



**HAL**  
open science

## Using metabolomics and predictive metabolomic to study the physiology of an old rustic oilseed plant.

Malo Le Boulch, Millena C. Barros Santos, Susana Silvestre, Dominik Grosskinsky, Anaïs da Costa, Frédérique Tellier, Amélie Flandin, Cédric Cassan, Paloma Leon, Sophie Colombié, et al.

### ► To cite this version:

Malo Le Boulch, Millena C. Barros Santos, Susana Silvestre, Dominik Grosskinsky, Anaïs da Costa, et al.. Using metabolomics and predictive metabolomic to study the physiology of an old rustic oilseed plant.. *Metabolomics* 2023, Jun 2023, Niagara falls, Canada. hal-04396004

**HAL Id: hal-04396004**

**<https://hal.inrae.fr/hal-04396004v1>**

Submitted on 15 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNTWIST will use the resilient European oilseed crop camelina in a multidisciplinary approach to decipher the diverse heat and drought stress response mechanisms associated with genetic diversity and use this knowledge to establish predictive models for the tangible improvement of cropping systems.

# Using metabolomics and predictive metabolomics to study the physiology of an old rustic oilseed plant

Malo Le Boulch<sup>1</sup>, Millena Barros-Santos<sup>1,2,3</sup>, Susana Silvestre<sup>4</sup>, Dominik Grosskinsky<sup>6</sup>, Anais Da Costa<sup>7</sup>, Frédérique Tellier<sup>7</sup>, Amélie Flandin<sup>1,2,3</sup>, Cédric Cassan<sup>1,2,3</sup>, Paloma Leon<sup>9</sup>, Sophie Colombié<sup>1,2,3</sup>, Javier Prieto<sup>9</sup>, Federica Zanetti<sup>8</sup>, Jean-Denis Faure<sup>7</sup>, Claudia Jonak<sup>6</sup>, Björn Usadel<sup>5</sup>, Richard Haslam<sup>4</sup>, Yves Gibon<sup>1,2,3</sup>, Pierre Pétriacq<sup>1,2,3\*</sup>, Sylvain Prigent<sup>1,2,3\*</sup>

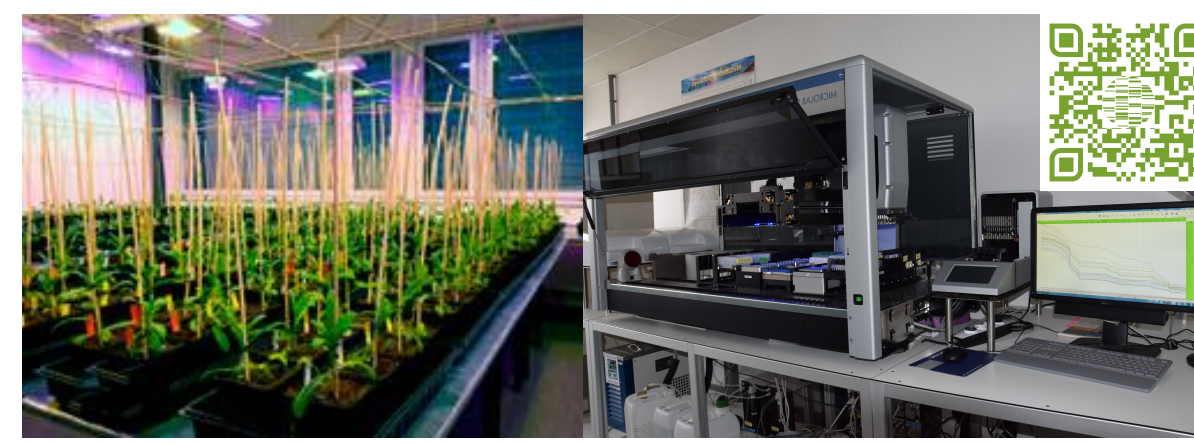
\* pierre.petriacq@inrae.fr, sylvain.prigent@inrae.fr

1: Univ. Bordeaux, INRAE, UMR1332 BFP, 33882 Villenave d'Ornon, France 2: Bordeaux Metabolome, MetaboHUB, PHENOME-EMPHASIS, 33140 Villenave d'Ornon, France 3: MetaboHUB, National Infrastructure of Metabolomics and Fluxomics, France. 4: Plant Sciences & Bioeconomy, Rothamsted Research, Harpenden, UK. 5: IBG-4 Bioinformatics, Forschungszentrum Jülich, 52428 Jülich, Germany 6: AIT Austrian Institute of Technology, Center for Health and Bioresources, Bioresources Unit, 3430 Tulln an der Donau, Austria 7: Université Paris-Saclay, INRAE, AgroParisTech, Institut Jean-Pierre Bourgin (IJPB), 78000 Versailles, France 8: Alma Mater Studiorum, Università di Bologna, DISTAL, Viale G. Fanin 44, 40127 Bologna, Italy 9: Camelina Company, Spain

## Introduction

- Targeted and untargeted metabolomics analyses on 54 *Camelina sativa* lines grown under three different conditions (control, heat, and drought stress).
- The same 54 lines were also phenotyped in field trials across Europe (France, England, Italy and Spain) and genotyped.
- We used machine-learning and data analysis to link greenhouse metabolome and field phenotypes of the different *Camelina sativa* lines.

## Methodology



**Greenhouse data acquisition**  
- On leaves of plants grown in greenhouses under control and stress conditions  
- Targeted and untargeted metabolomics  
- Genome sequencing of the 54 lines



**Phenotyping data acquisition in fields**  
- Under agronomical conditions  
- In four different locations across Europe

54 lines  
3 conditions  
950 samples

Blanks QC

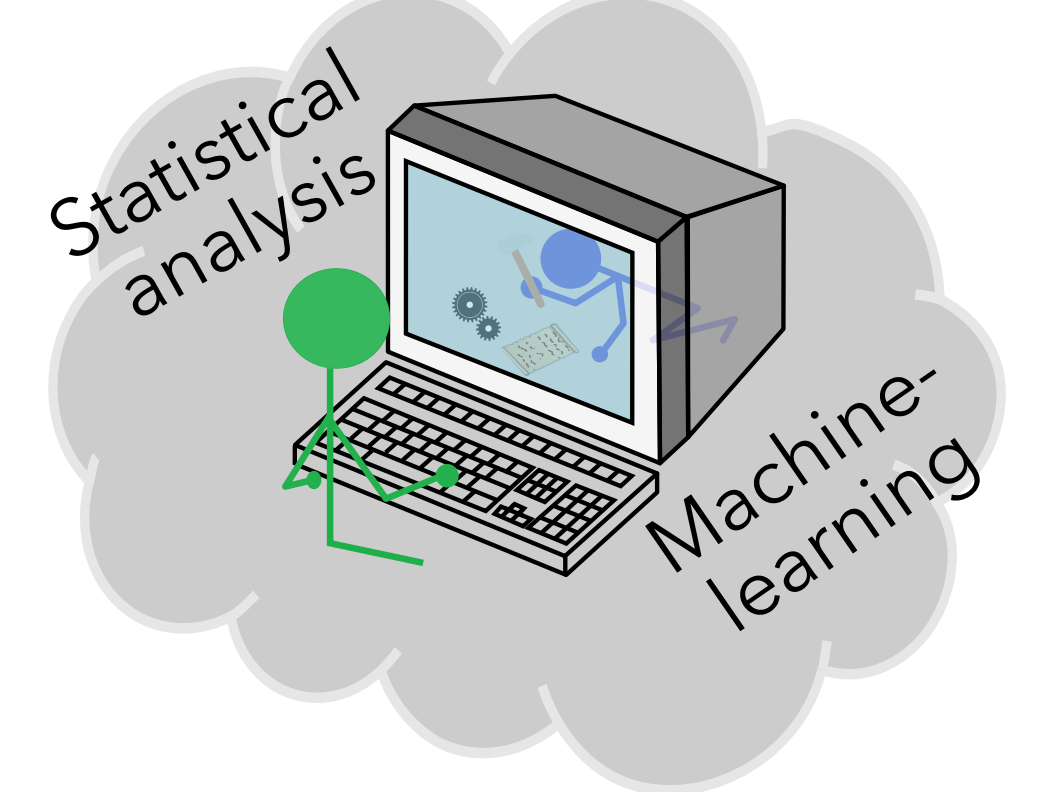
ESI-

Full Scan

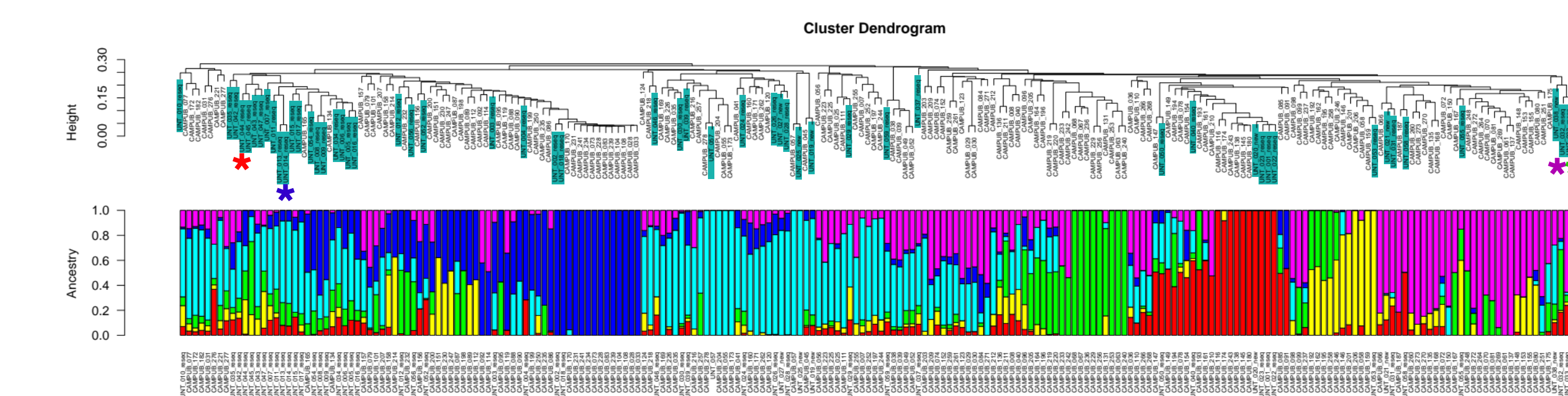
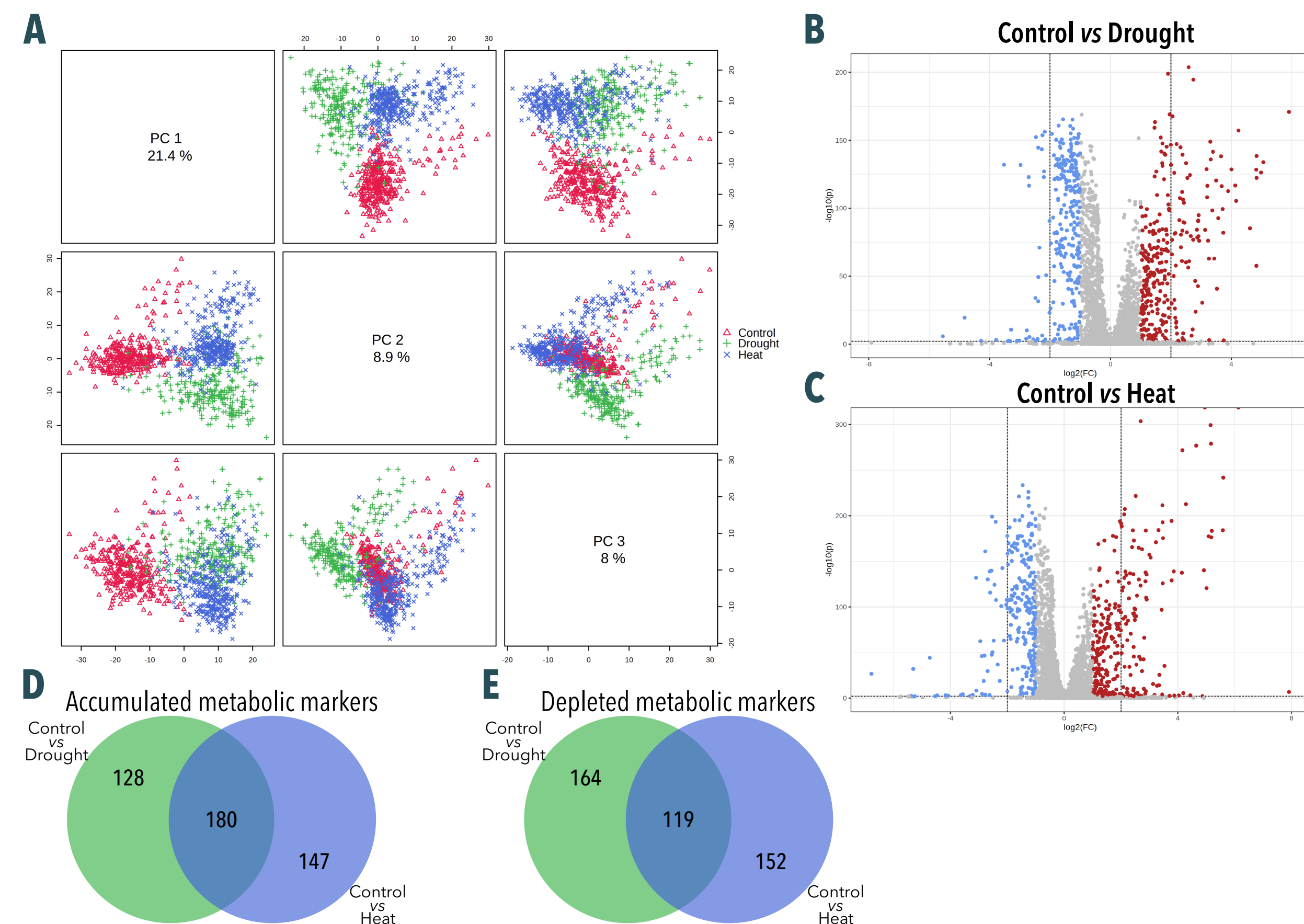
DDA

UHPLC-LTQ-Orbitrap mass spectrometry (thermoScientific)

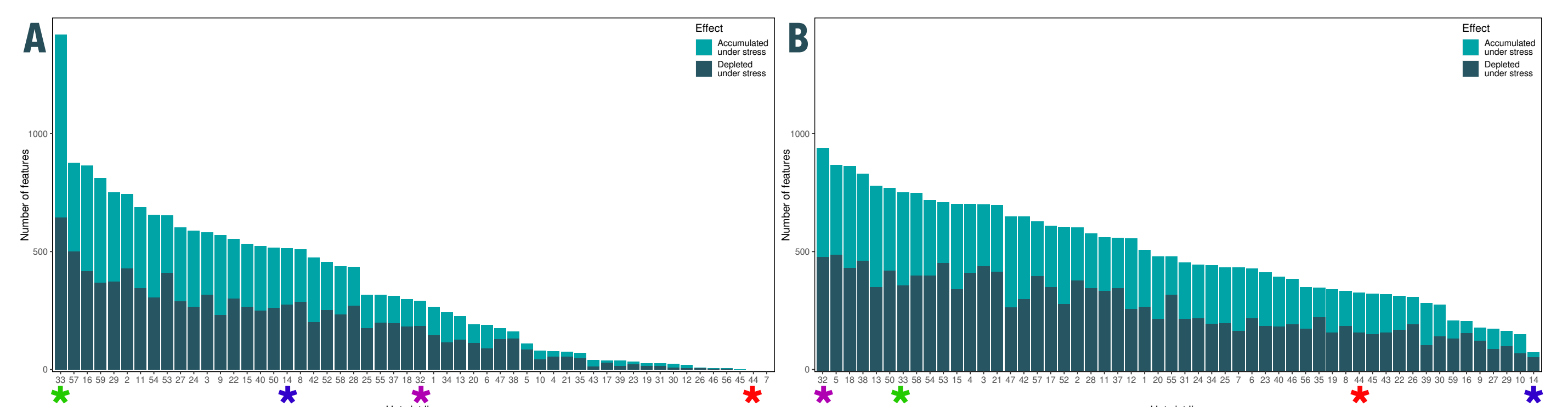
**Untargeted LC-MS:** (Luna et al. 2020)  
- Raw features : 11680  
- After SN > 10 : 10869  
- After CV QC < 30% : 3016 final features  
**Targeted measurements:**  
- 9 major biomass components  
- Starch, citrate, glucose, sucrose, malate, chlorophyll A&B, protein and amino-acid



## Results

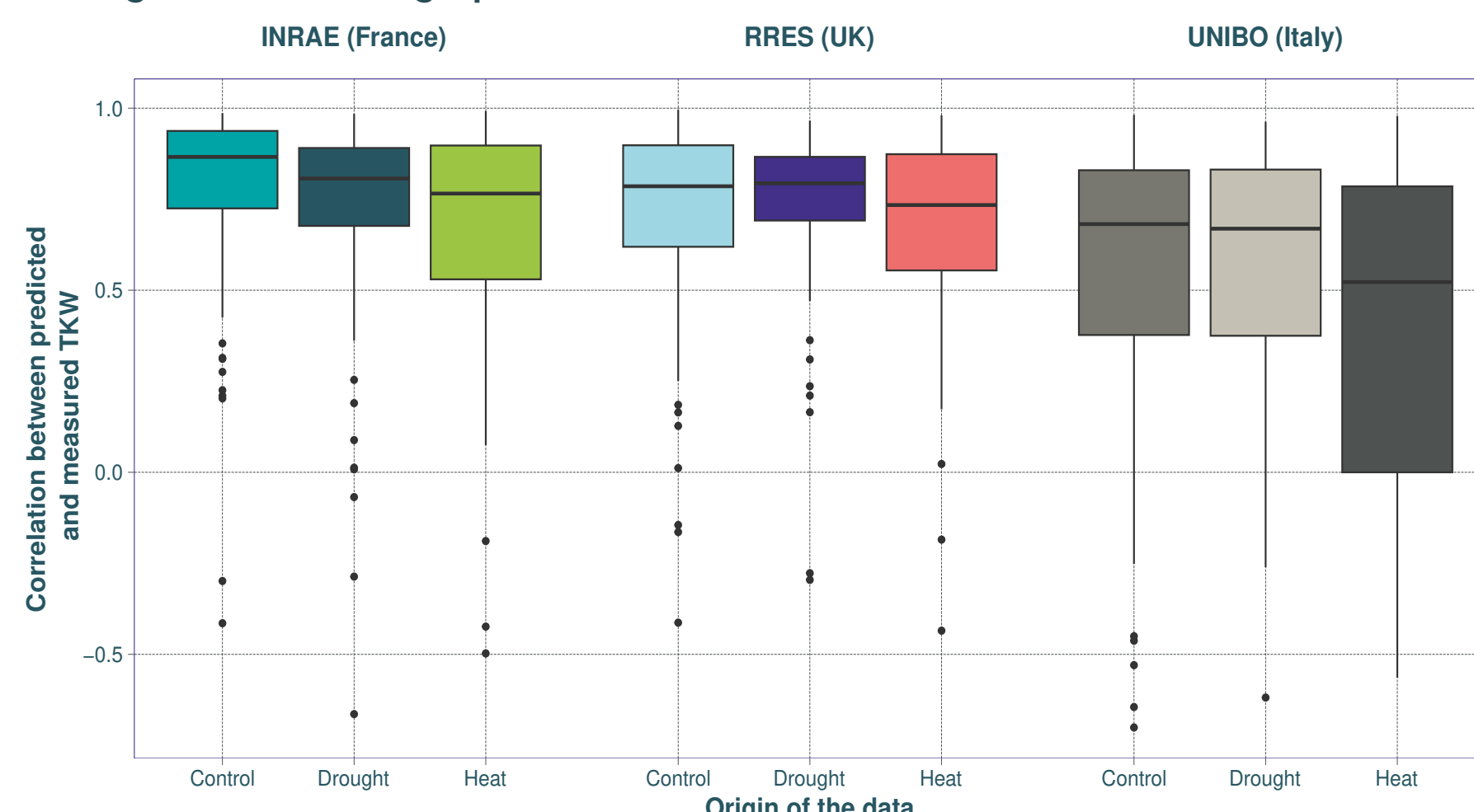


Hierarchical clustering and ADMIXTURE analysis of the Untwist lines and publicly available accessions. IBG distances were hierarchically clustered and 9 sub-populations were assumed.



**Metabolomic responses of individual camelina lines to drought and heat stress.** Volcano plots for each individual *Camelina* line were constructed. The resulting accumulated and depleted markers were stacked for drought (A) and heat stress (B). Colored stars correspond to highlighted lines in the genetic clustering.

**Global metabolic responses of Camelina lines to drought and heat stress**  
(A) PCA score plots of normalised untargeted features (B-C) Volcano plots of metabolic markers that were accumulated (red) or depleted (blue) in response to drought (B) or heat stress (C). (D-E) Venn diagrams showing specific and common metabolic markers identified by volcano plots.



**Predictive metabolomics for Thousand Kernel Weight in three locations using three datasets.** Machine-learning models were built using Ridge, elastic-net and LASSO regression models splitting randomly the dataset into 80/20% one hundred time, using a 3 time 10-fold CV for training. Best models were chosen based on RMSE and correlation between measured and predicted values is displayed.

## Conclusion & Perspectives

- Camelina sativa* lines have very diverse metabolic responses to different stresses
- Stress response from metabolic point of view seems not entirely linked to the genetic background of the lines
- Predictive metabolomics enabled the prediction of an agronomic trait of interest using data coming from greenhouses
- More predictions are ongoing using other machine-learning models and to predict other physiological variables
- Annotation of the different metabolic markers in progress