Analysis and integration of omics data in a context of plant abiotic stress: an example of workflow with the `mixOmics` package

Harold Duruflé

Chargé de Recherches (INRAE, BioForA)

Contribution of an integrative study to the understanding of plant adaptation to their environment: A focus on plant cell walls.

Supervisors:

Pr. Christophe DUNAND

Pr. Philippe BESSE & Sebastien DÉJEAN

OXFORD

Briefings in Bioinformatics, 00(00), 2020, 1–13

doi: 10.1093/bib/bbaa166

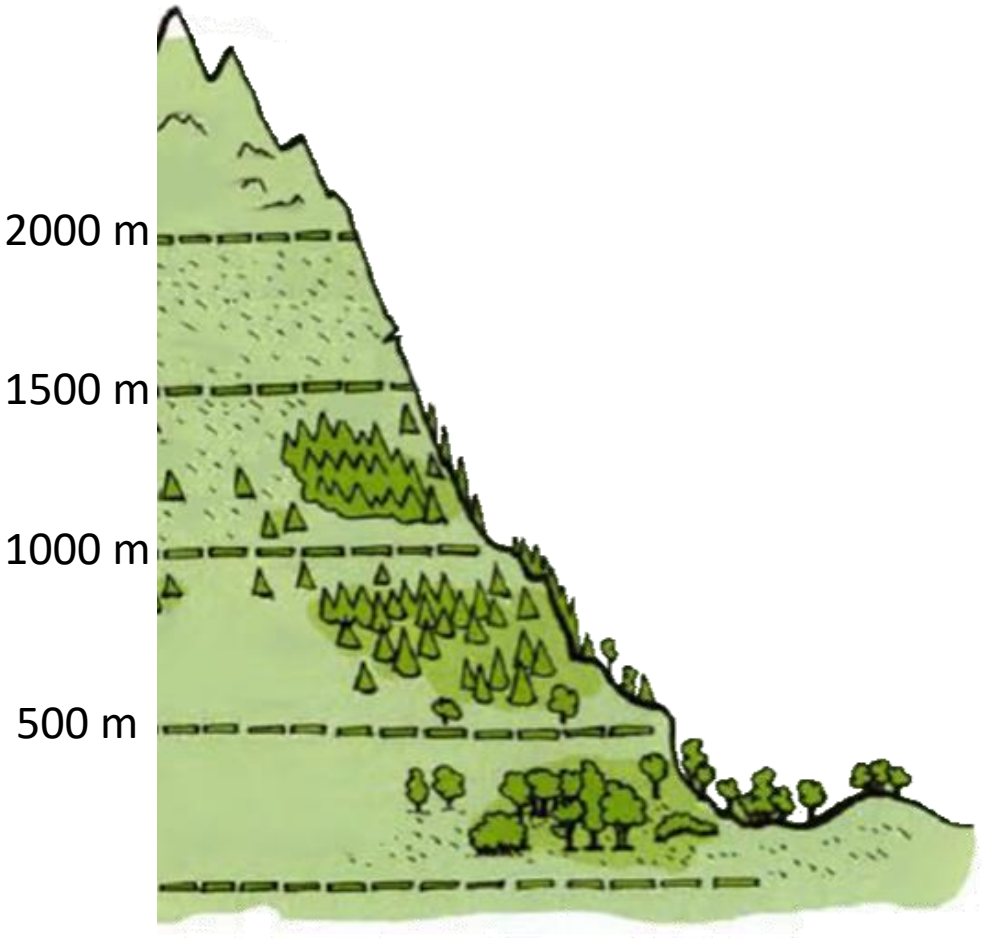
Problem Solving Protocol

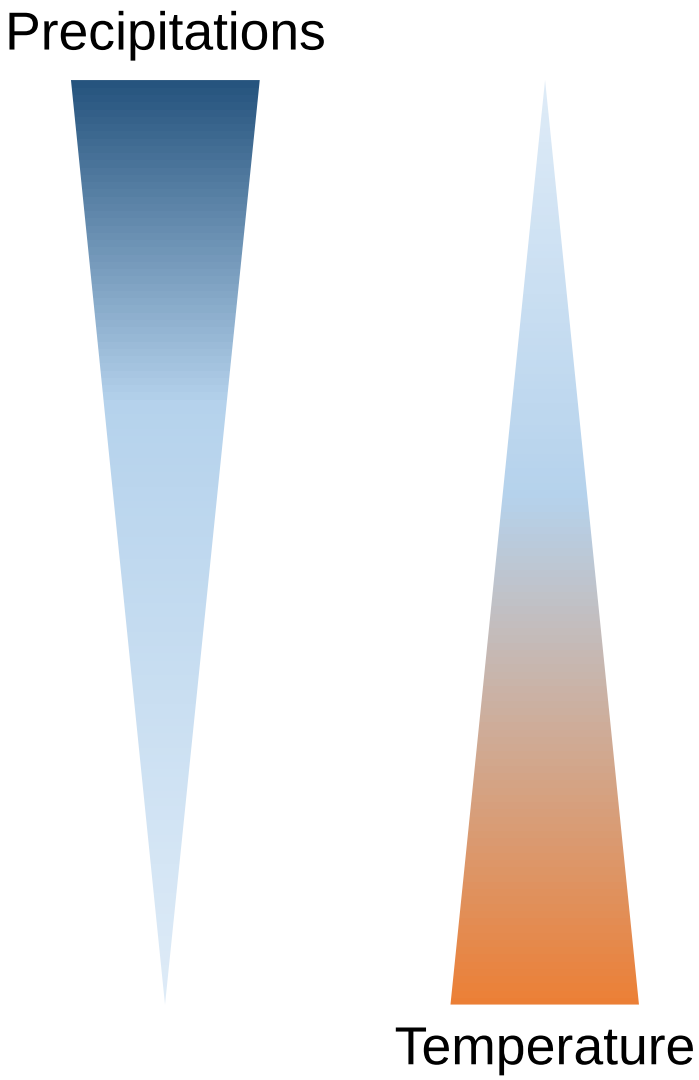
A powerful framework for an integrative study with heterogeneous omics data: from univariate statistics to multi-block analysis

Harold Duruflé, Merwann Selmani, Philippe Ranocha, Elisabeth Jamet, Christophe Dunand and Sébastien Déjean

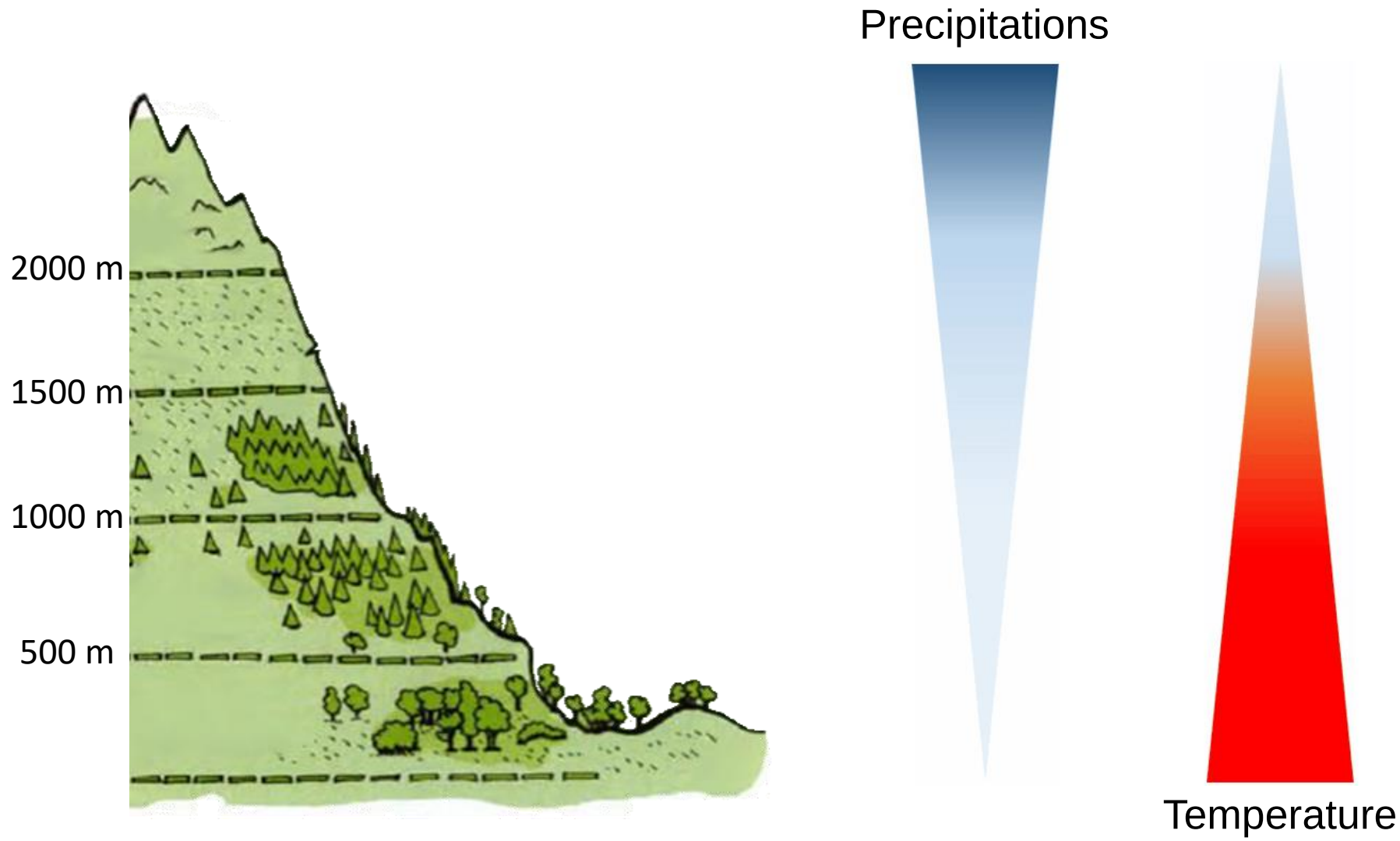
Objectives & strategies

Mountains as areas of study



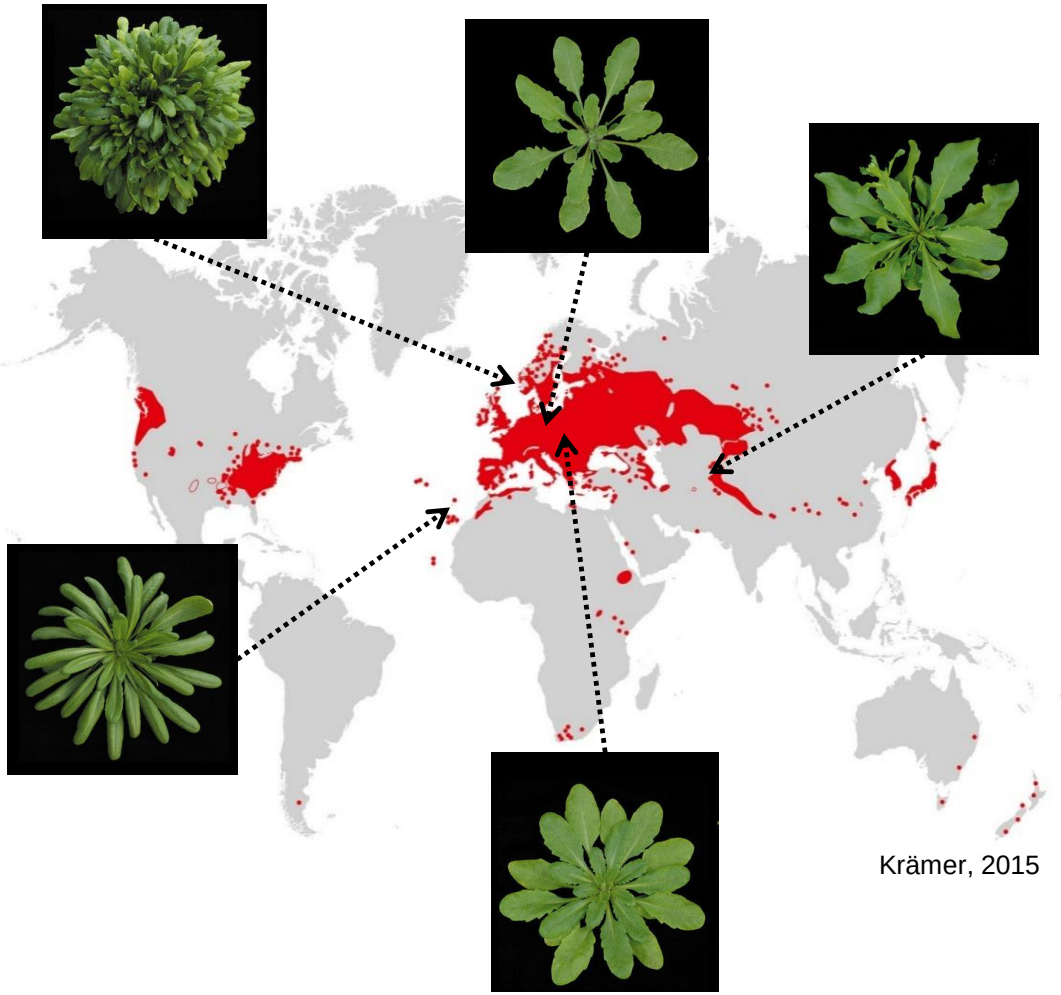


Mountains as areas of study



The model plant: *Arabidopsis thaliana*

PlantScreen Compact System

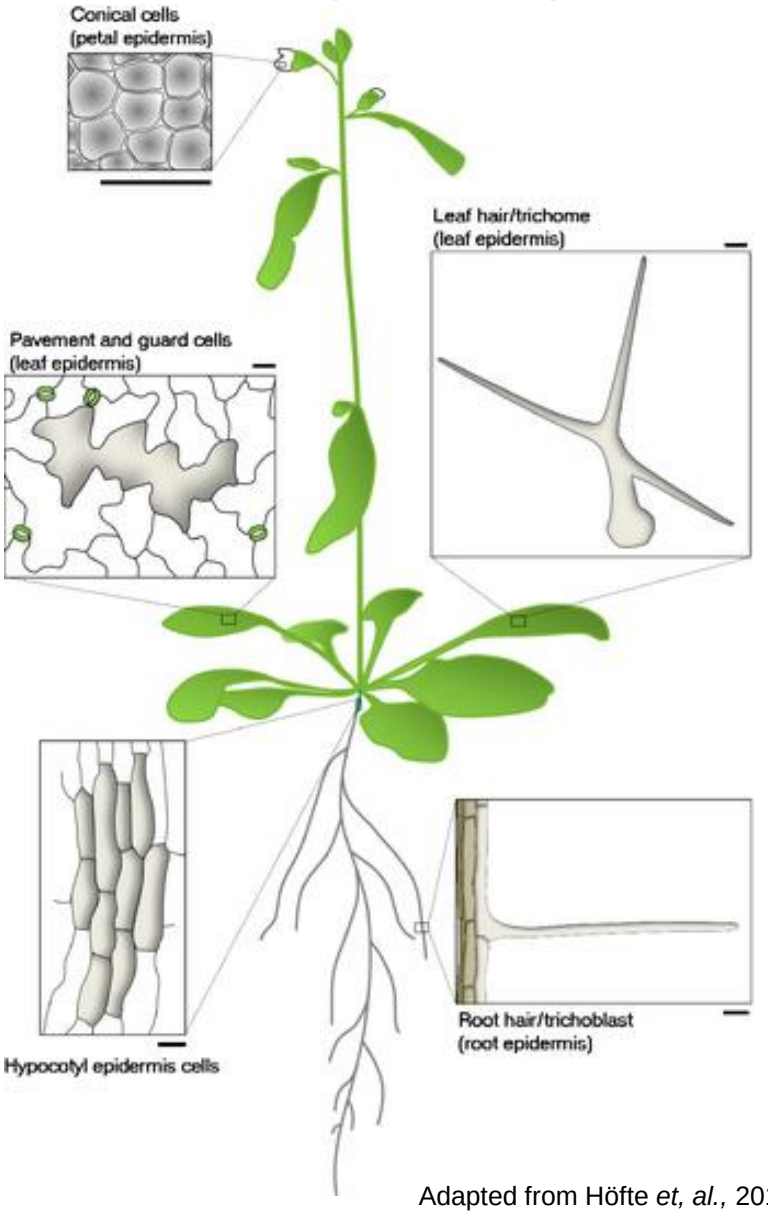


Krämer, 2015

Adapted from Savolainen & Lascoux *2016*

The cell wall: the plant skeleton

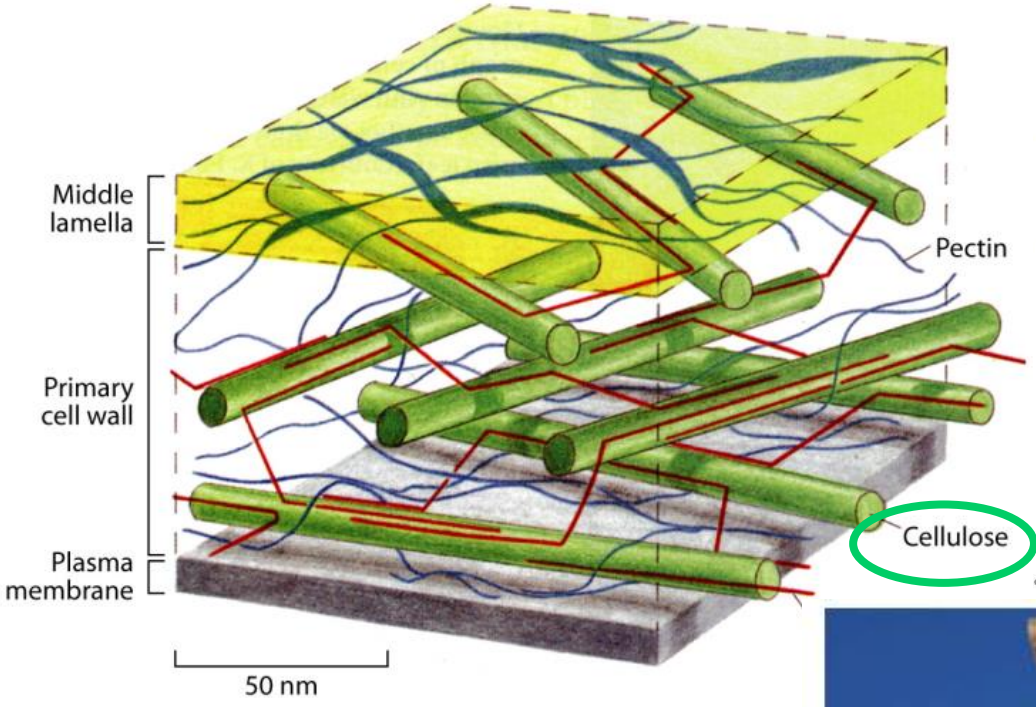
- The cell wall contributes to the cell and the plant shapes
- Different functions
- Different compositions

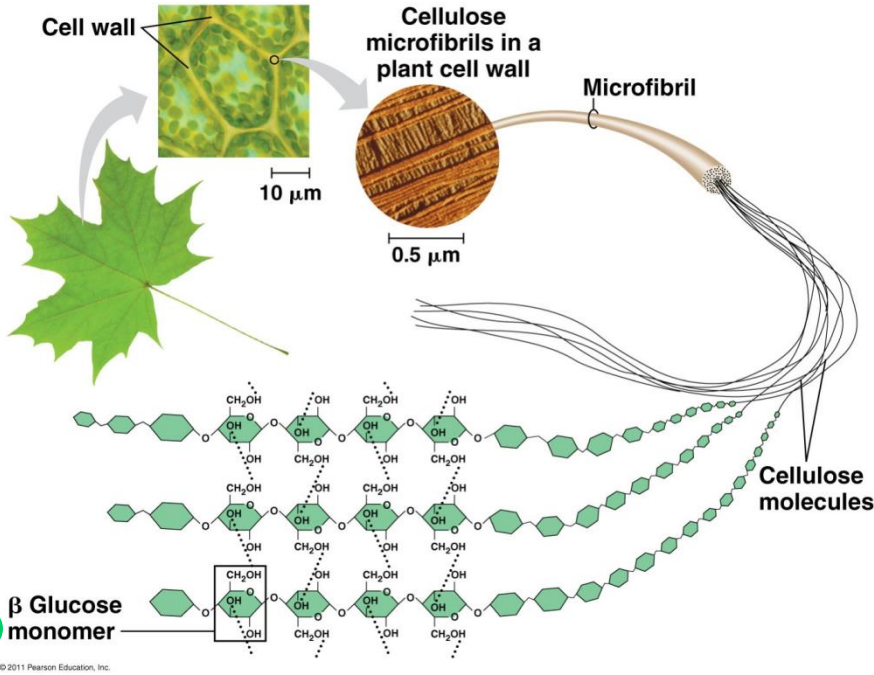


Adapted from Höfte *et, al.,* 2017 & Braidwood *et al.* 2013

The cell wall: main constituents

Cell wall polysaccharides

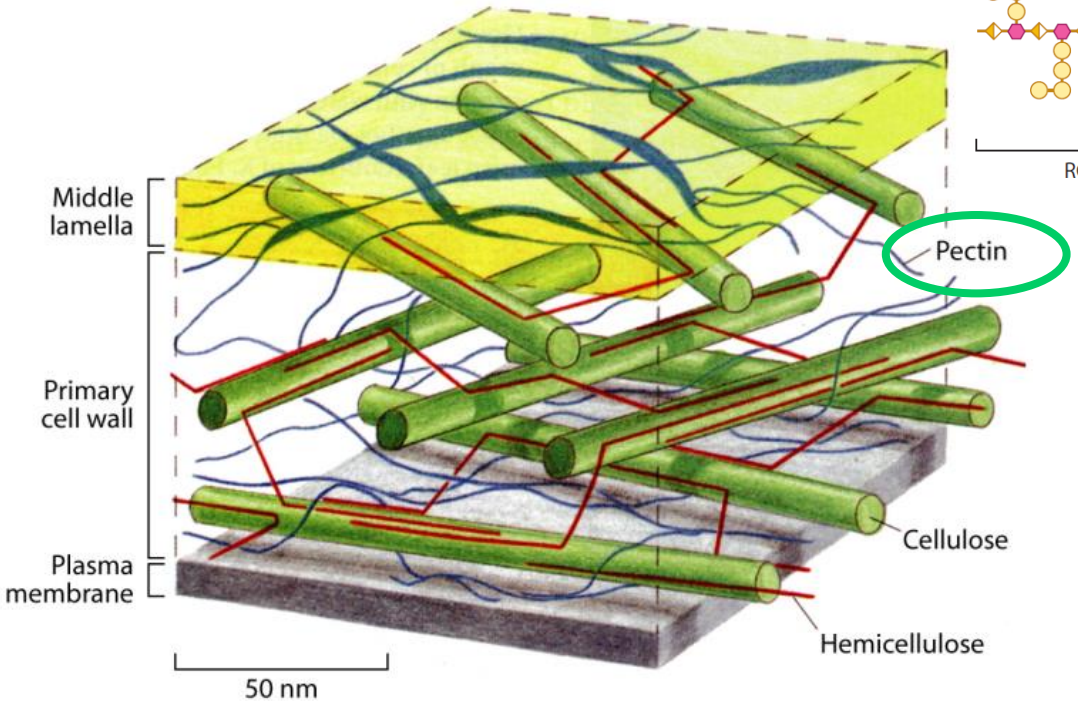


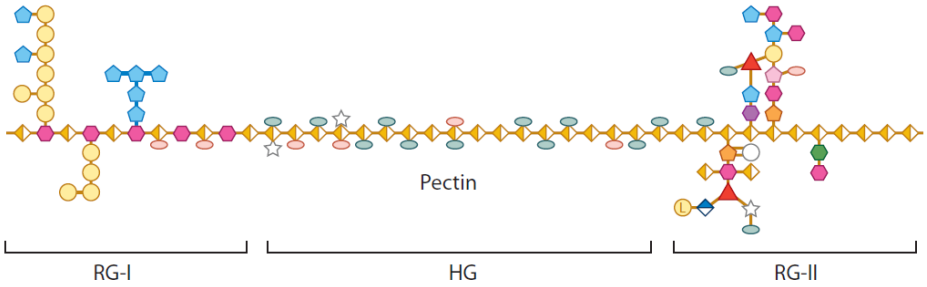


Scheller a

The cell wall: main constituents

Cell wall polysaccharides

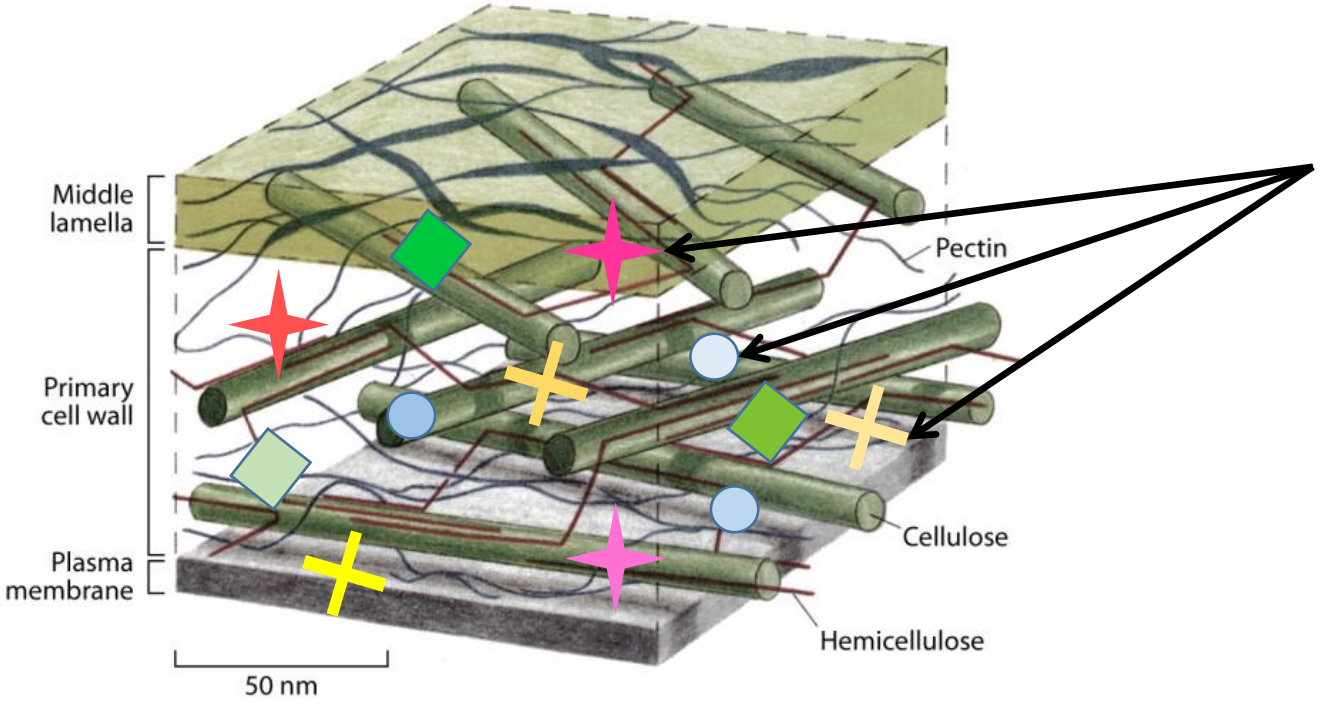




Scheller and Ulvskov 2010

The cell wall: main constituents

Cell wall polysaccharides



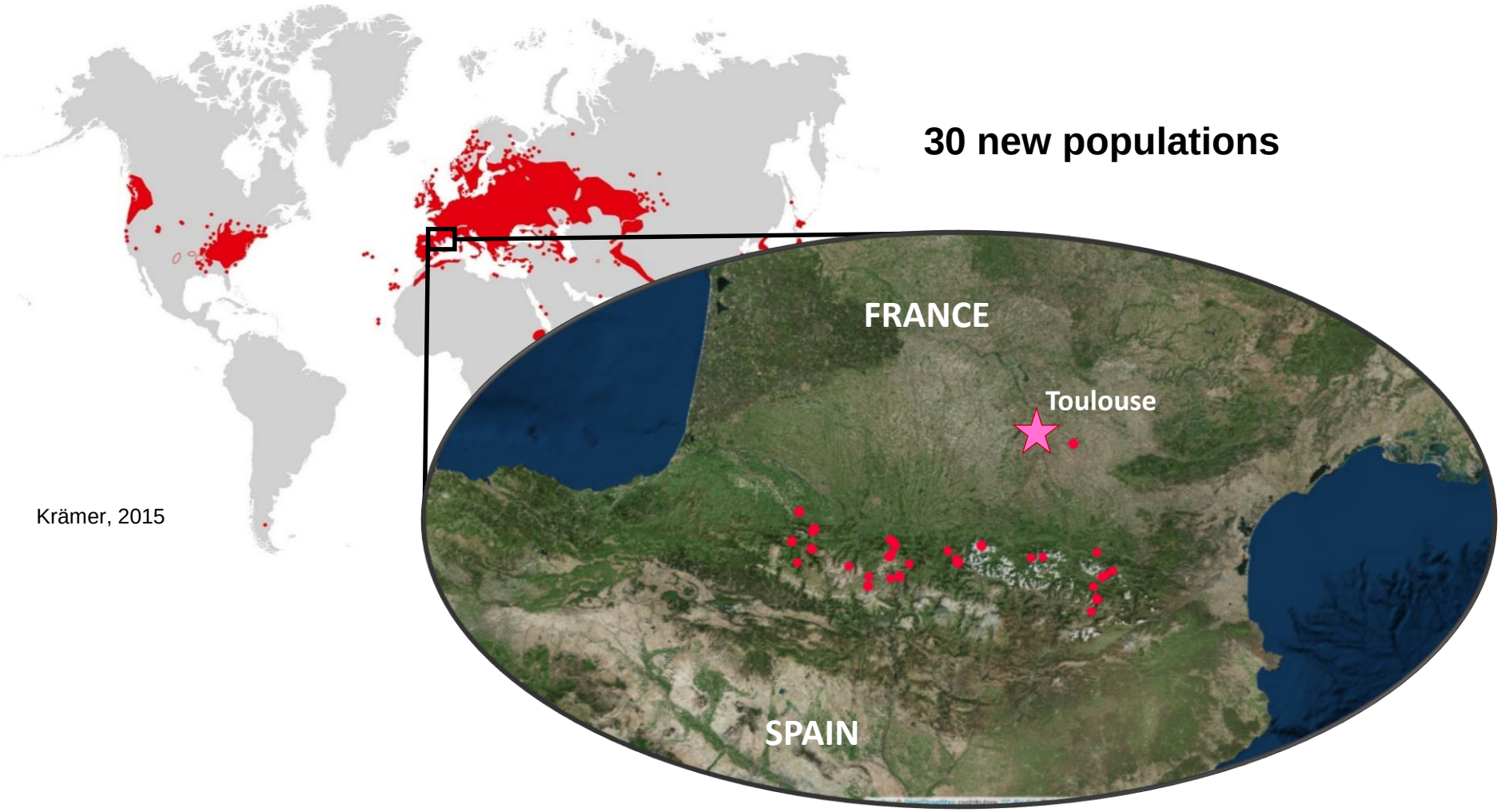
Cell wall proteins

Scheller and Ulvskov 2010

- Interlaced networks that can be reorganized at any time
  - Dynamic and plastic
- Proteins contribute to assemble and remodel the cell wall
  - Different functional class

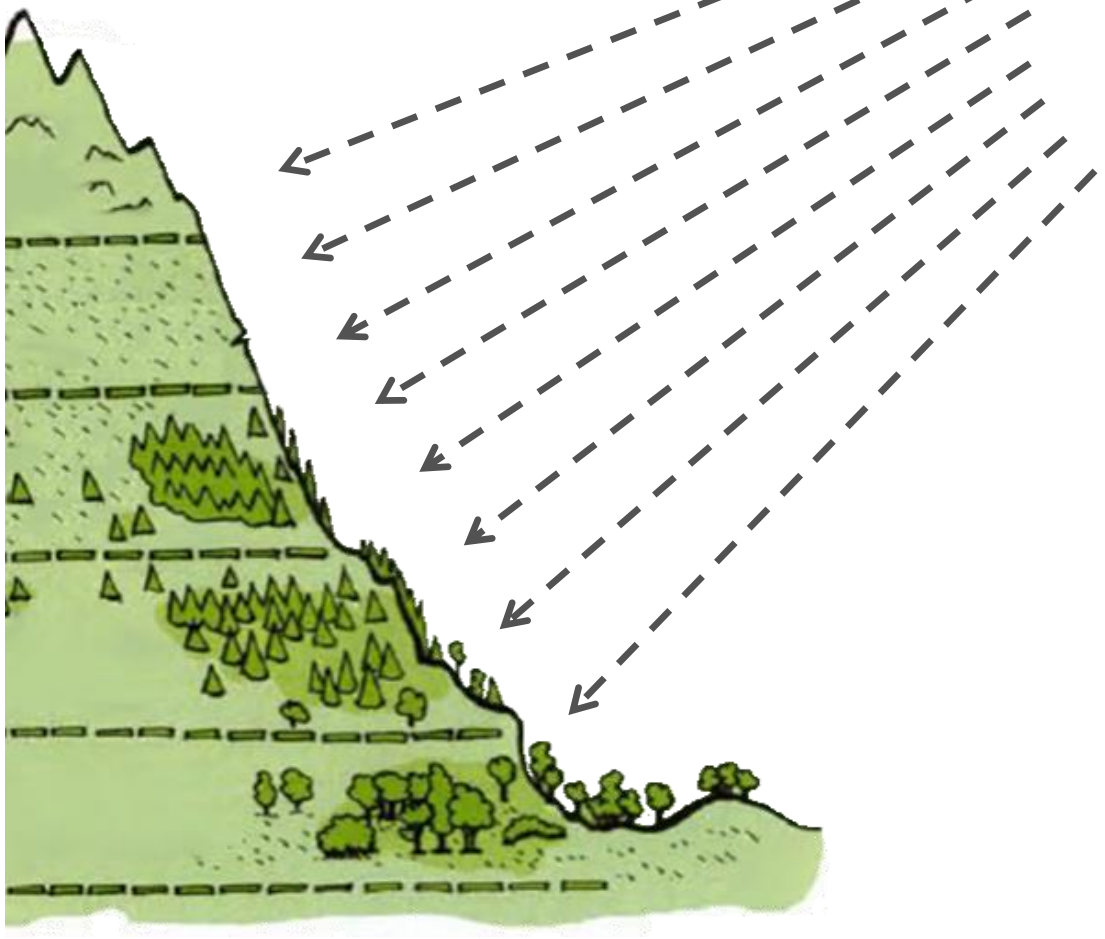
Study of natural populations

Highlighting the natural diversity of *A. thaliana* populations in the Pyrenees.



Duruflé, *et. al.,* Front Plant Sci (2019)

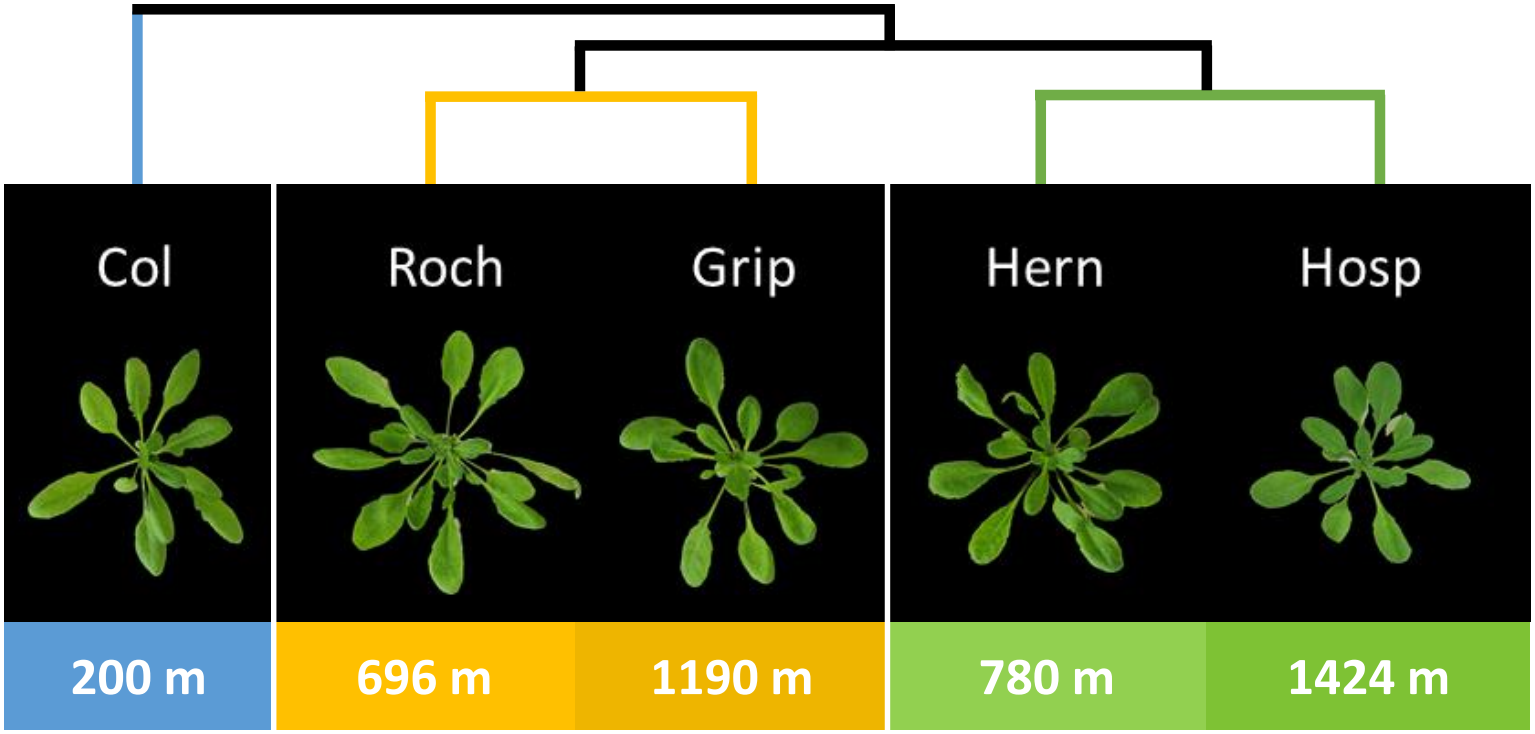
Natural populations from contrasted growth conditions



- Different altitude levels
- Different environments

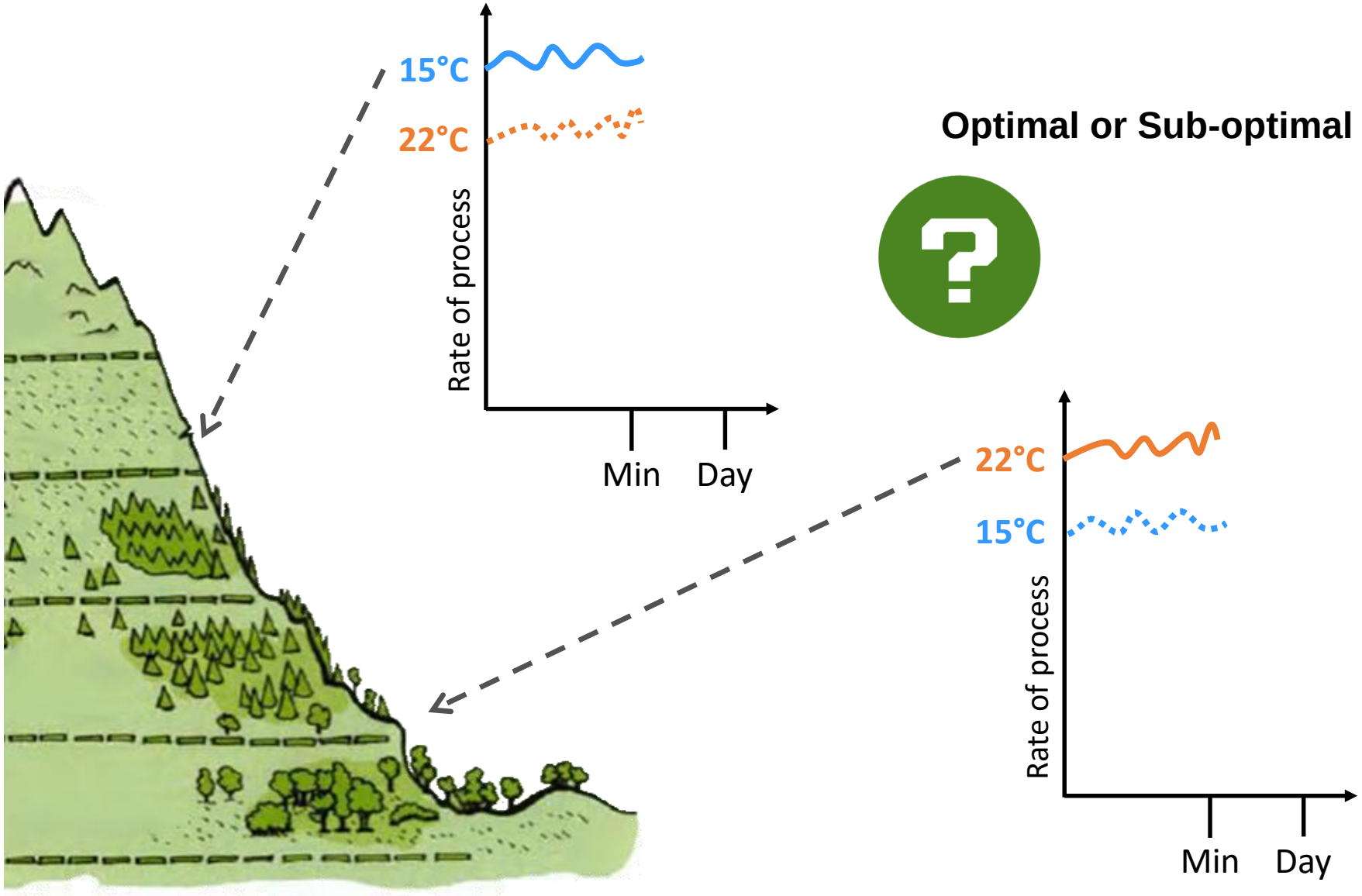
Adaptation ?

Environmental adaptation of *A. thaliana*



- 3 genetic clusters
- 2 contrasted altitudes

Choice of the temperatures growth conditions



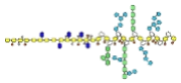
A system biology approach

Two organs

Omics analysis

Phenomics (Macro- and micro- phenotypic analyses)

- 5 and 4 phenotype on the rosette and the floral stems



Metabolomics

- 6 cell wall polysaccharides

Cell wall proteomics

- 364 and 414 cell wall proteins (CWPs) on rosette and floral stems

Transcriptomics

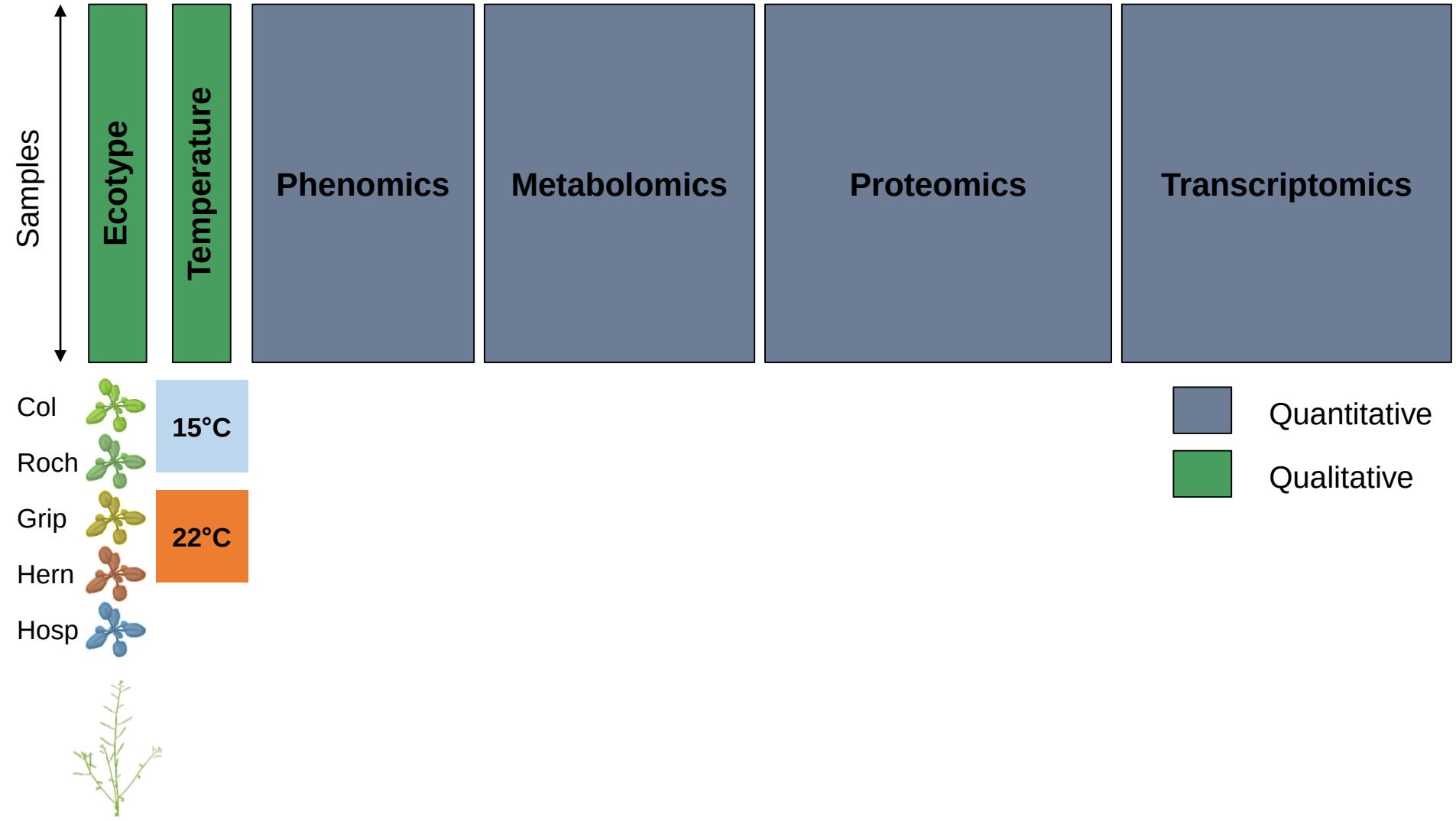
- 19,763 and 22,570 transcripts on rosette and floral stems

R Datasets Package "`WallOmicsData`"

Soon available (CRAN) for users needing benchmarking

- 3 biological replicates
- 20 plants per sample

A system biology approach: principle of blocks



mixOmics workflow

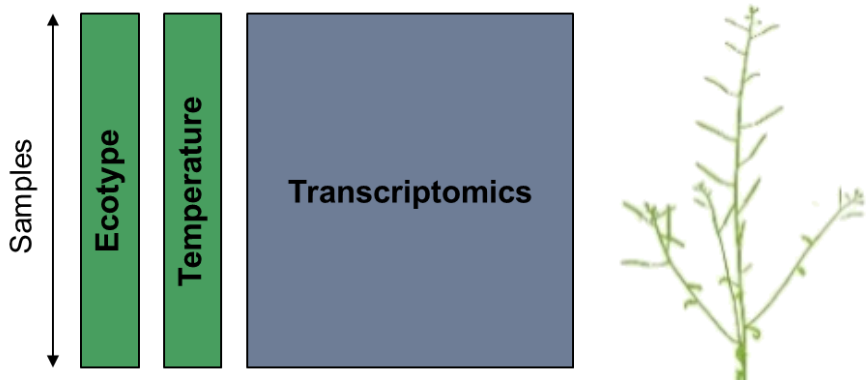
1) Ask a biological question

2) Run a method: `pca(), pls(), spls(), plsda(), block.pls(),…`

3) Represent individuals: `plotIndiv()`

4) Represent variables: `plotVar(), plotLoadings(), cim(),…`

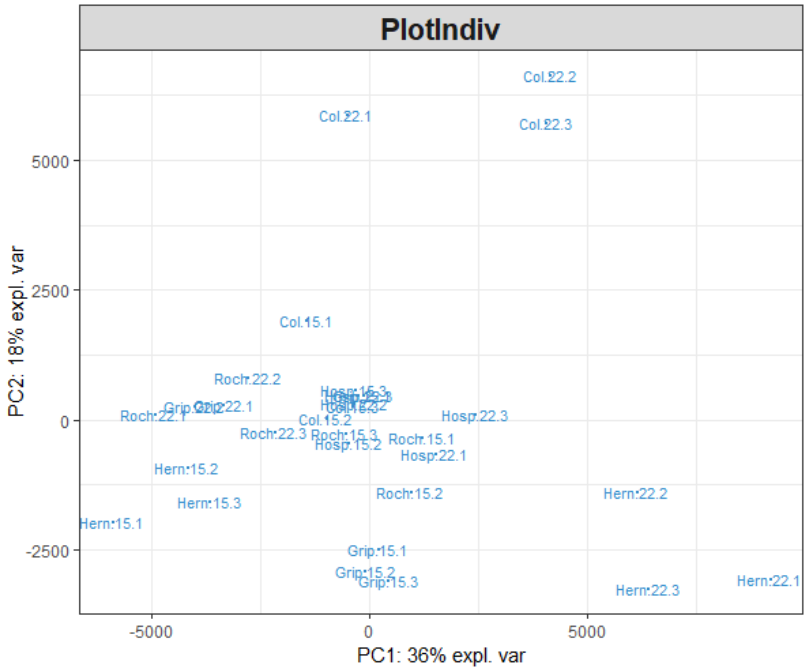
Case study
focused on the floral stem

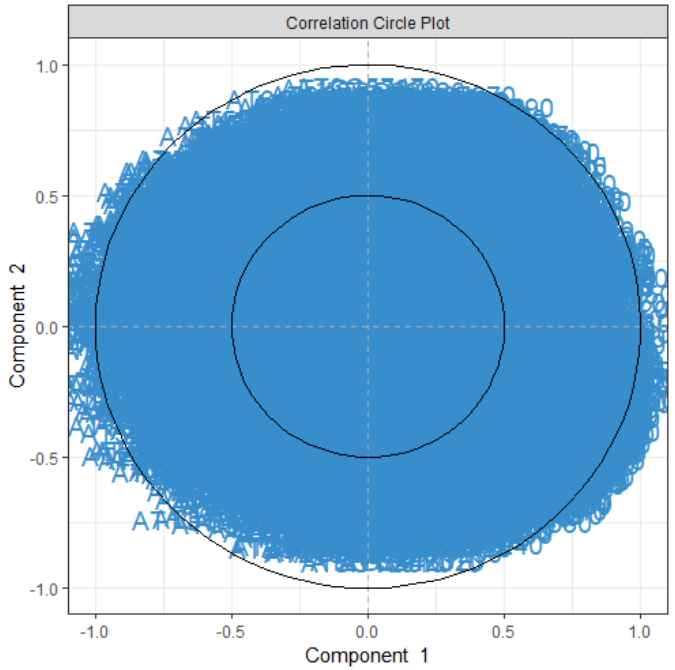


Can we observe on the transcriptomics data, with no prior, the effect of different environmental growth conditions or different ecotypes?

→ Perform Principal Component Analysis

```
Result_PCA_stems_transcriptomics <- pca(Transcriptomics_Stems)
plotIndiv(Result_PCA_stems_transcriptomics)
plotVar(Result_PCA_stems_transcriptomics)
```

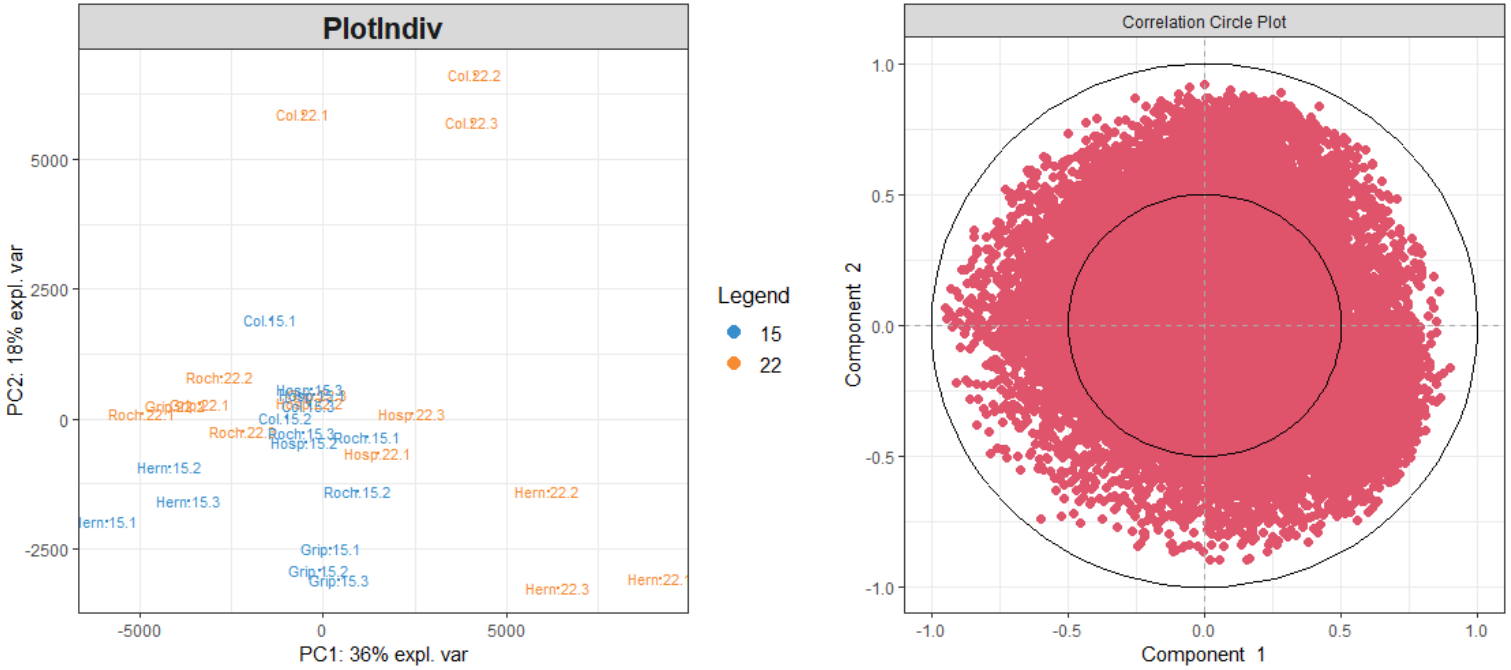




Can we observe on the transcriptomics data, with no prior, the effect of different environmental growth conditions or different ecotypes?

→ Perform Principal Component Analysis

```
plotIndiv(Result_PCA_stems_transcriptomics, group = Temperature, legend = TRUE)
plotVar(Result_PCA_stems_transcriptomics, var.names = FALSE, pch = 16, cex = 2, col = 2)
```

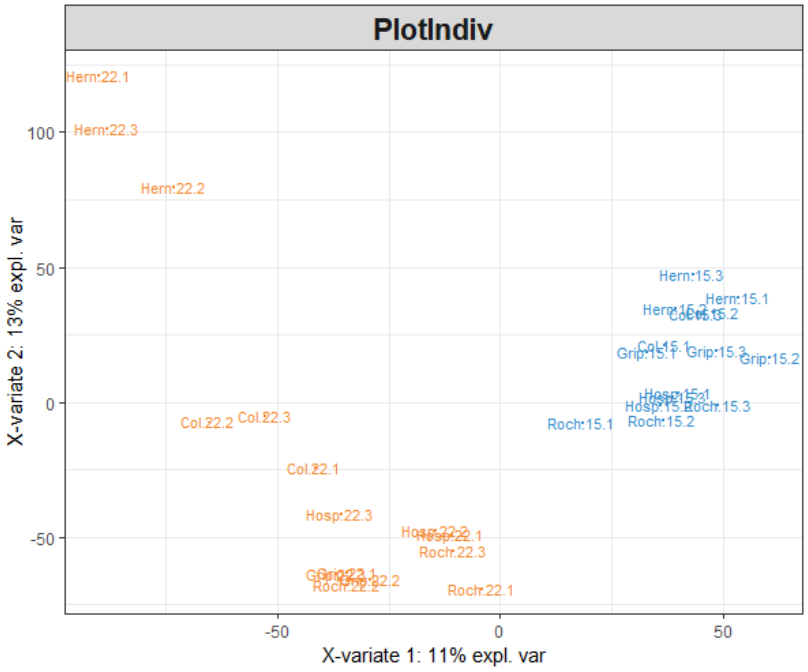


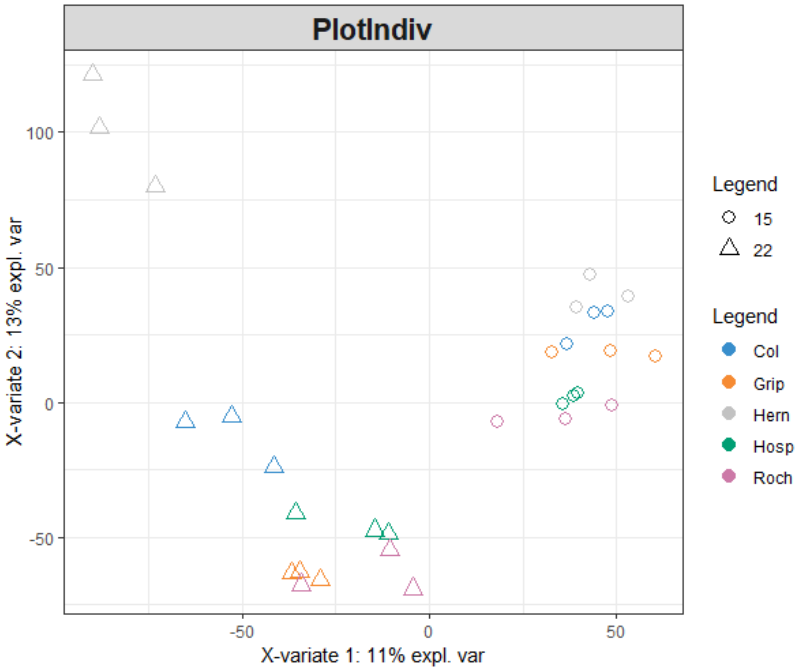
Can we observe a global effect of temperature on the different ecotypes according to their transcriptomics profiles?

→ Perform Projection to Latent Structures - Discriminant Analysis

```
Result_PLSDA_stems_transcriptomics_temperature <- plsda(X=Transcriptomics_Stems,Y= Temperature)
plotVar(Result_PLSDA_stems_transcriptomics_temperature)
plotIndiv(Result_PLSDA_stems_transcriptomics_temperature)
```

```
plotIndiv(Result_PLSDA_stems_transcriptomics_temperature,
          ind.names = FALSE, pch = Temperature, group = Ecotype, legend = TRUE)
```

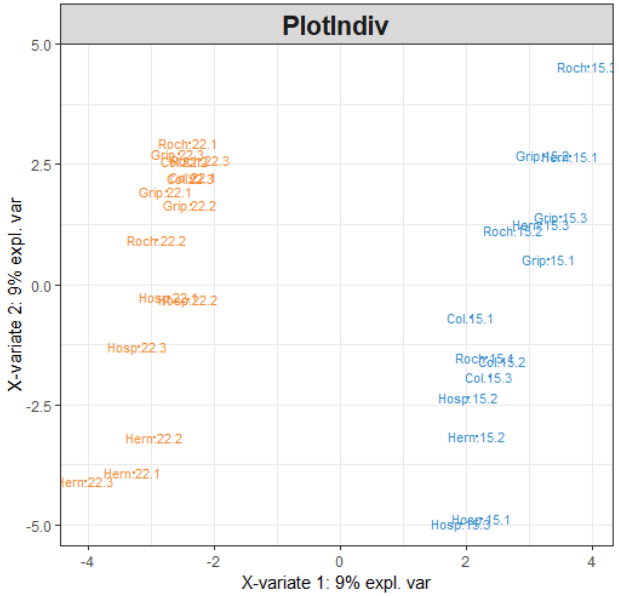


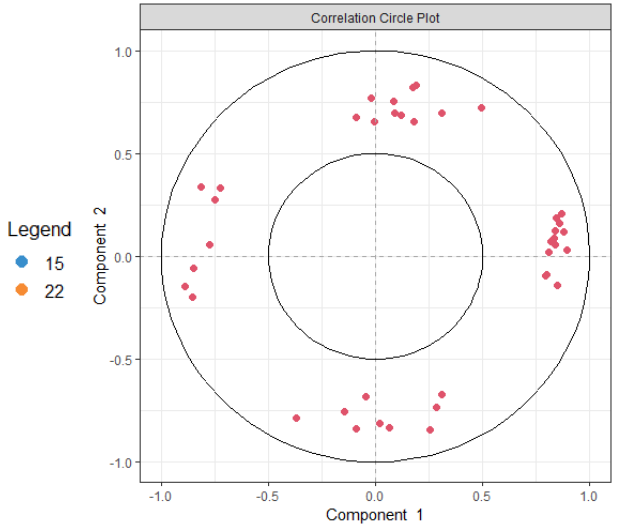


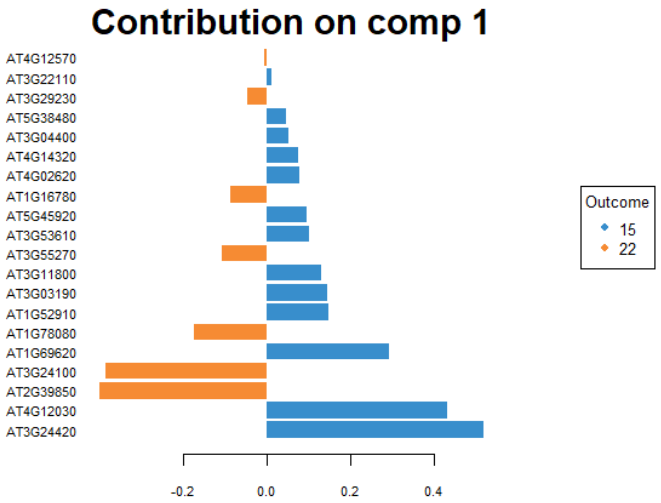
How to know the best candidate genes for the global effect of temperature?

→ Perform Sparse Projection to Latent Structures - Discriminant Analysis

```
Result_sPLSDA_stems_transcriptomics_temperature <- splsda(X = Transcriptomics_Stems,
                                                          Y = Temperature, keepX = c(20,20))
plotIndiv(Result_sPLSDA_stems_transcriptomics_temperature)
plotVar(Result_sPLSDA_stems_transcriptomics_temperature, var.names = FALSE, pch = 16, cex = 2, col = 2)
plotLoadings(Result_sPLSDA_stems_transcriptomics_temperature, contrib = 'max', method = 'mean')
```

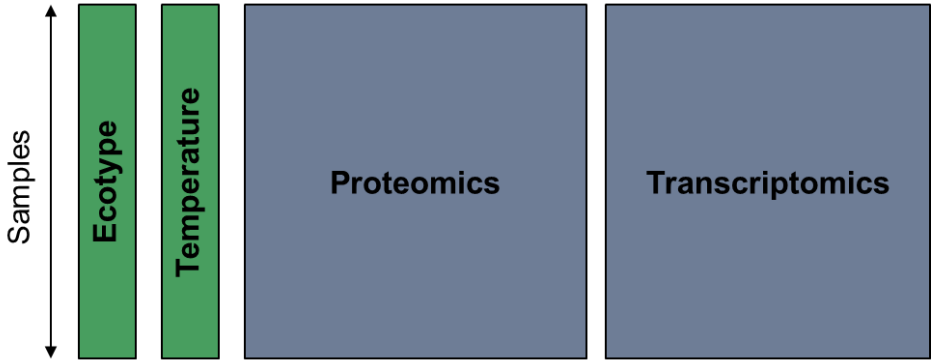







```
Selected_transcripts_temperature <-selectVar(Result_sPLSDA_stems_transcriptomics_temperature,comp=1)
```

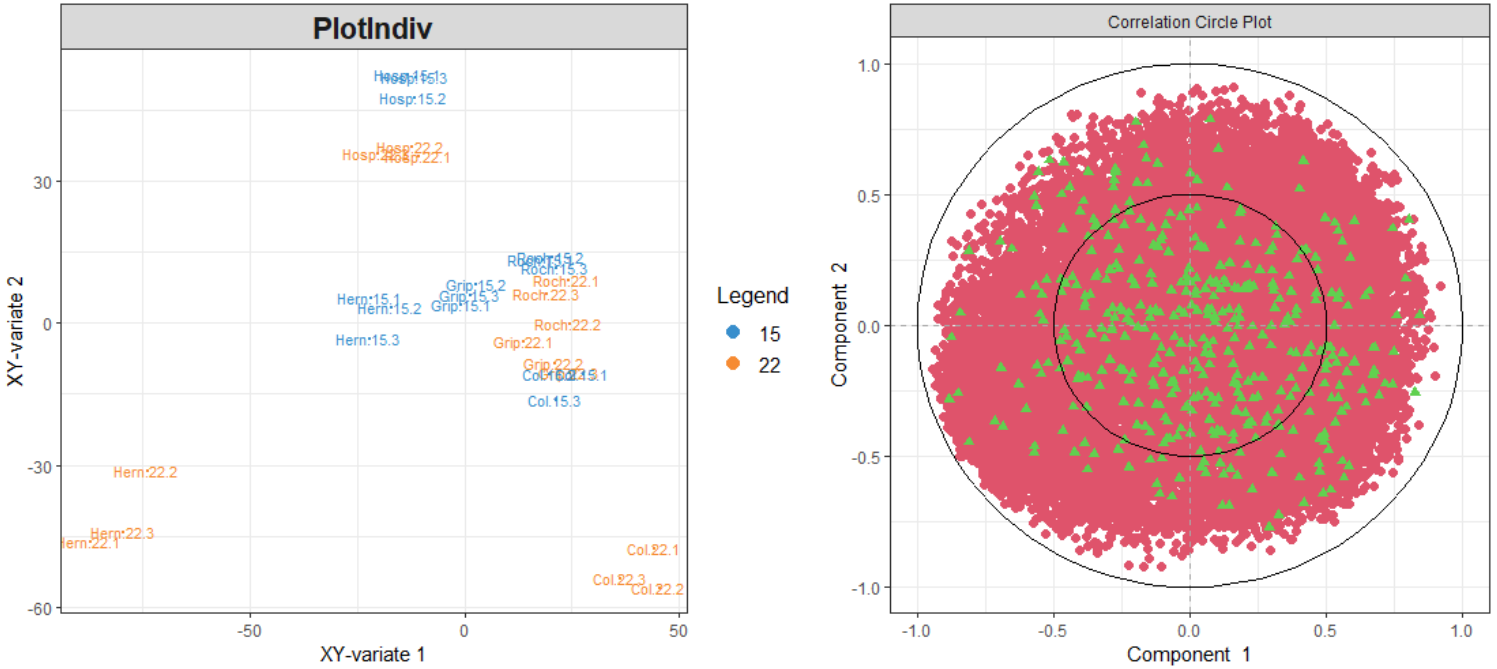
Horizontal integration



Can we highlight relationships between cell wall proteins and transcripts in stems?

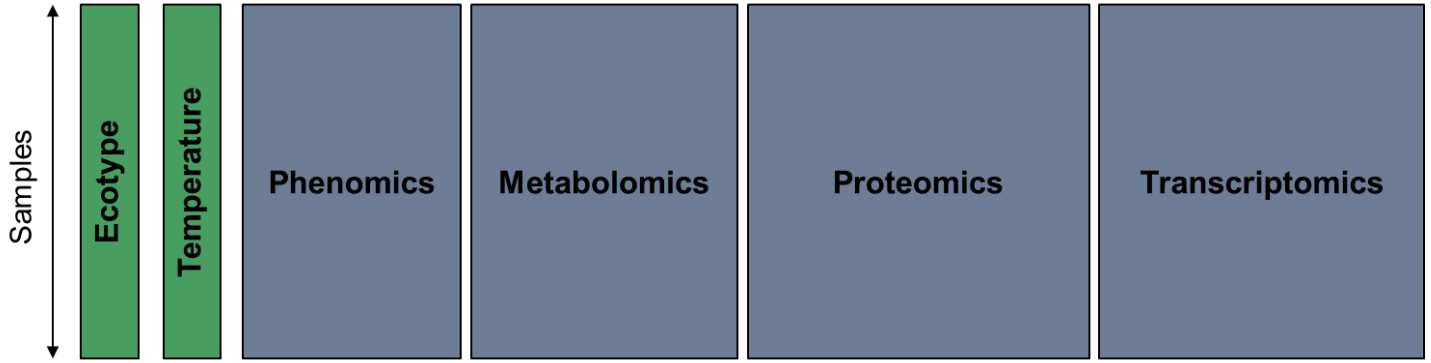
→ Perform Projection to Latent Structures

```
Result_PLS_stems_transcriptomics <- pls(X = Transcriptomics_Stems, Y = Proteomics_Stems_CW)
plotIndiv(Result_PLS_stems_transcriptomics, rep.space = "XY-variate", group = Temperature,legend = TRUE)
plotVar(Result_PLS_stems_transcriptomics, var.names = c(FALSE,FALSE), pch = c(16,17),
        cex = c(2,2), col = c(2,3))
```




```
Result_sPLS_stems_transcriptomics <- spls(X = Transcriptomics_Stems, Y = Proteomics_Stems_CW,
                                          keepX = c(10,10), keepY = c(10,10))
plotLoadings(Result_sPLS_stems_transcriptomics, contrib = 'max')
```

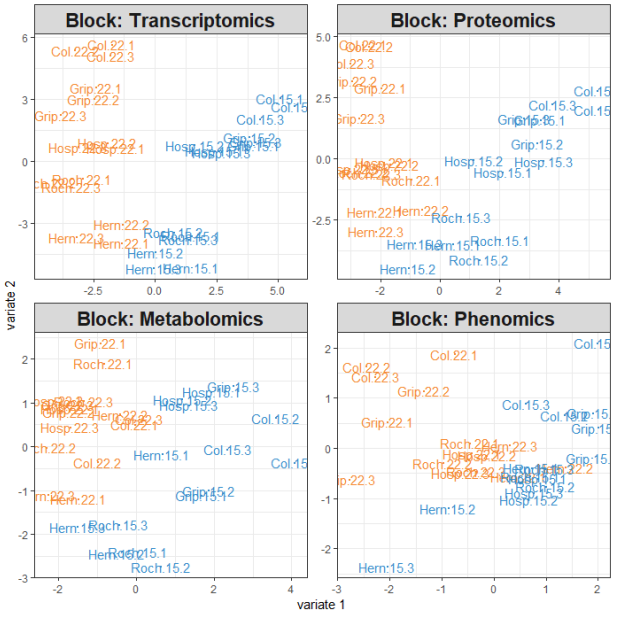
Horizontal integration

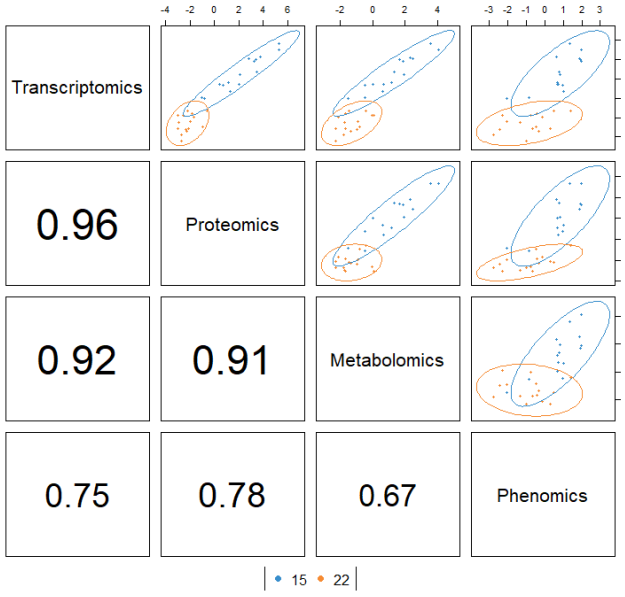


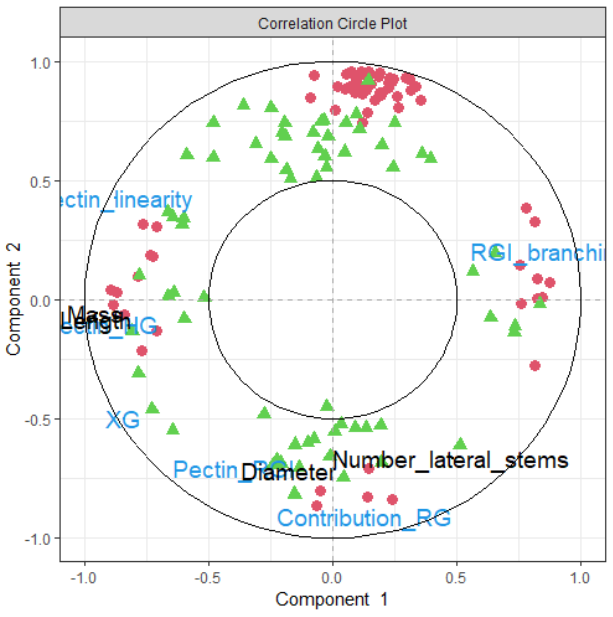
Can we determine a multi-omics signature to classify ecotypes?

→ Perform multi-block Sparse Projection to Latent Structure - Discriminant Analysis (DIABLO)

```
Data_Stems <- list(Transcriptomics = Transcriptomics_Stems,
                   Proteomics = Proteomics_Stems_CW,
                   Metabolomics = Metabolomics_Stems,
                   Phenomics = Phenomics_Stems)

Keepdata_Data_Stems <- list(Transcriptomics = c(20, 20),
                            Proteomics = c(20, 20),
                            Metabolomics = c(6, 6),
                            Phenomics = c(4, 4))

Result_DIABLO_stems <- block.splsda(X = Data_Stems, Y = Temperature, keepX = Keepdata_Data_Stems)
plotIndiv(Result_DIABLO_stems, cex=4)
plotDiablo(Result_DIABLO_stems)
plotVar(Result_DIABLO_stems, var.names = c(FALSE, FALSE, TRUE, TRUE), pch = c(16, 17, NA, NA),
        cex = c(3, 3, 5, 5), col = c(2, 3, 7, 1))
```



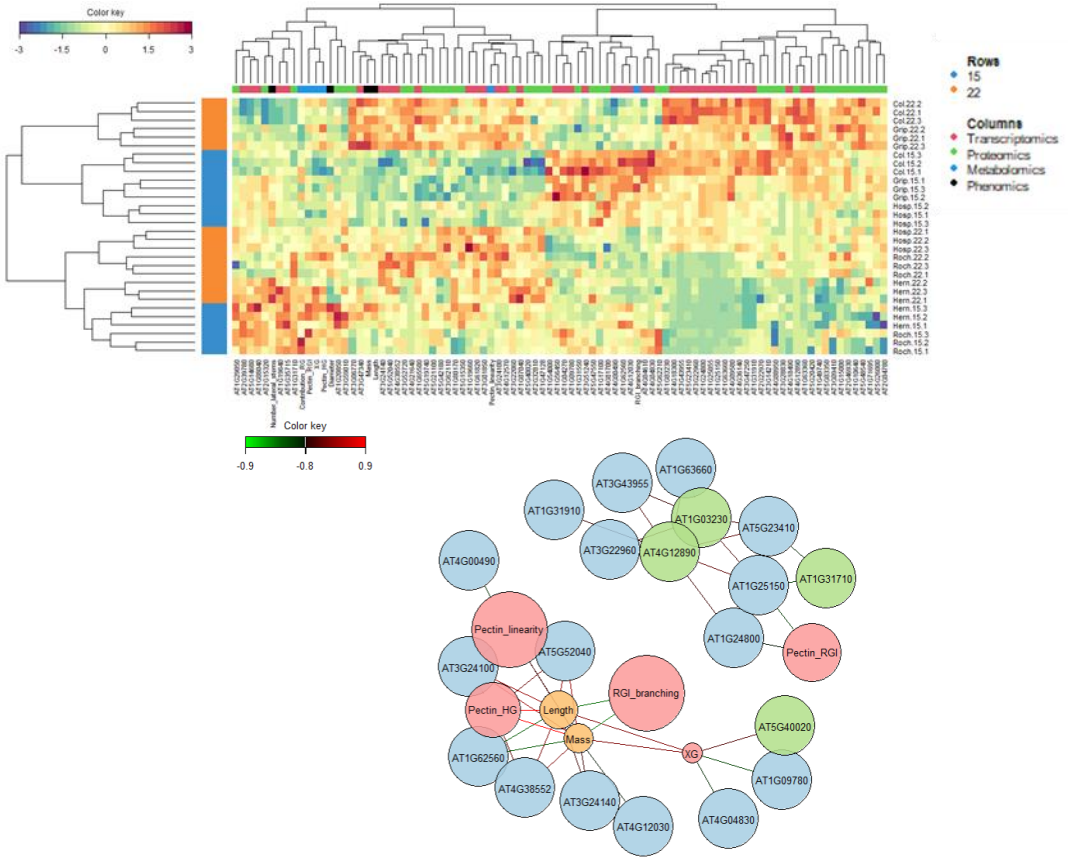


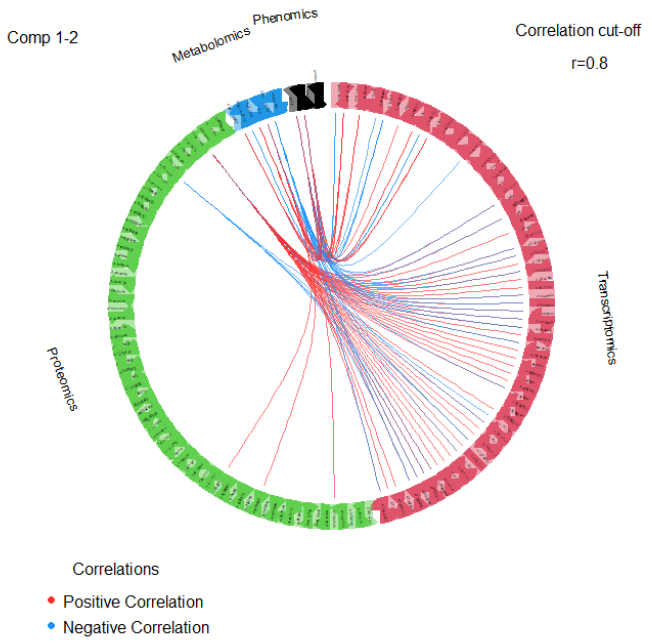


Can we determine a multi-omics signature to classify ecotypes?

→ Perform multi-block Sparse Projection to Latent Structure - Discriminant Analysis (DIABLO)

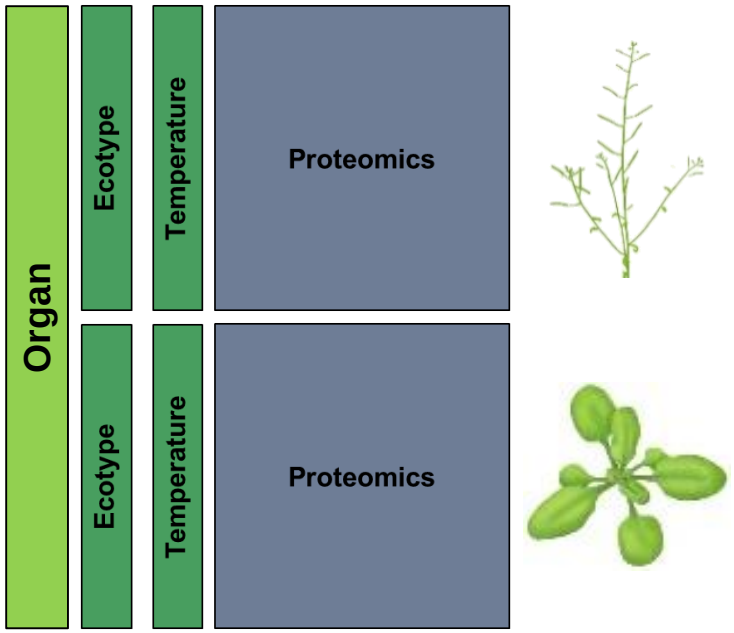
```
cimDiablo(Result_DIABLO_stems, margins = c(8,10), size.legend = 0.5, color.blocks = c(2, 3, 7, 1),
          legend.position = "topright")

circosPlot(Result_DIABLO_stems, cutoff = 0.8, size.legend = 1, size.variables = 0.2, size.labels = 1,
           color.blocks = c(2, 3, 7, 1))
network(Result_DIABLO_stems, blocks = c(1,2,3,4),block.var.names = c(T,T,T,T),cutoff = 0.8)
```






```
Candidates_Stems <- selectVar(Result_DIABLO_stems)
Transcripts_Candidates <- Candidates_Stems$Transcriptomics$name
```

Vertical integration

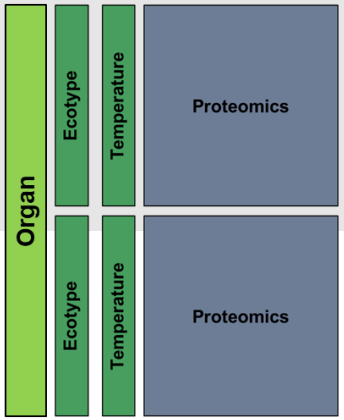


Can we identify on the proteomics data behaviors that do not depend on the organ?

→ Perform Multivariate INTegrative Method (MINT)

Need to format the data to assemble the proteomic data of these two organs

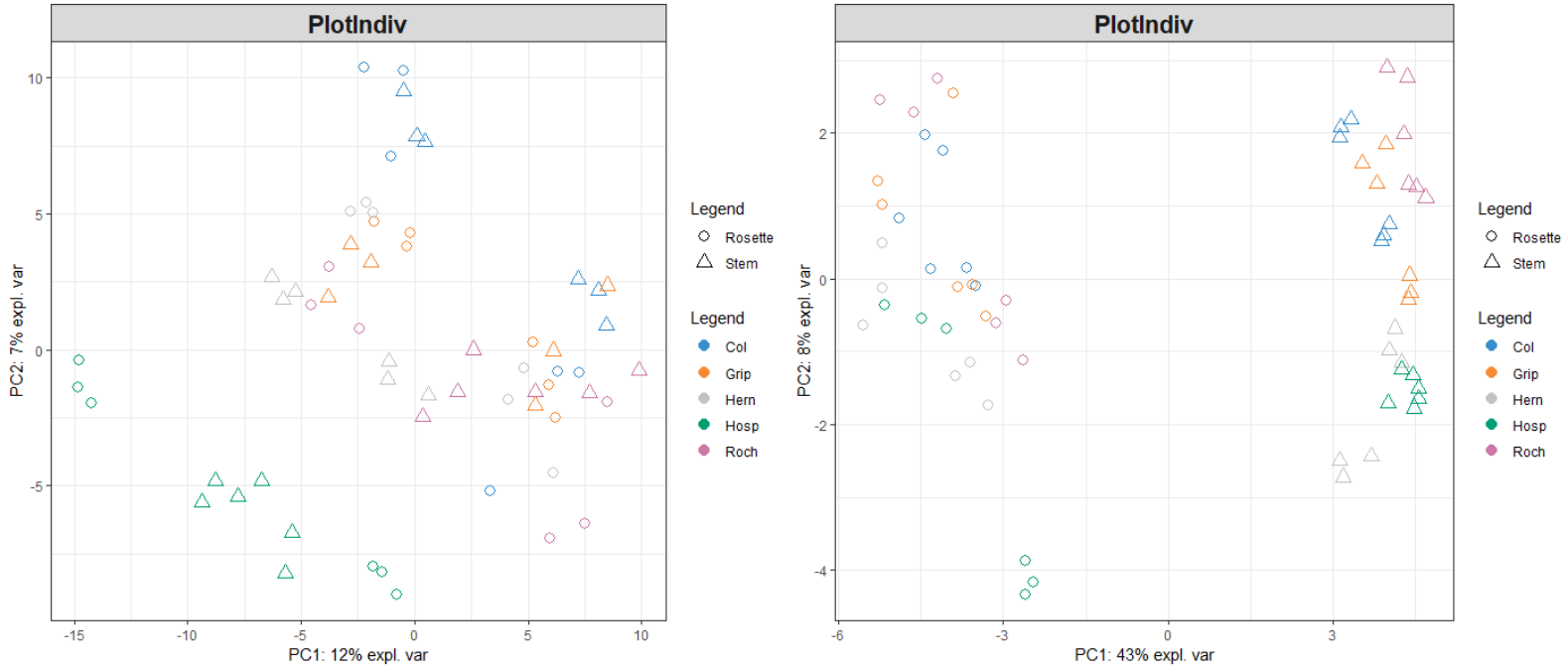
```
# To retrieve the list of common proteins between Stems and Rosettes
Common_List_Prot_Stem_Rosette <- intersect(colnames(Proteomics_Stems_CW),
                                           colnames(Proteomics_Rosettes_CW))

length(Common_List_Prot_Stem_Rosette) # 304 common variables

# To build one single dataset with stem and rosette data
Data_Prot_Mint <- rbind.data.frame(Proteomics_Rosettes_CW[,Common_List_Prot_Stem_Rosette],
                                   Proteomics_Stems_CW[,Common_List_Prot_Stem_Rosette])

# To add factors
Organ_Mint <- as.factor(rep(c("Rosette", "Stem"), each = 30))
Ecotype_Mint <- rep(Ecotype, 2)
Genetic_Cluster_Mint <- rep(Genetic_Cluster, 2)
Altitude_Cluster_Mint <- rep(Altitude_Cluster, 2)

# To make the rownames more explicit and not duplicated
rownames(Data_Prot_Mint)[31:60]<-paste0("Stem.", rownames(Data_Prot_Mint)[1:30])
rownames(Data_Prot_Mint)[1:30]<-paste0("Rosette.", rownames(Data_Prot_Mint)[1:30])
```



With no prior, what are the main effects of different environmental growth conditions or different ecotypes, when controlling the variations due to the organ?

→ Perform MINT-PCA

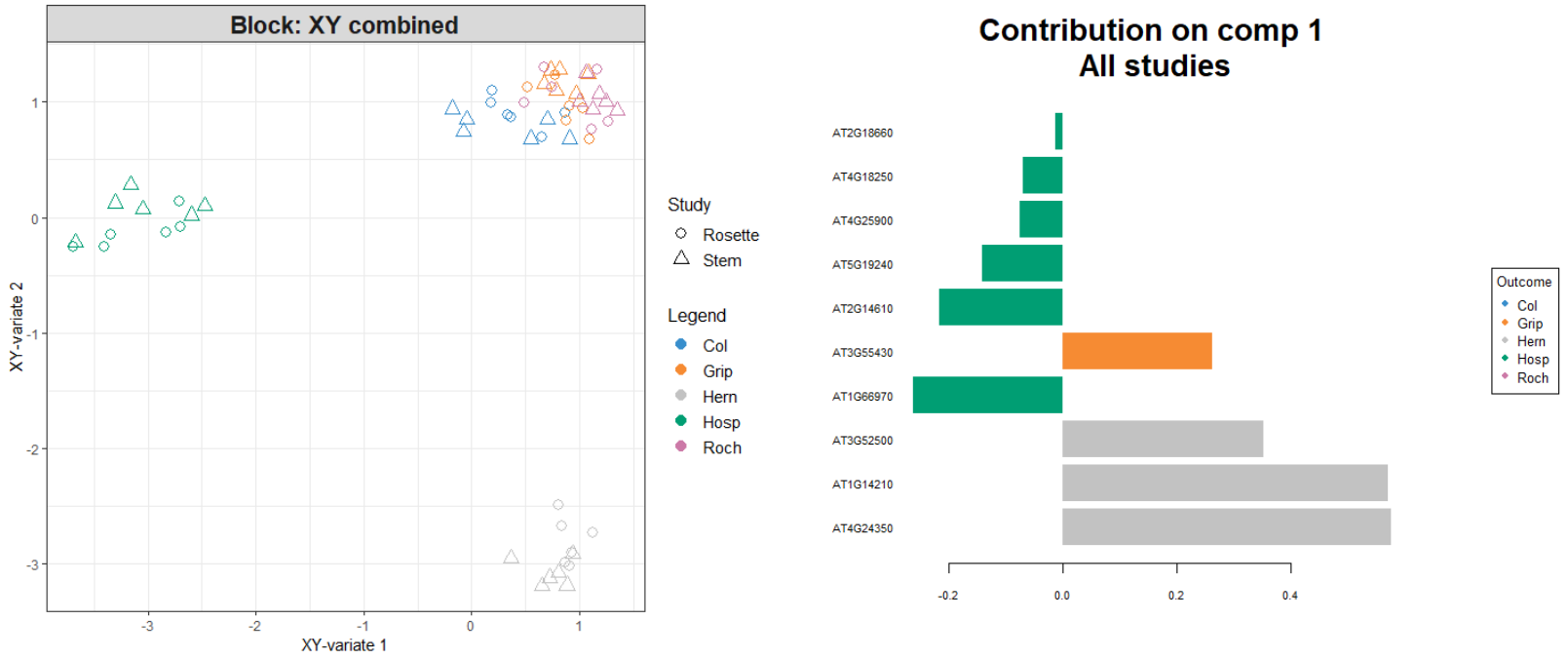
```
res_mint_pca <- mint.pca(X = Data_Prot_Mint, study = Organ_Mint, ncomp = 3)
plotIndiv(res_mint_pca, legend = TRUE, ind.names = FALSE, pch = Organ_Mint, group = Ecotype_Mint)

res_pca_no_mint <- pca(X = Data_Prot_Mint, ncomp = 3)
plotIndiv(res_pca_no_mint, legend = TRUE, ind.names = FALSE, pch = Organ_Mint, group = Ecotype_Mint)
```



Can we determine a proteomics signature of the 5 ecotypes controlling the variations due to the organ?

→ Perform MINT-sPLS-DA

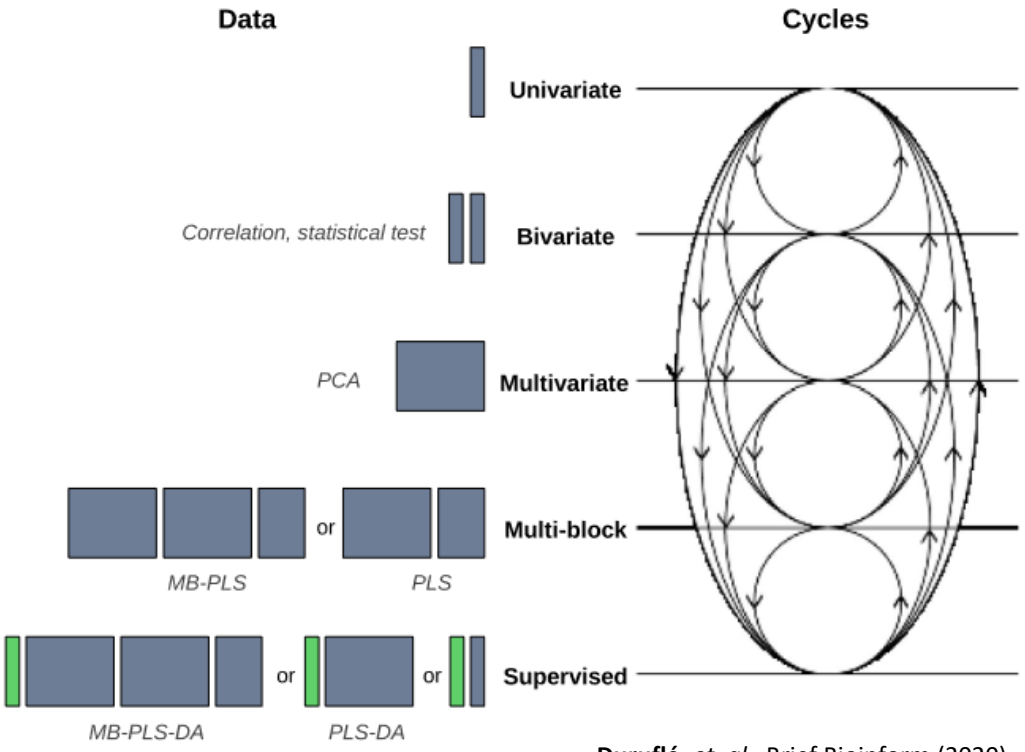
```
res_mint_splsda <- mint.splsda(X = Data_Prot_Mint, Y = Ecotype_Mint, study = Organ_Mint, ncomp = 3,
                               keepX = c(10,10,10))
plotIndiv(res_mint_splsda, legend = TRUE, rep.space = "XY-variate")
plotLoadings(res_mint_splsda, comp = 1, method = "mean", contrib = "max")
```



CONCLUSIONS

Conclusions

- Practice on your own data! The best way to understand what a method has to tell you.
- Do not bypass the elementary analyses (univariate, bivariate, multivariate single data set)
- Clearly identify the biological question to use the most appropriate methods



Duruflé, *et. al.,* Brief Bioinform (2020)

Thanks for your attention

Christophe Dunand
Élisabeth Jamet
Philippe Ranocha
Vincent Burlat
Maxime Bonhomme

Sébastien Déjean
Philippe Besse

Nathalie Escaravage
Monique Burrus

Michel Zivy
Thierry Balliau