squares-Discriminant Analysis, OPLS-DA) models were constructed on generated data with the aim to identify the best HRMS technique enabling a reliable differentiation of a country of origin. The best prediction ability, up to 100%, was obtained by processing the data generated by HPLC-HRMS. Alliin, phosphatidylcholine (16:0/18:2), arginine, dehydroalanine, phosphatidylethanolamine (16:0/22:6), L-γ-Glutamyl-*S*-allyl-L-cysteine and choline glycerolphosphate, were identified as compounds most contributing to a correct classification of the samples.

Thank you for your kind attention…

jana.hajslova@vscht.cz