Microbial Community Analysis

Bioinformatic tools

Corentin Maslard

October 12, 2026

unine
Université de Neuchâtel
Faculty of Science



Quick poll: pe.app/cmaslar3

One question from Session 3 before we move on.

Why analyse the microbial communities of plants and soil?

Microbial key functions in the plant–soil system

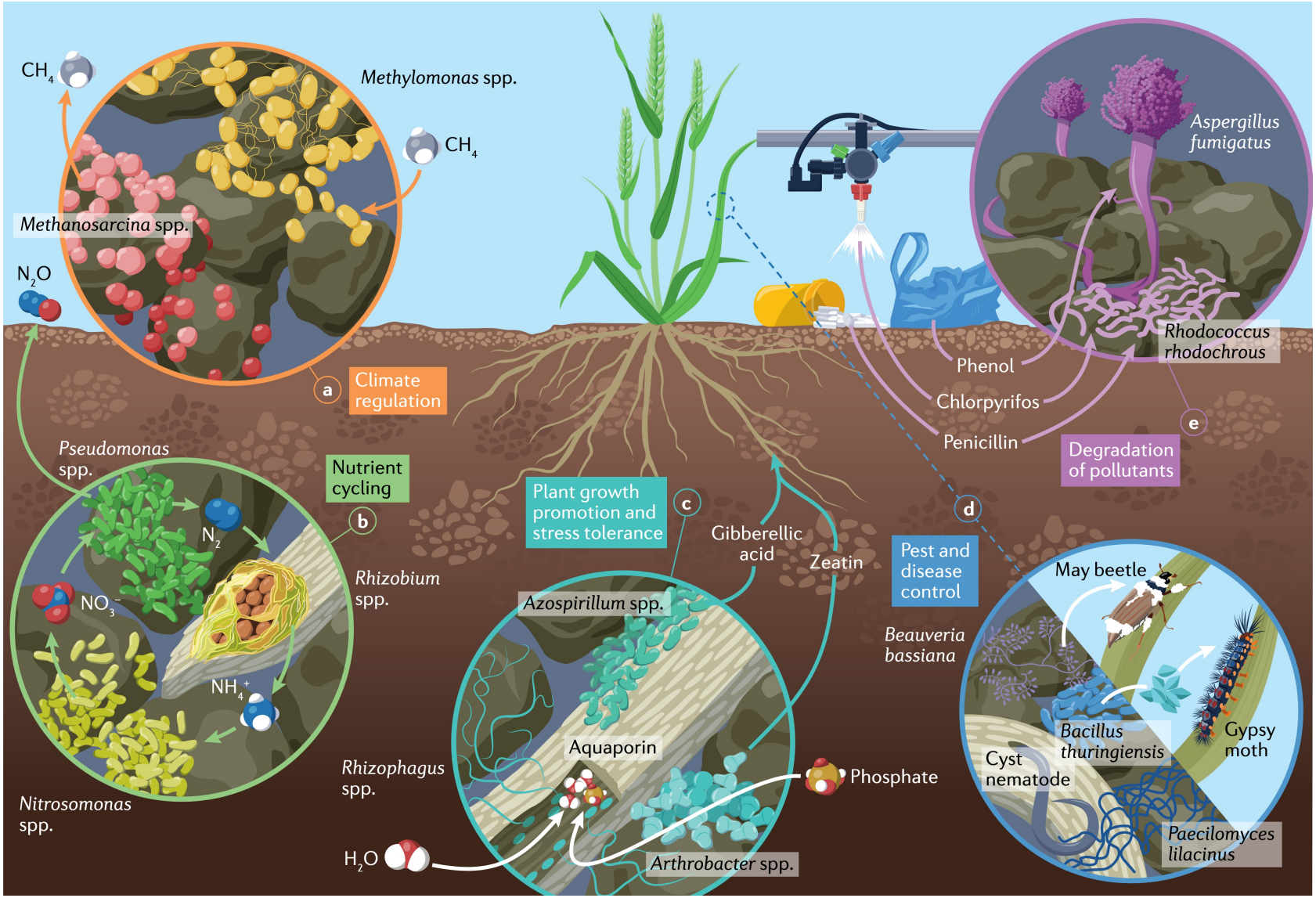
a. Climate regulation

b. Microbial nutrient cycling

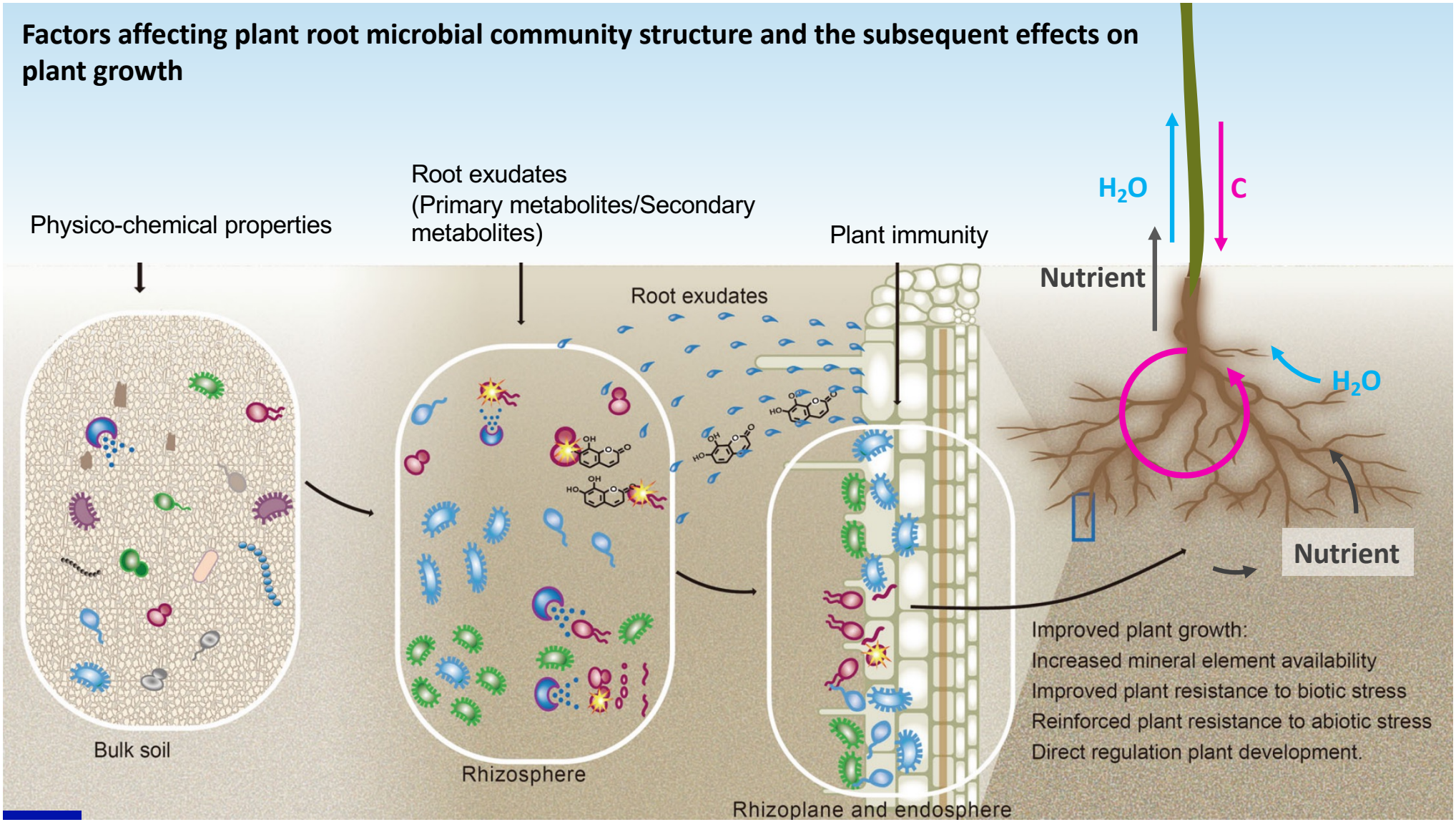
c. Plant growth promotion and abiotic stress tolerance

d. Pest and disease control

e. Toxin and pollutant degradation.



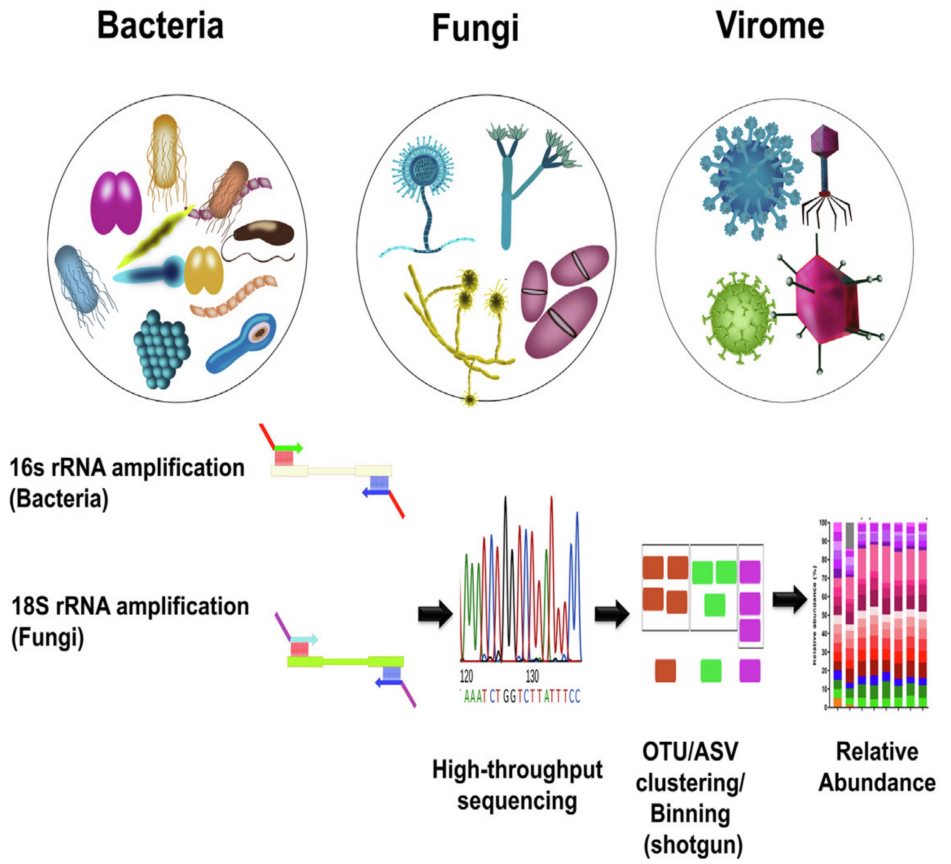
(Hartmann and Six, 2022)


Factors affecting plant root microbial community structure and the subsequent effects on plant growth
Physico-chemical properties
Root exudates
(Primary metabolites/Secondary metabolites)
Plant immunity
Root exudates
H_2O
C
Nutrient
H_2O
Nutrient
Bulk soil
Rhizosphere
Rhizoplane and endosphere
Improved plant growth:
Increased mineral element availability
Improved plant resistance to biotic stress
Reinforced plant resistance to abiotic stress
Direct regulation plant development.

What type of analysis can be performed to measure microbial communities?

The different types of microbial community analysis

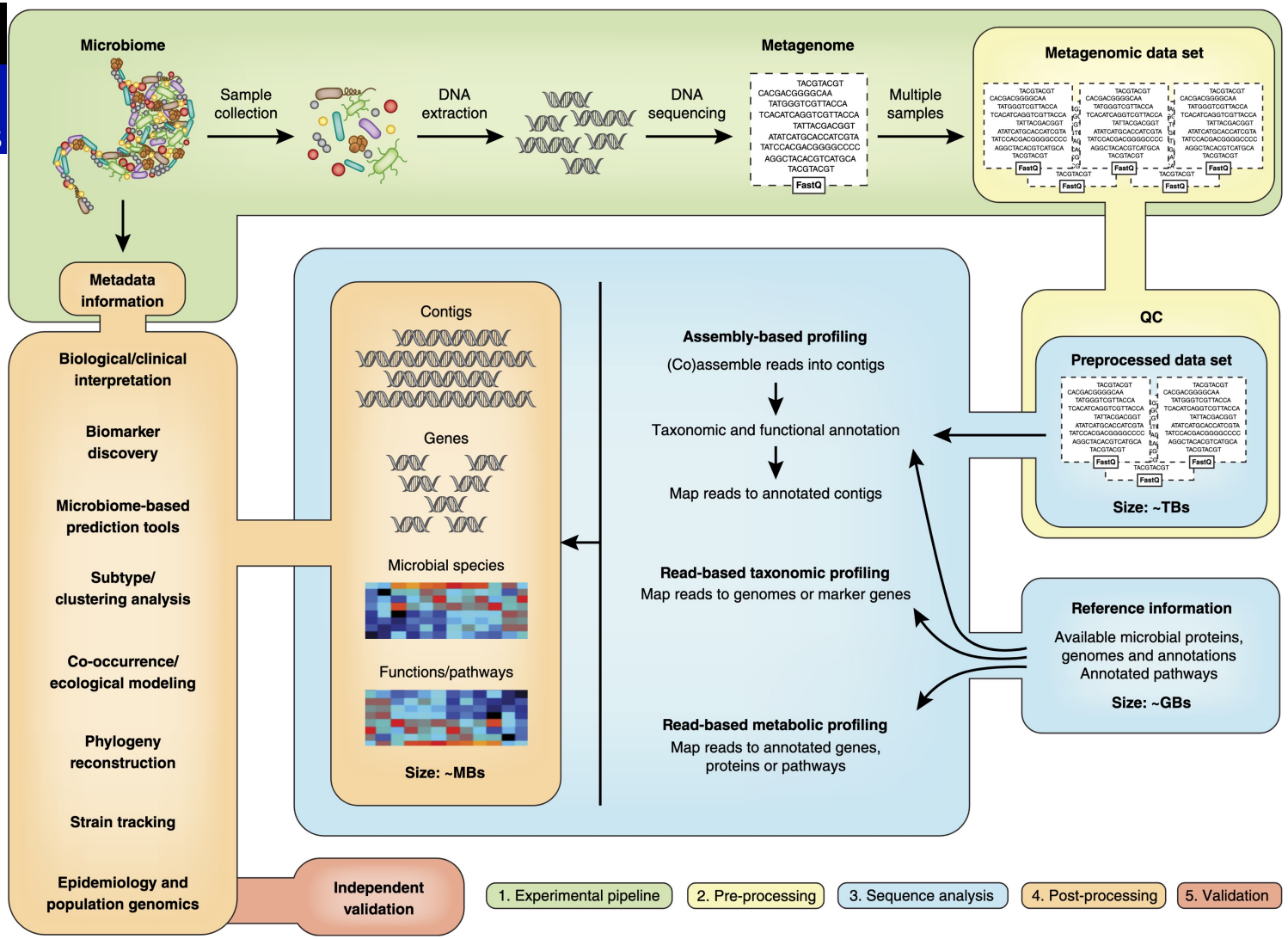
- **Amplicon sequencing** → diversity & composition.



(Matchado et al., 2021)

The different types of microbial community analysis

- **Amplicon sequencing** (diversity & composition)
- **Shotgun/metagenomics** (functional potential)

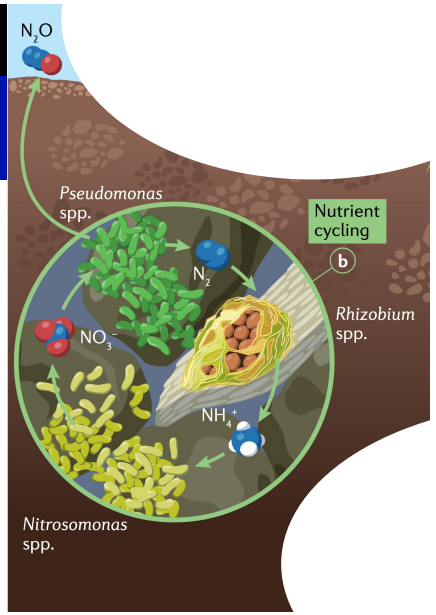


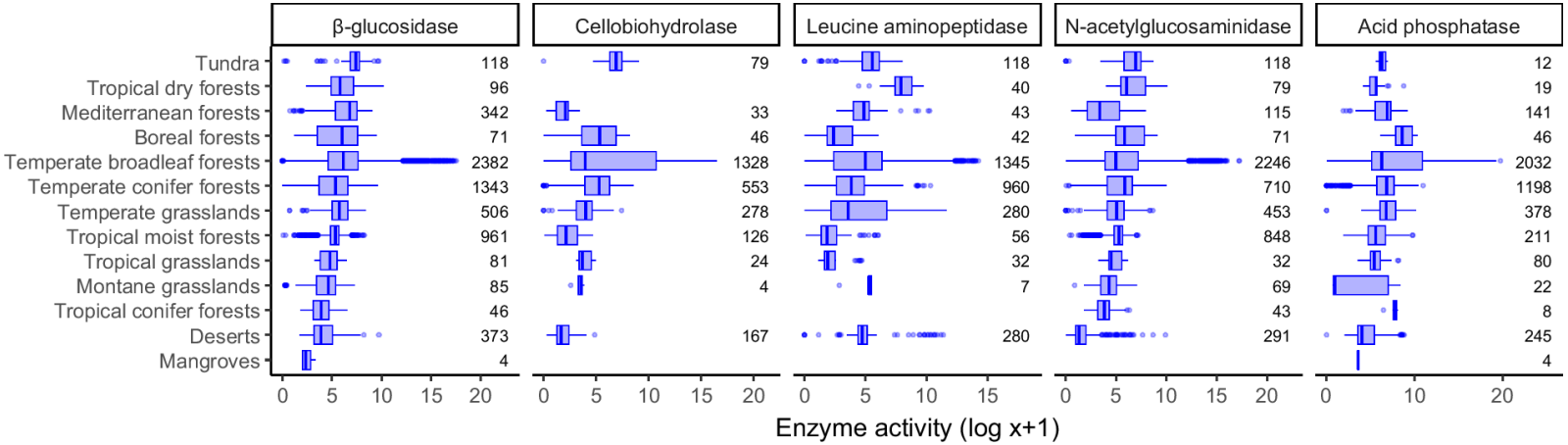
(Quince et al., 2017)

The different types of microbial community analysis

- **Amplicon sequencing** → diversity & composition.
- **Shotgun/metagenomics** → functional potential.
- **Enzyme assays** → global activity indicators.

The type of enzyme	Pathway	Process
Ammonia monooxygenase (AMO)	$NH_4^+ - N \rightarrow NH_2OH$	Nitrification
Hydroxylamine oxidoreductase (HAO)	$NH_2OH \rightarrow NOH$	Nitrification
Nitrite reductase (NXR)	$NO_2^- \rightarrow NO$	Denitrification
Nitric oxide reductase (NOR)	$NO \rightarrow N_2O$	Denitrification
Nitrous oxide reductase (NOS)	$N_2O \rightarrow N_2$	Denitrification

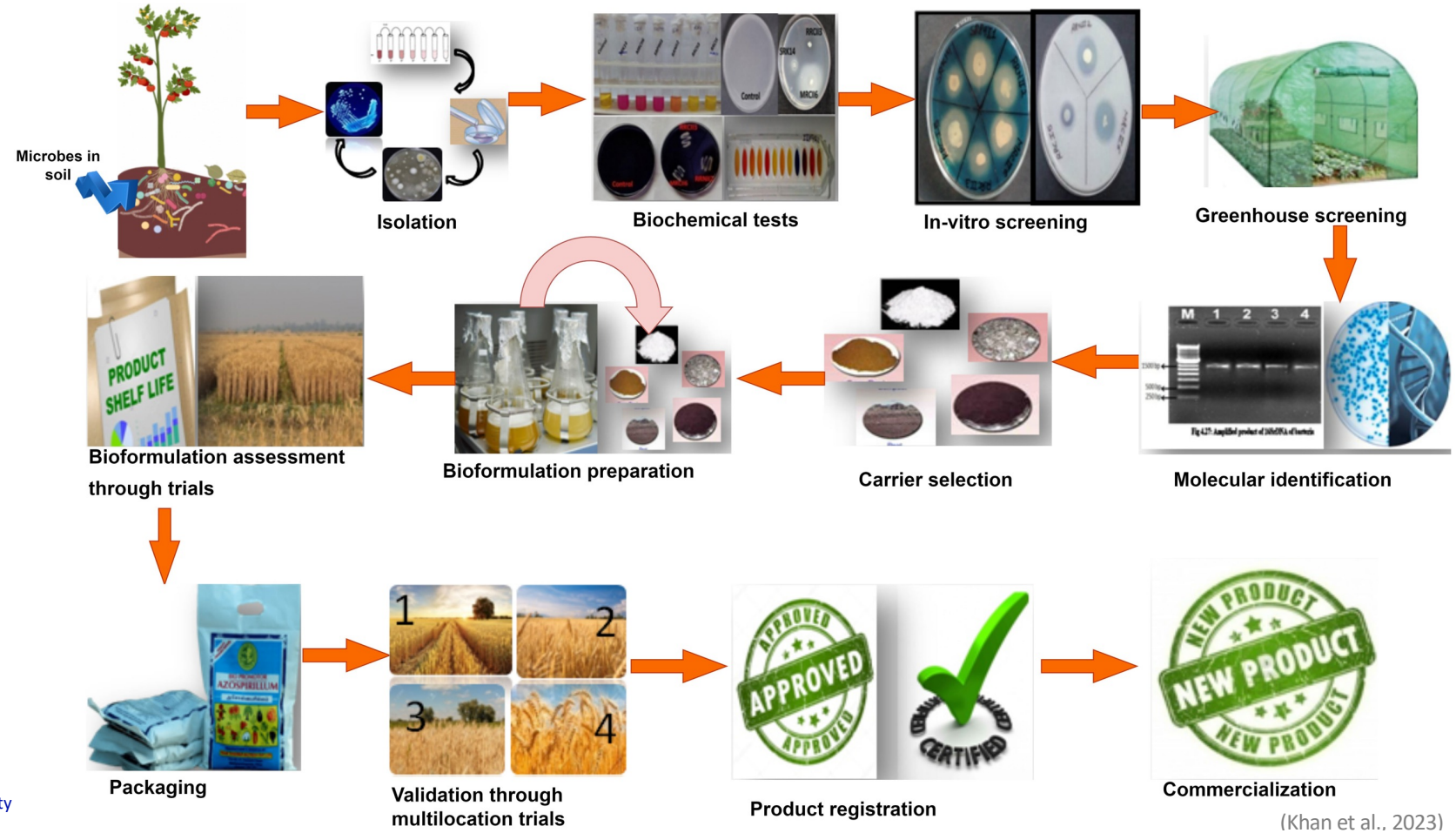




(van Galen et al., 2025), (Hartmann and Six, 2022)

The different types of microbial community analysis

- **Amplicon sequencing** (diversity & composition)
- **Shotgun/metagenomics (**functional potential)
- **Enzyme assays** (global activity indicators)
- **Cultivation & functional tests** → (validation with individual microbes)

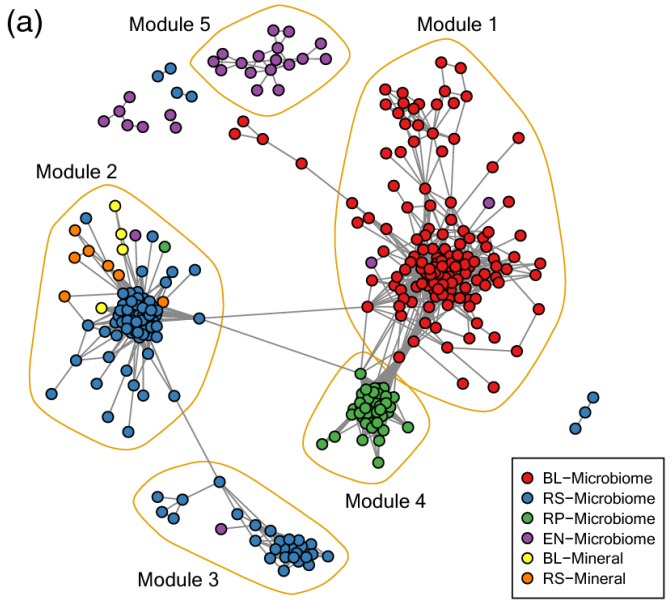


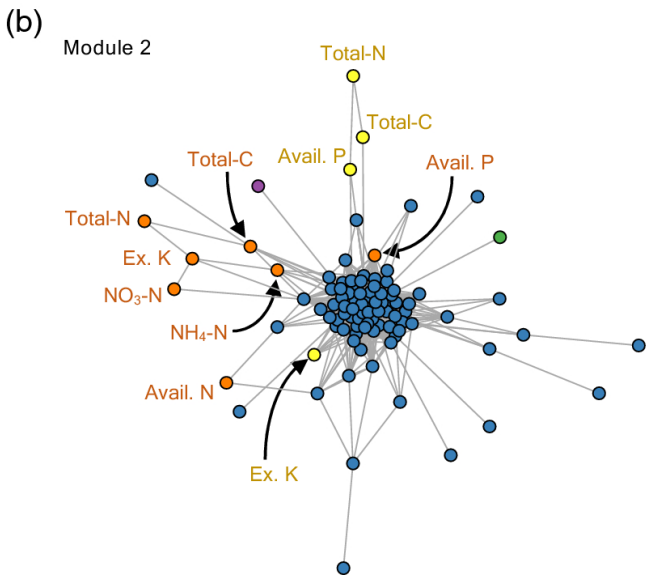
(Khan et al., 2023)

The different types of microbial community analysis

- **Amplicon sequencing** → diversity & composition
- **Shotgun/metagenomics** → functional potential
- **Enzyme assays** → global activity indicators
- **Cultivation & functional tests** → validation with individual microbes
- **Multi-omics** → plant–microbe interactions at multiple scales

BL, bulk soil
RS, rhizosphere
RP, rhizoplane
EN, endosphere

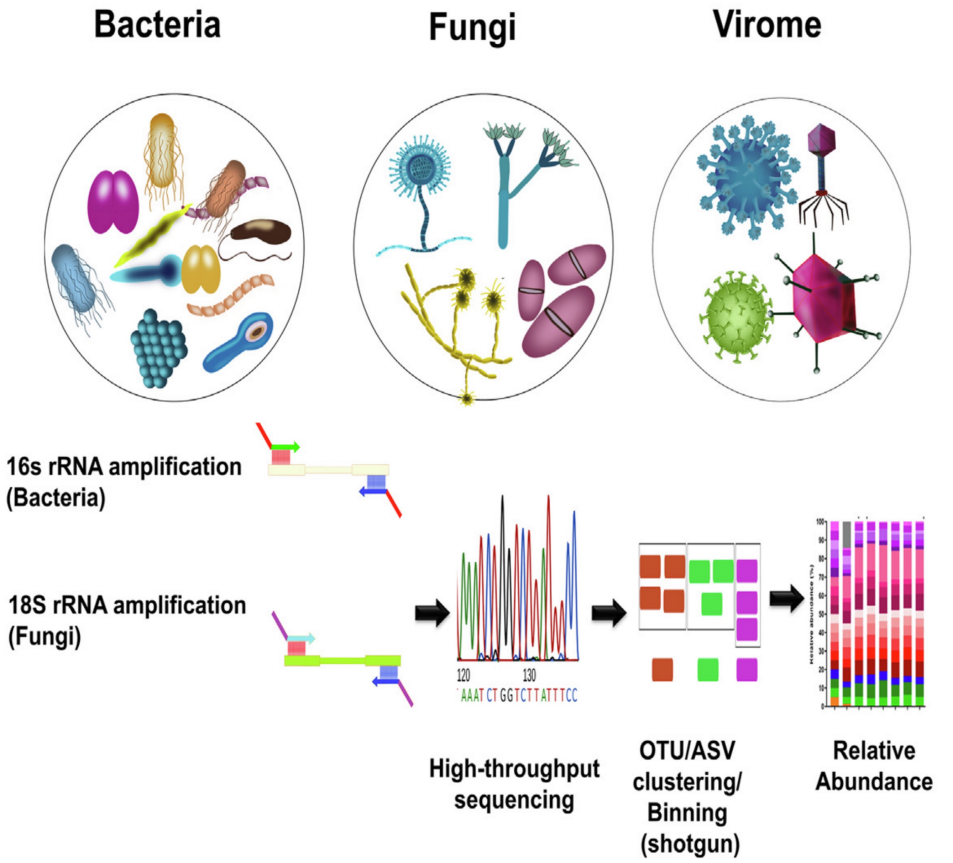




(Yamazaki et al., 2021)

The different types of microbial community analysis

- **Amplicon sequencing** → diversity & composition
- **Shotgun/metagenomics** → functional potential
- **Enzyme assays** → global activity indicators
- **Cultivation & functional tests** → validation with individual microbes
- **Multi-omics** → plant–microbe interactions at multiple scales

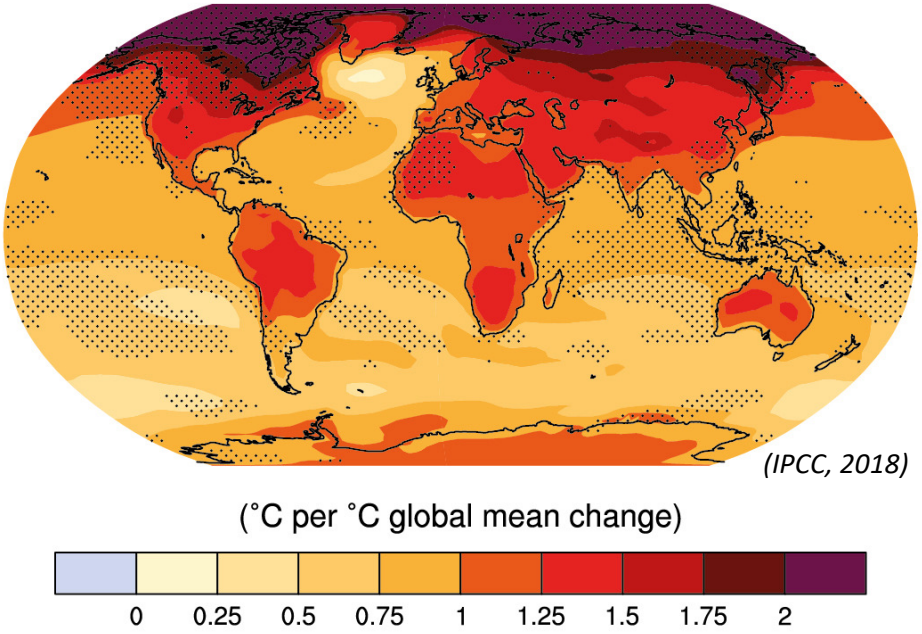


(Matchado et al., 2021)

How does global warming alter plant microbial communities?

Predictions indicate a global average temperature increase of up to 6°C by 2100

Evolution of the **average surface temperature**
(between 1986-2005 and 2081-2100)

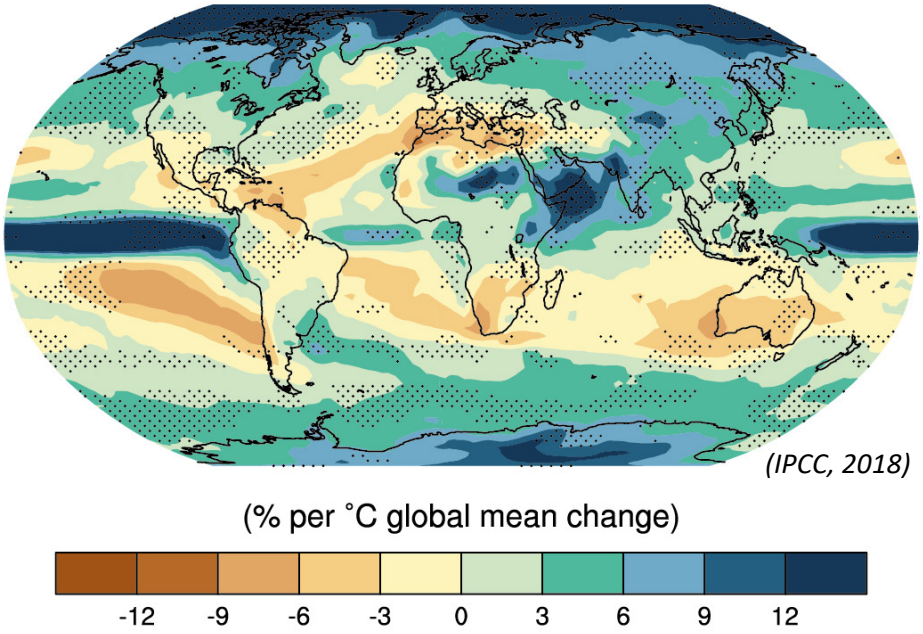


(IPCC, 2018)

Some mid-latitude regions, such as the Mediterranean, will experience more intense drought events

Evolution of **average precipitation** (between 1986-2005 and 2081-2100)

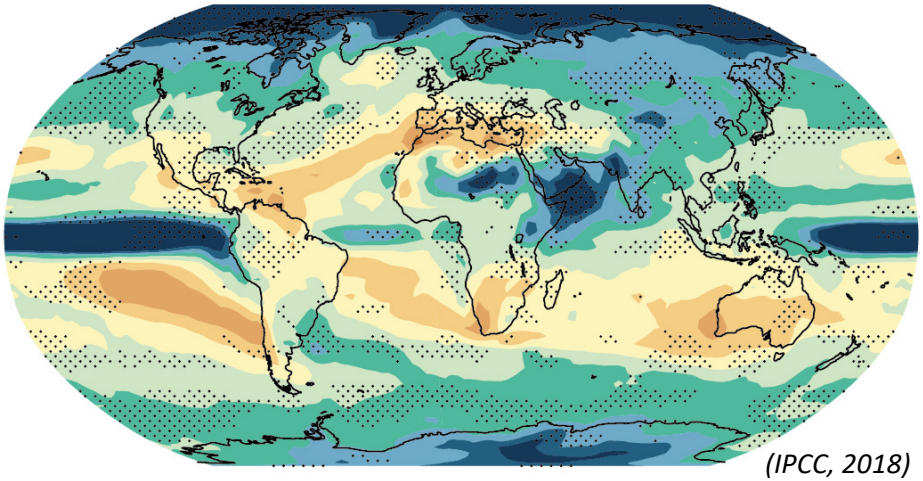
2100: Models predict +2°C, 35 rainless days in Europe, causing **soybean yields to drop by 50%.**



(IPCC, 2018)

Some mid-latitude regions, such as the Mediterranean, will experience more intense drought events

Evolution of **average precipitation** (between 1986-2005 and 2081-2100)



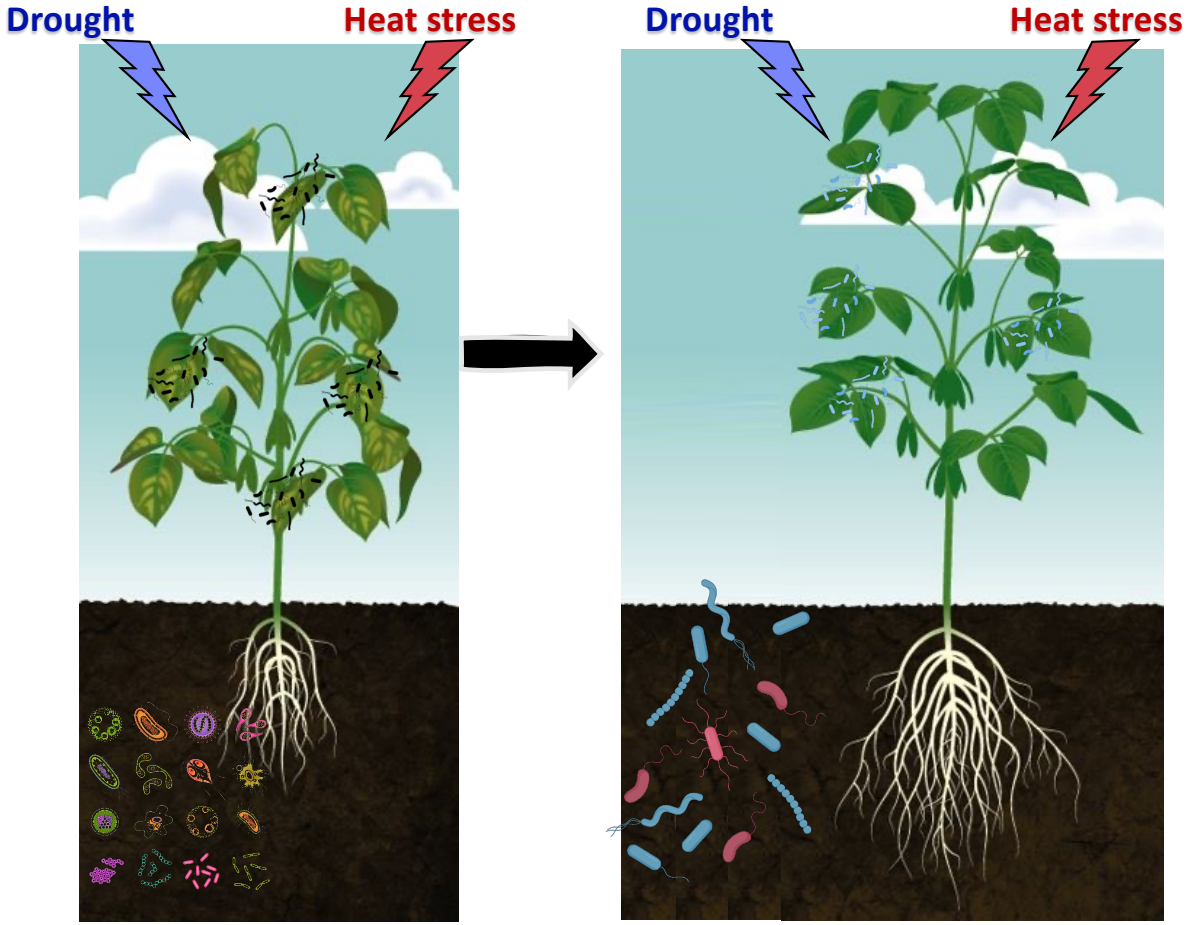
(IPCC, 2018)

2100: Models predict +2°C, 35 rainless days in Europe, causing **soybean yields to drop by 50%.**

Studying the possibility of improving plant resistance by examining the interaction between plant life and micro-organisms

Hypothesis :

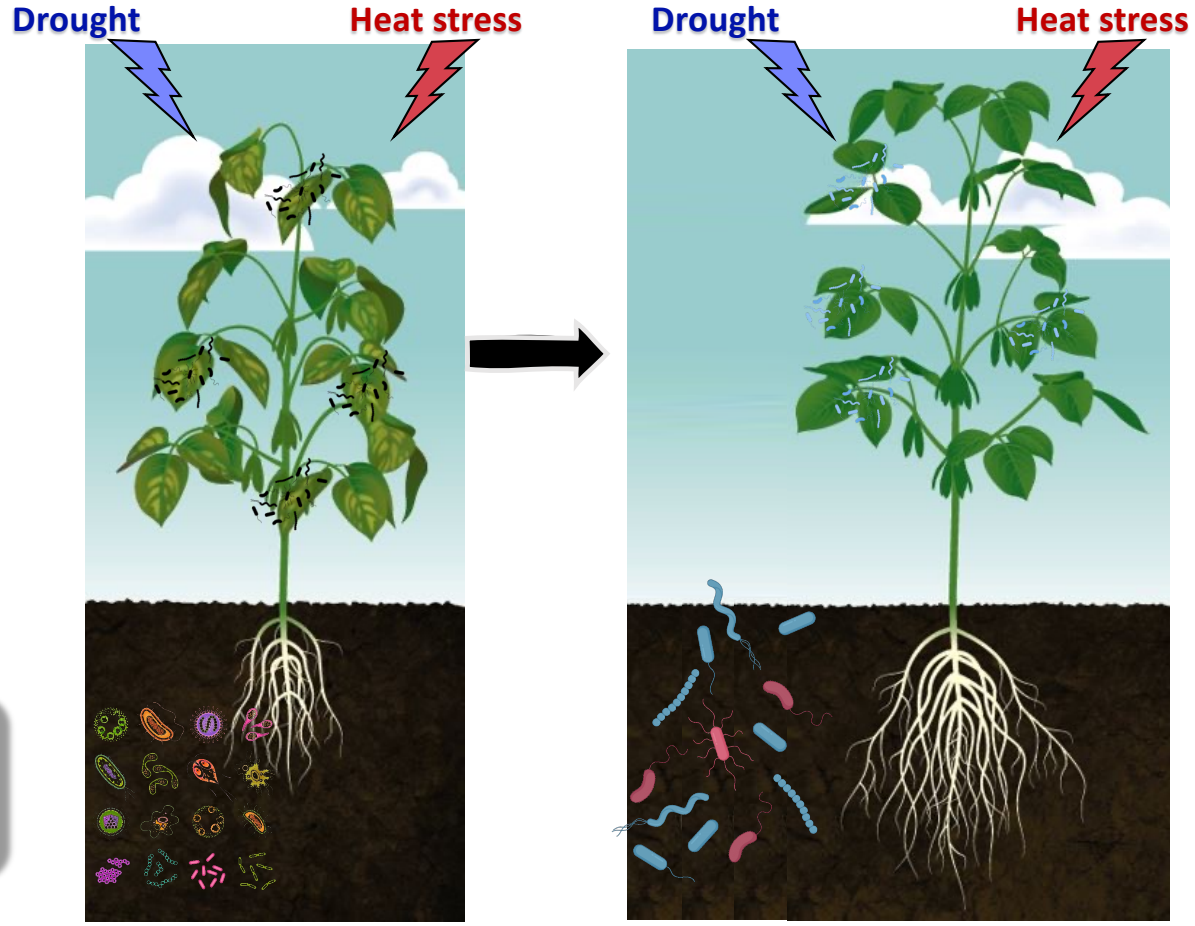
- Microbial communities exhibit unique responses to **heat** and/or **water** stress.
- The disappearance or appearance of certain taxa under double stress may not be a simple additive effect of the two stresses.
- Each plant compartment reacts uniquely and interdependently to each stress.
- Microbial communities may exist exclusively on plants with increased resistance or sensitivity to water and/or heat stress.



Hypothesis :

- Microbial communities exhibit unique responses to **heat** and/or **water** stress.
- The disappearance or appearance of certain taxa under double stress may not be a simple additive effect of the two stresses.
- Each plant compartment reacts uniquely and interdependently to each stress.
- Microbial communities may exist exclusively on plants with increased resistance or sensitivity to water and/or heat stress.

How to verify these hypotheses ?



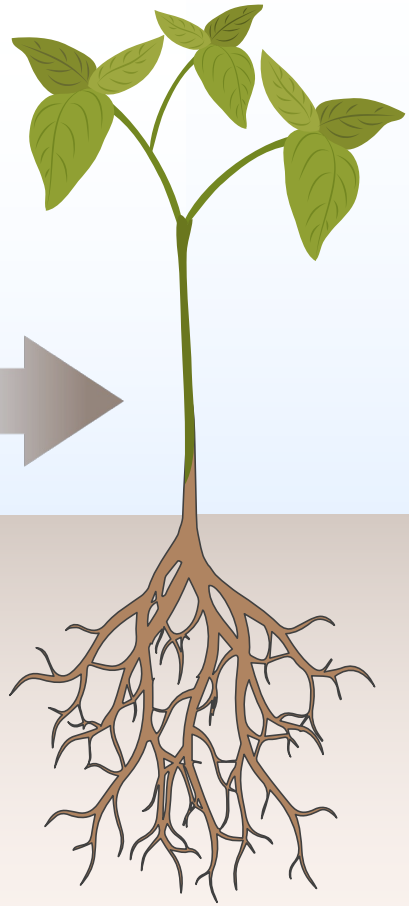
4 climatic conditions

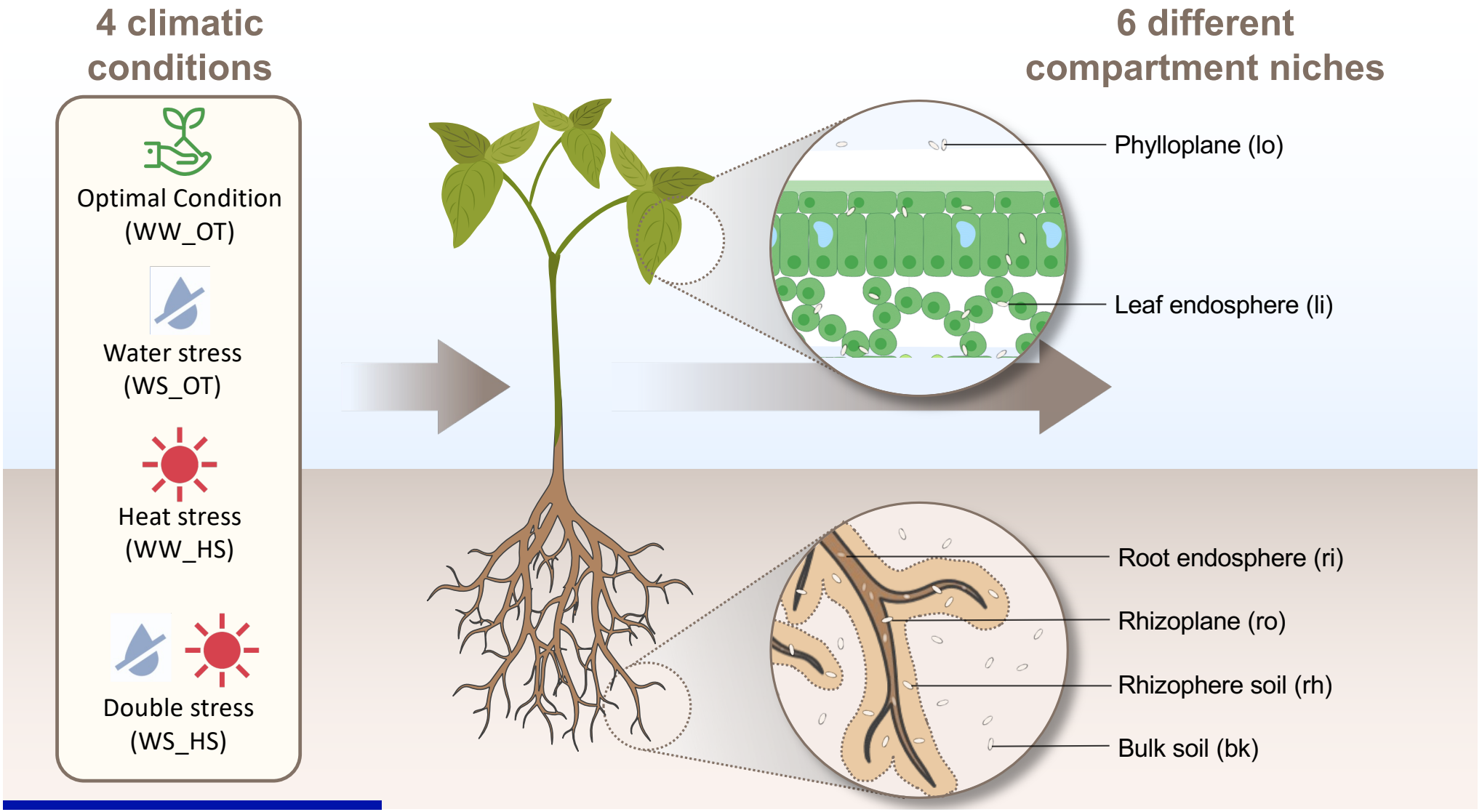
Optimal Condition
(WW_OT)

Water stress
(WS_OT)

Heat stress
(WW_HS)

Double stress
(WS_HS)




4 climatic conditions
Optimal Condition (WW_OT)
Water stress (WS_OT)
Heat stress (WW_HS)
Double stress (WS_HS)
6 different compartment niches
Phylloplane (lo)
Leaf endosphere (li)
Root endosphere (ri)
Rhizoplane (ro)
Rhizophere soil (rh)
Bulk soil (bk)

4 climatic conditions

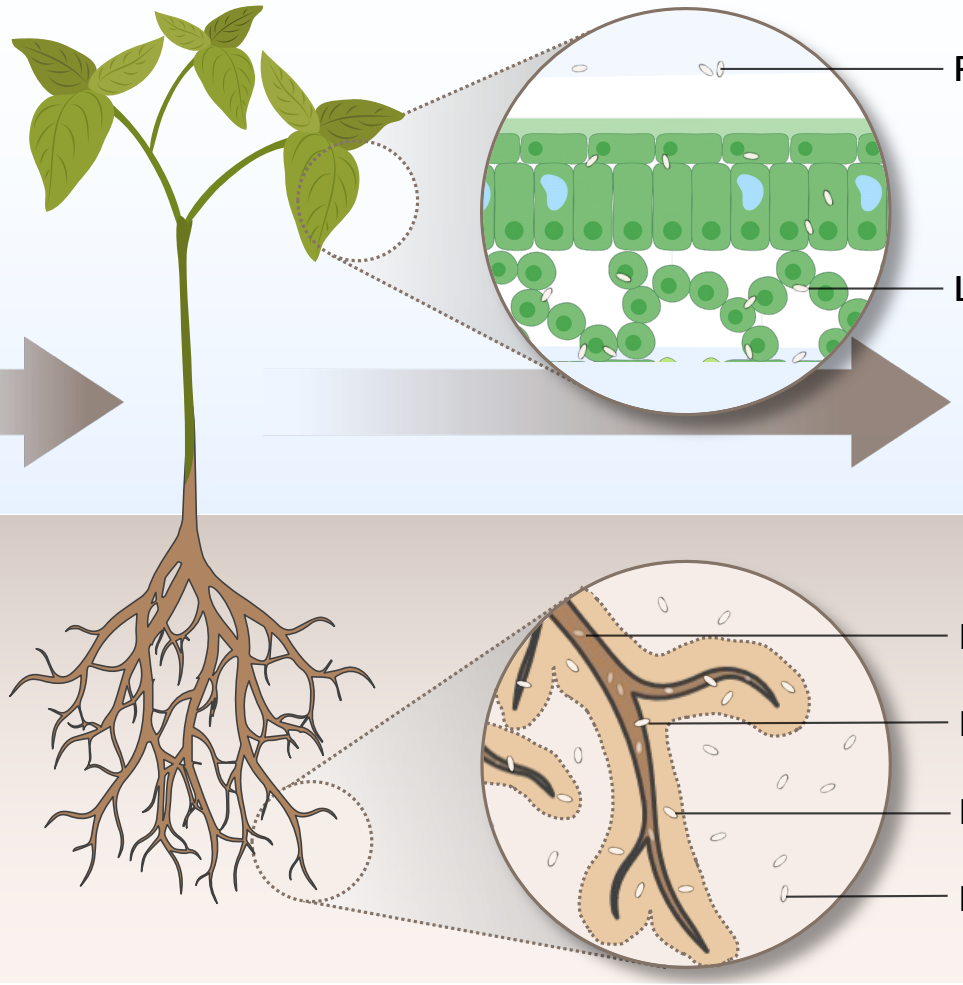
Optimal Condition (WW_OT)

Water stress (WS_OT)

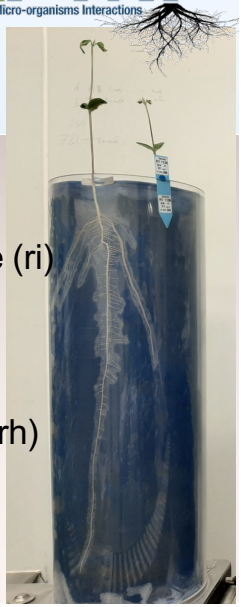
Heat stress (WW_HS)

Double stress (WS_HS)

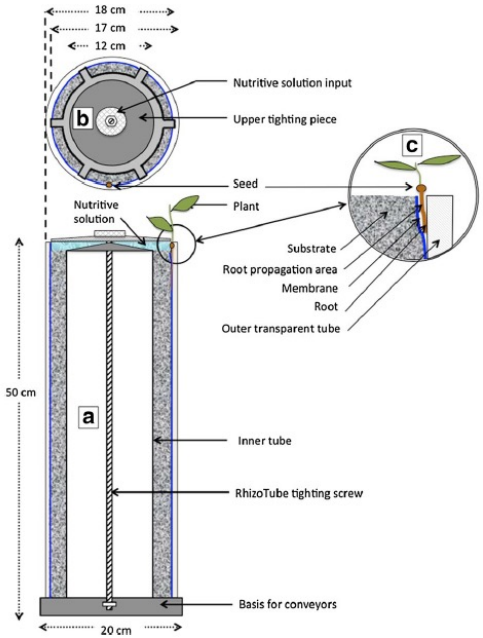
6 different compartment niches



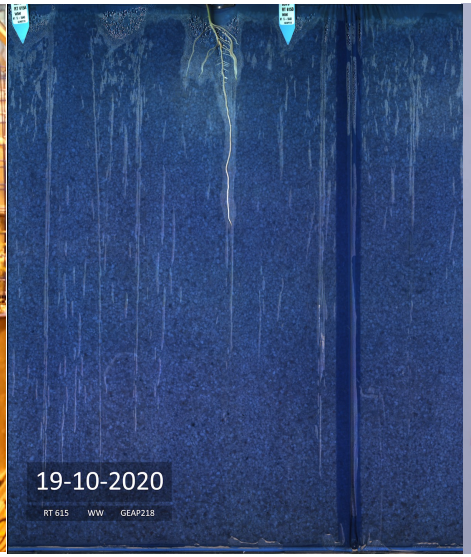
4PMI Plant Phenotyping Platform
Plant Micro-organisms Interactions



Architecture measurement with the 4pmi platform in Dijon, France

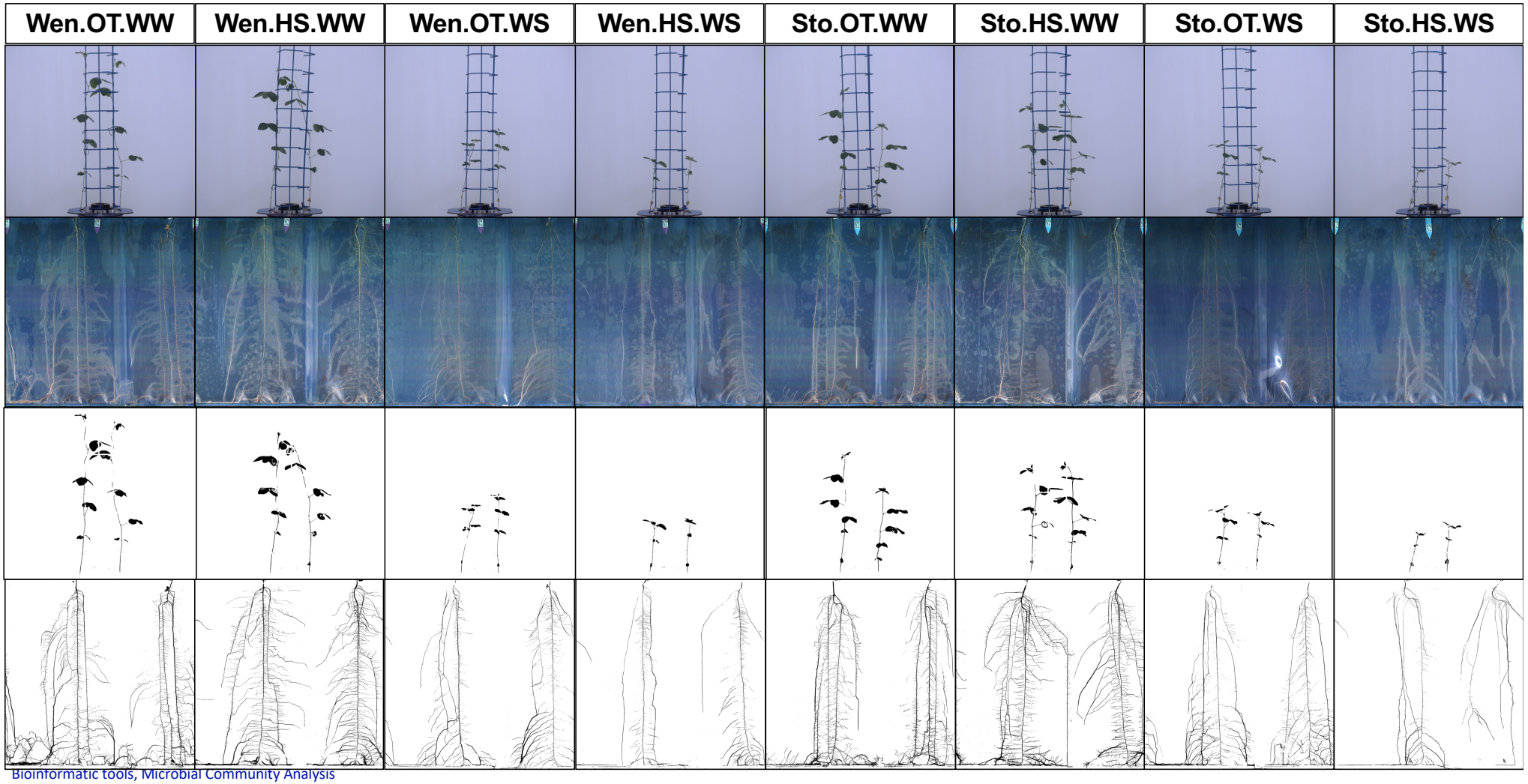




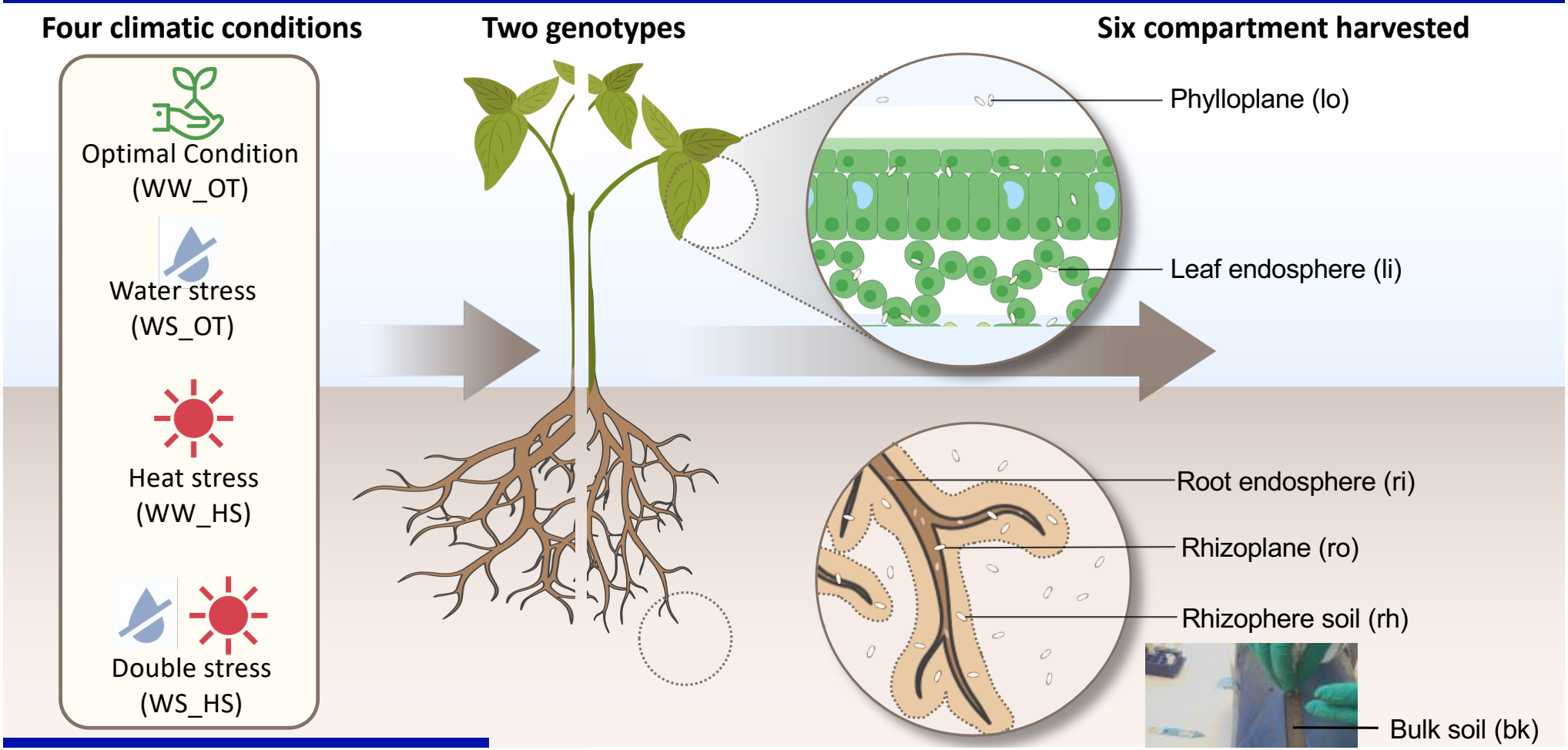




Non-destructive phenotyping of roots over time, precise control of watering and greenhouse climate


Wen.OT.WW
Wen.HS.WW
Wen.OT.WS
Wen.HS.WS
Sto.OT.WW
Sto.HS.WW
Sto.OT.WS
Sto.HS.WS

Collection of 6 different compartments for bacterial and fungal community analysis



General analysis pipeline

Harvesting and extraction DNA microorganisms



DNA extraction and PCR

Bacterial 16S ribosomal DNA (rDNA) and fungal ITS2 region



Sequencing

NovaSeq platform and 250 bp paired-end reads were generated



Bioinformatics analysis

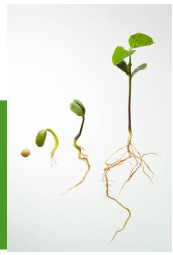
DADA2 Amplicon Sequencing. Exactly.

qiime2

R Vegan Cytoscape phyloseq

General analysis pipeline

Harvesting and extraction DNA microorganisms



DNA extraction and PCR

Bacterial 16S ribosomal DNA (rDNA) and fungal ITS2 region



Sequencing

NovaSeq platform and 250 bp paired-end reads were generated



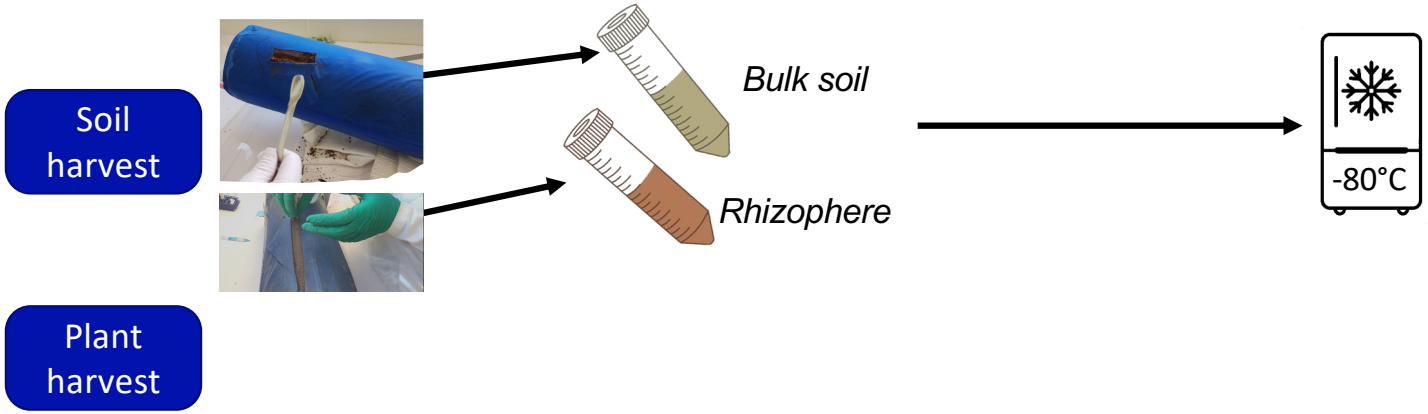
Bioinformatics analysis

DADA2 Amplicon Sequencing. Exactly.

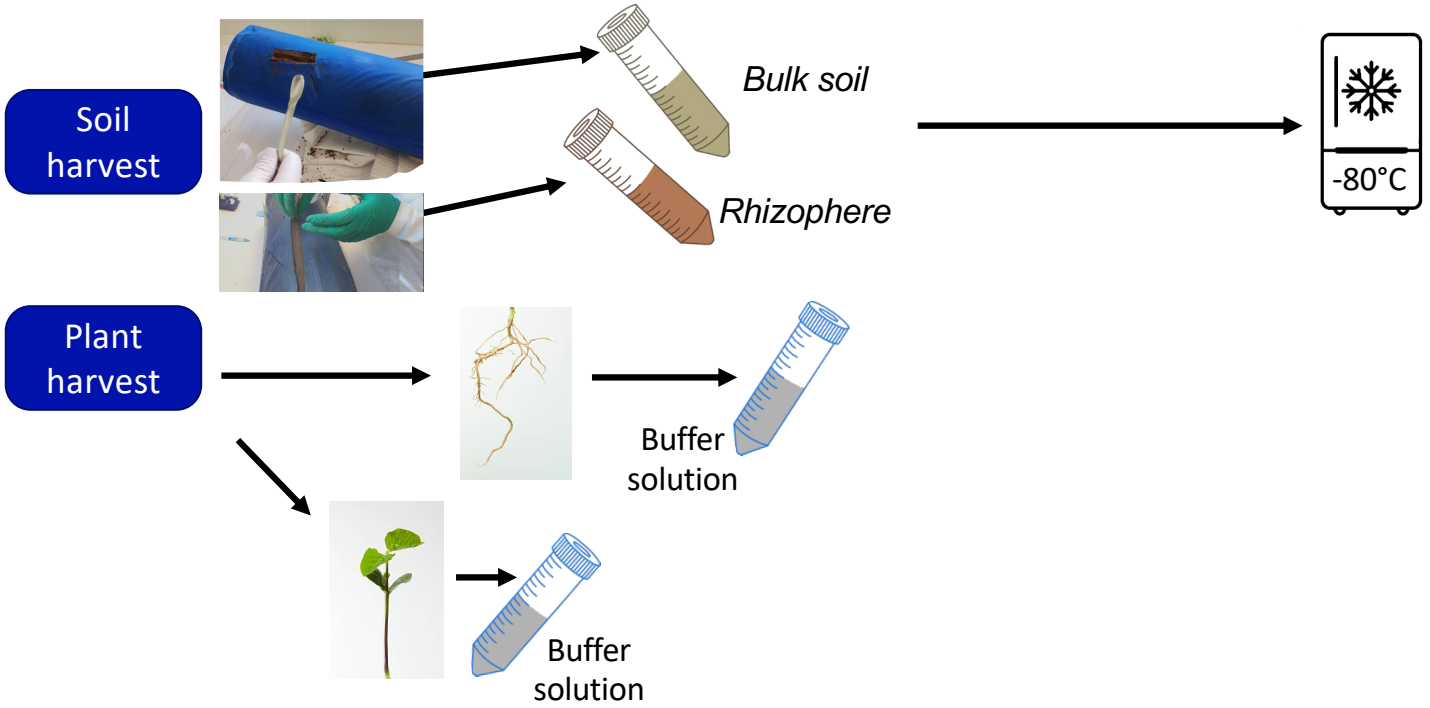
qiime2

R Vegan Cytoscape phyloseq

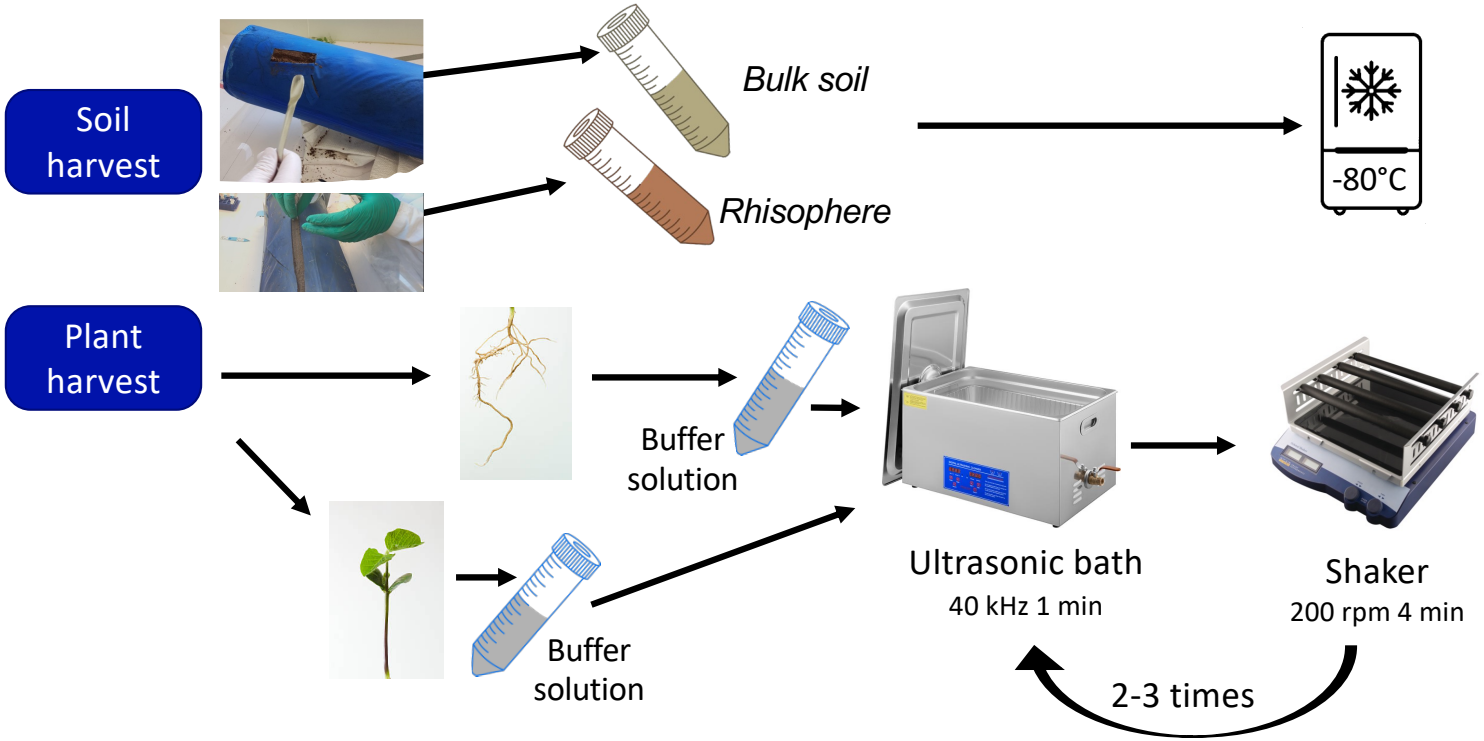
Sample collection method



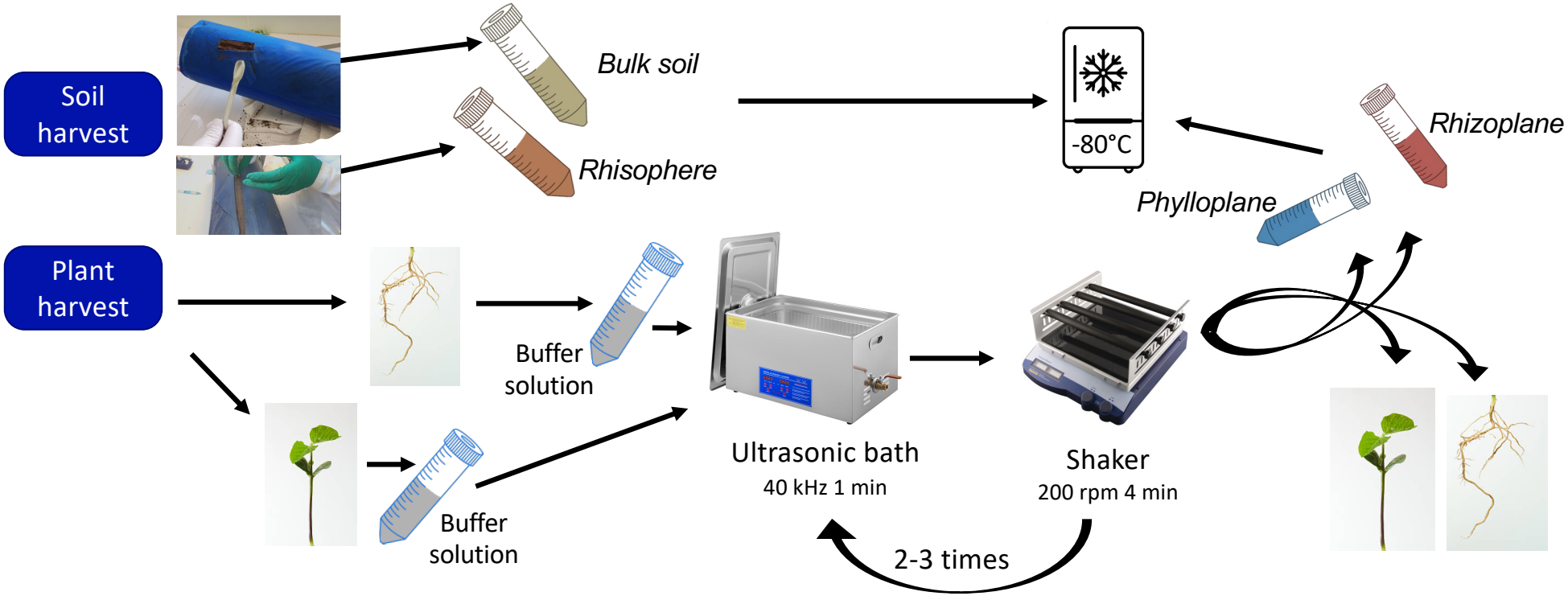
Sample collection method



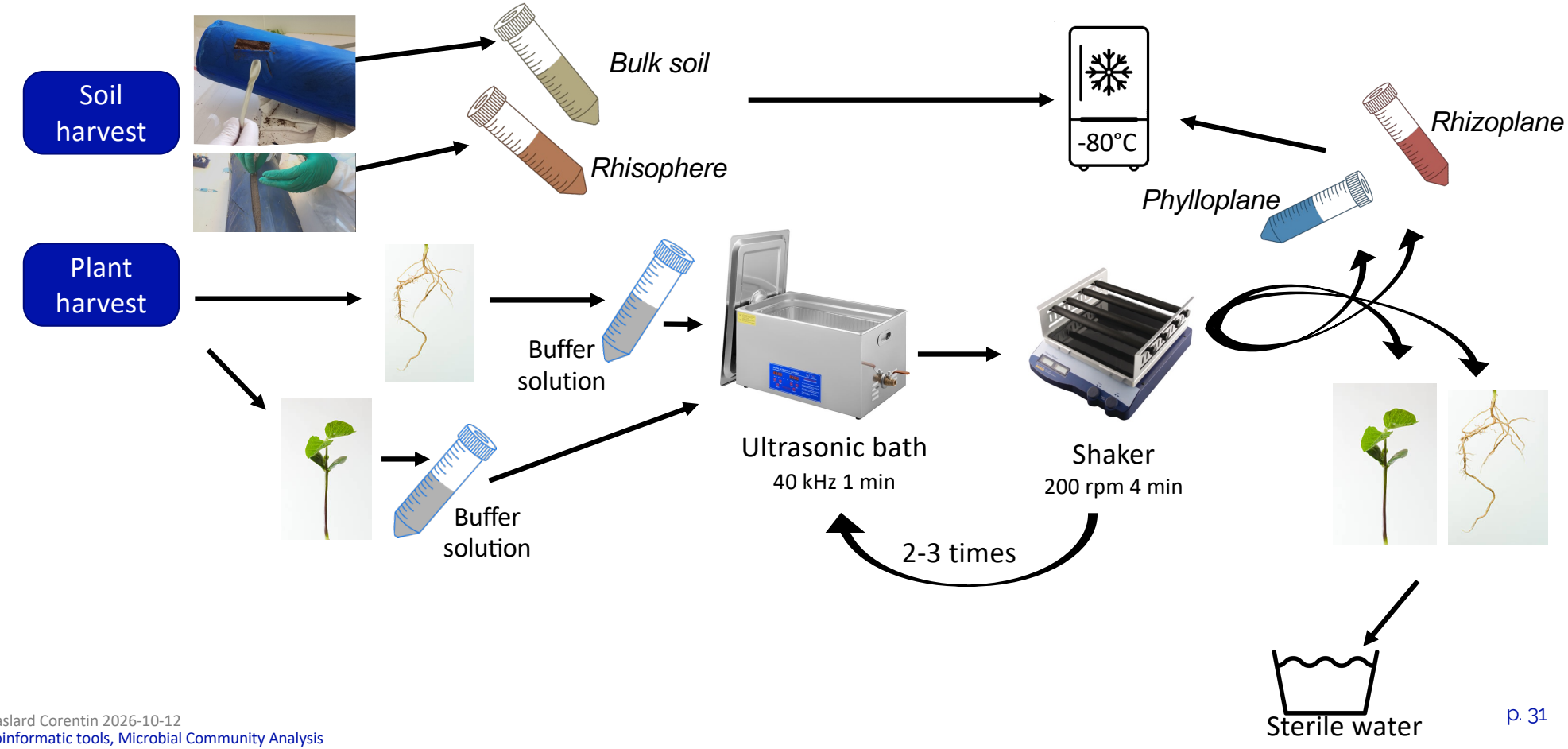
Sample collection method



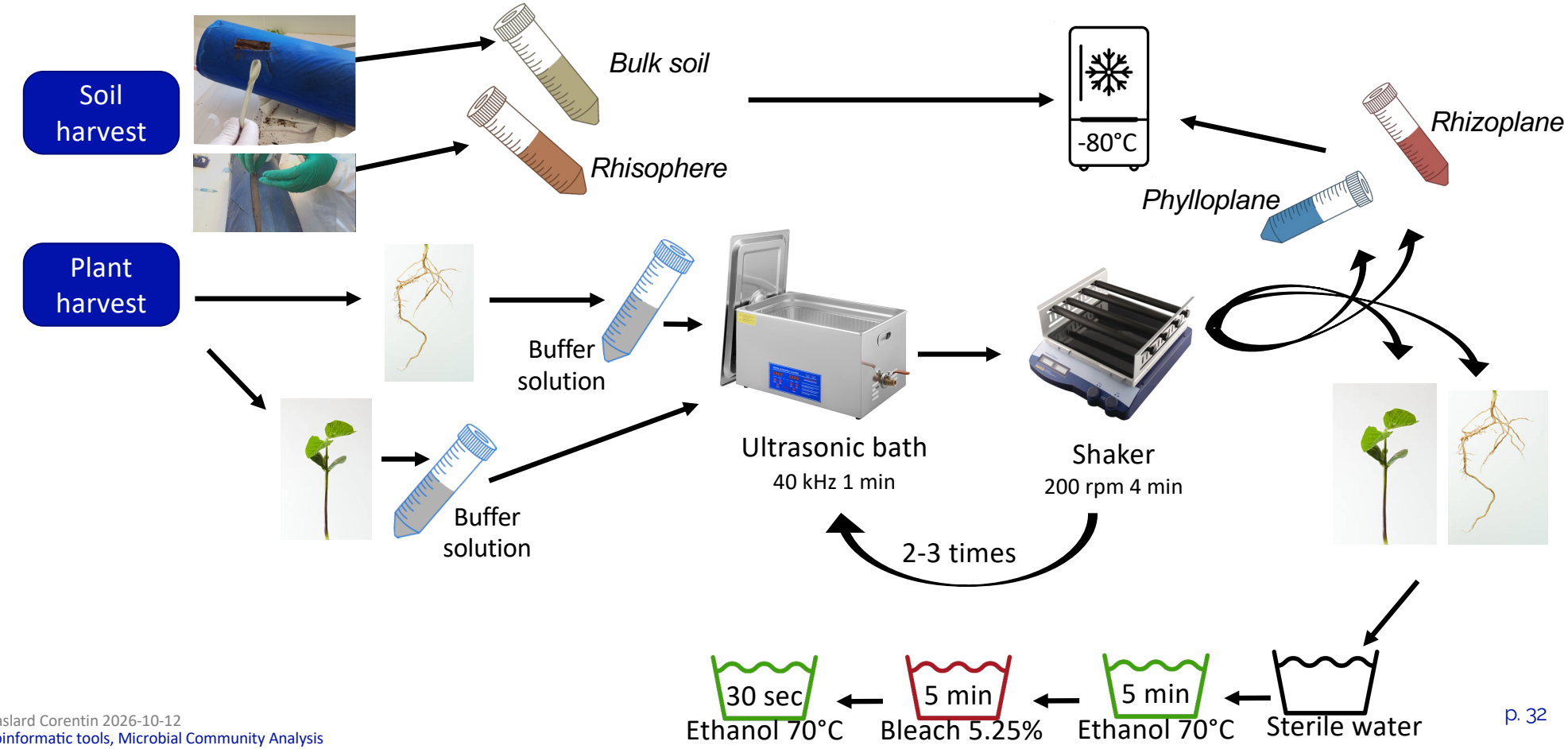
Sample collection method



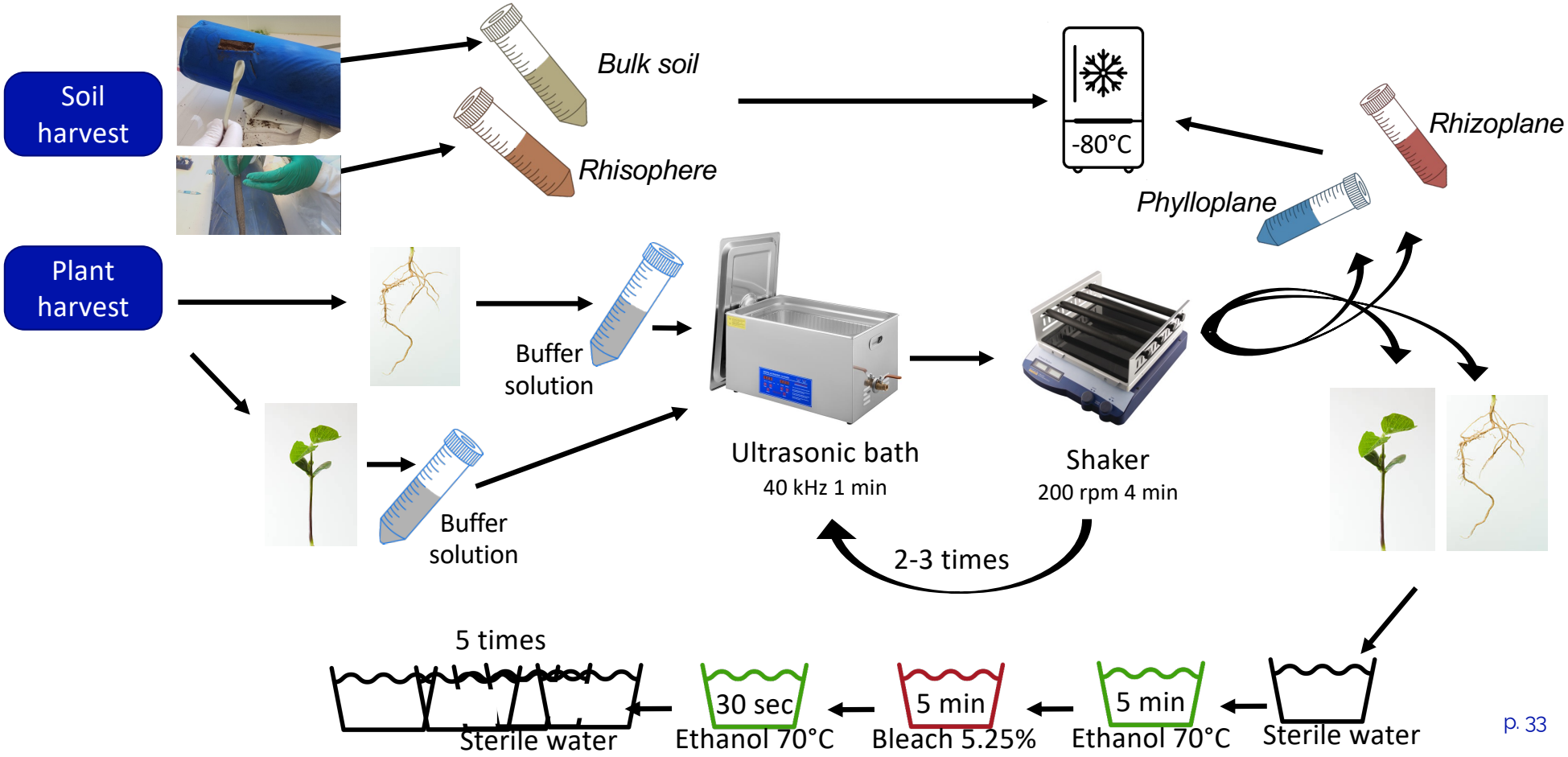
Sample collection method



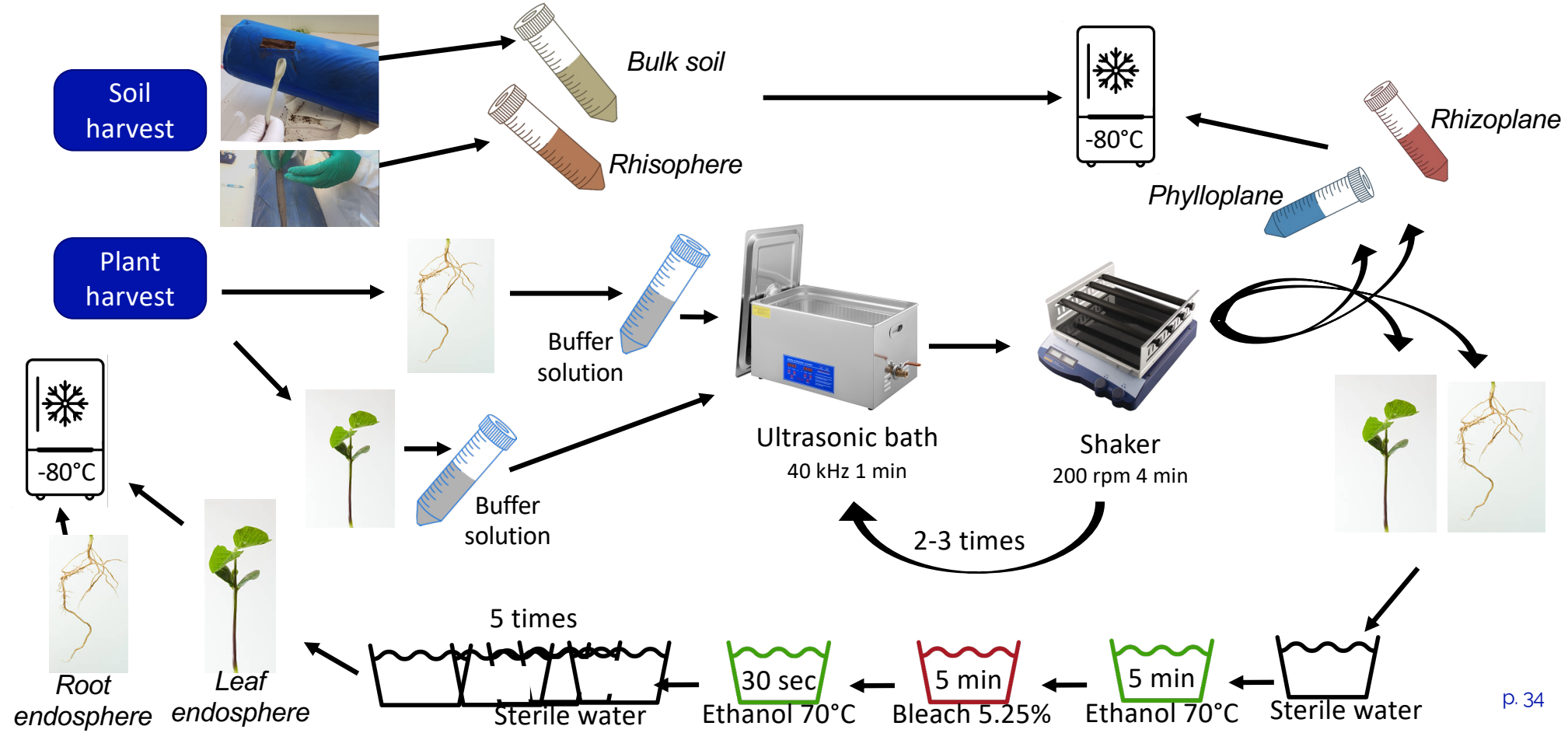
Sample collection method



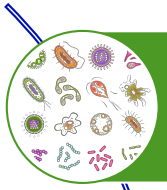
Sample collection method



Sample collection method



General analysis pipeline



Harvesting and extraction DNA microorganisms





DNA extraction and PCR

Bacterial 16S ribosomal DNA (rDNA) and fungal ITS2 region

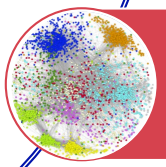


Novogene

Sequencing

NovaSeq platform and 250 bp paired-end reads were generated





Bioinformatics analysis

DADA2 Amplicon Sequencing. Exactly.

qiime2

R

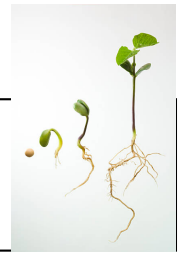
Vegan

Cytoscape

phyloseq

General analysis pipeline

Harvesting and extraction DNA microorganisms



DNA extraction and PCR

Bacterial 16S ribosomal DNA (rDNA) and fungal ITS2 region



Sequencing

NovaSeq platform and 250 bp paired-end reads were generated



Bioinformatics analysis

DADA2 Amplicon Sequencing. Exactly.

qiime2

R Vegan Cytoscape phyloseq

The four steps always used in bioinformatics

- Data loading
- Cleaning
- Analysis
- Visualisation

What does the data set look like after sequencing ?

Head 20_bk.effective.fastq

```
cmaslard@LEGcompute3:~/bioinformatic_tools_mcom_analyse/data/sequencing_data/bacteria_16S$ head 20_bk.effective.fastq
@20_bk A00783:1396:HKVK7DRX3:2:2101:25048:3004 1:N:0:TACGACGT+CCATGAAC
GTAGTCCACGCCGTAAACTATGGGTGCTAGCCGTCGGGGAGCATGCTCTTCGGTGGCGCAGCAAACGCATTAAGCACCCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAGGCAACGCGAAGAACCTTACCAGCCCTTGACTTCCCGATCGGGGGACCAGAGATGGACCTCTTCAATTGGTCGGTGAC
AGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGCCATTAGTTGCCATCATTAAGTTGGGCACTCTAGTGGGACTGCCGGTGATAAGCCGGA
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF,,F:FFFFFFFF
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFF::FFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFF
@20_bk A00783:1396:HKVK7DRX3:2:2101:16902:3020 1:N:0:TACGACGT+CCATGAAC
GTAGTCCACGCCGTAAACGATGGGCACTAGGTGTGGGAGGTGTCGACTCCTCCCGTGCCGGCGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTGGGCTTGACATGTAACTGACAGCCGTGGAAACACGGTCTTCCTTCGGGACAGTT
ACACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGTCCTATGTTGCCAGCATTAAGTTGGGGACTCATAGGAGACTGCCGGTGACAAACCGGA
+
FFFFFFFFFFFFFFFFF:FFFFFF:F:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFF::FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF,:FFFFFFFFF;FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF,FFFFFFFFFFFFFFF
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@20_bk A00783:1396:HKVK7DRX3:2:2101:24948:3051 1:N:0:TACGACGT+CCATGAAC
GTAGTCCACGCCGTAAACTATGGGTGCTAGCCGTCGGGGAGCATGCTCTTCGGTGGCGCAGCAAACGCATTAAGCACCCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAGGCAACGCGAAGAACCTTACCAGCCCTTGACTTCCCGATCGGGGGACCAGAGATGGACCTCTTCAATTGGTCGGTGAC
AGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGCCATTAGTTGCCATCATTAAGTTGGGCACTCTAGTGGGACTGCCGGTGATAAGCCGGA
cmaslard@LEGcompute3:~/bioinformatic_tools_mcom_analyse/data/sequencing_data/bacteria_16S$
```

Structure of a fastq file

Each *read* is always composed of **four consecutive lines**:

Head 20_bk.effective.fastq

```
cmaslard@LEGcompute3:~/bioinformatic_tools_mcom_analyse/data/sequencing_data/bacteria_16S$ head 20_bk.effective.fastq
@20_bk A00783:1396:HKVK7DRX3:2:2101:25048:3004 1:N:0:TACGACGT+CCATGAAC
GTAGTCCACGCCGTAAACTATGGGTGCTAGCCGTCGGGGAGCATGCTCTTCGGTGGCGCAGCAAACGCATTAAGCACCCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAGGCAACGCGAAGAACCTTACCAGCCCTTGACTTCCCGATCGGGGGACCAGAGATGGACCTCTTCAATTGGTCGGTGAC
AGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGCCATTAGTTGCCATCATTAAGTTGGGCACTCTAGTGGGACTGCCGGTGATAAGCCGGA
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF,,F:FFFFFFFF
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFF::FFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFF
@20_bk A00783:1396:HKVK7DRX3:2:2101:16902:3020 1:N:0:TACGACGT+CCATGAAC
GTAGTCCACGCCGTAAACGATGGGCACTAGGTGTGGGAGGTGTCGACTCCTCCCGTGCCGGCGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTGGGCTTGACATGTAACTGACAGCCGTGGAAACACGGTCTTCCTTCGGGACAGTT
ACACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGTCCTATGTTGCCAGCATTAAGTTGGGGACTCATAGGAGACTGCCGGTGACAAACCGGA
+
FFFFFFFFFFFFFFFFF:FFFFFF:F:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF::FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF,:FFFFFFFFF;FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF,FFFFFFFFFFFFFF
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@20_bk A00783:1396:HKVK7DRX3:2:2101:24948:3051 1:N:0:TACGACGT+CCATGAAC
GTAGTCCACGCCGTAAACTATGGGTGCTAGCCGTCGGGGAGCATGCTCTTCGGTGGCGCAGCAAACGCATTAAGCACCCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAGGCAACGCGAAGAACCTTACCAGCCCTTGACTTCCCGATCGGGGGACCAGAGATGGACCTCTTCAATTGGTCGGTGAC
AGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGCCATTAGTTGCCATCATTAAGTTGGGCACTCTAGTGGGACTGCCGGTGATAAGCCGGA
cmaslard@LEGcompute3:~/bioinformatic_tools_mcom_analyse/data/sequencing_data/bacteria_16S$
```

Structure of a fastq file

Each *read* is always composed of **four consecutive lines**:

Lines	Content	Example
1	**Header line** — the read identifier	@20_bk A00783:1396:HKVK7DRX3:2:2101:25048:3004 1:N:0:TACGACGT+CCATGAAC
2	**Nucleotide sequence** (A, T, C, G)	GTAGTCCACGCCGTAAACTATGGGTGCTAGCCG…
3	**Separator line** — starts with a + symbol (sometimes repeats the ID)	+
4	**Quality scores** — one ASCII character per base	FFFFFFFFFFFFFFFFFFFF::FFFFFFFFFF…

- @20_bk → sample name (matches your file name)
- A00783 → Illumina machine ID
- 1396:HKVK7DRX3:2:2101:25048:3004 → flowcell and cluster coordinates (unique read ID)
- 1:N:0: → indicates **read 1** (from paired-end sequencing), not filtered (N)
- TACGACGT+CCATGAAC → dual-index barcodes used for multiplexing

Find in the literature a tool for analysing this type of data

QIIME2 « chaïm two »

QIIME 2 is a bioinformatics platform used to analyze **microbial community sequencing data,** like your **16S** (bacteria) and **ITS** (fungi) datasets.

It takes raw FASTQ files and lets you:
•clean and denoise reads (DADA2),
•assign taxonomy (SILVA or UNITE),
•calculate diversity (alpha/beta),
•and create interactive plots (barplots, PCoA, etc.).

In short, QIIME 2 helps you go **from raw sequences to biological interpretation,** identifying which microbes are present and how communities differ between your samples.

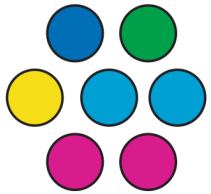
Let us use the results from QIIME2

Alpha diversity

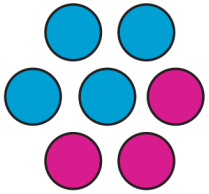
Alpha Diversity:

Refers to the diversity of organisms sharing the same community or habitat. A combination of species richness and eveness is used to represent diversity within a community or habitat

Alpha diversity: biodiversity within a sample



High alpha diversity

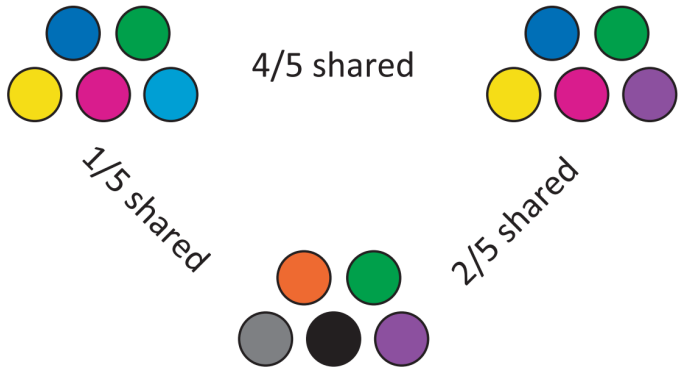


Low alpha diversity

Beta diversity

Beta Diversity:
Indicates diversity between communitites. Species frequently change when habitat or community changes

Beta diversity: similarity between samples



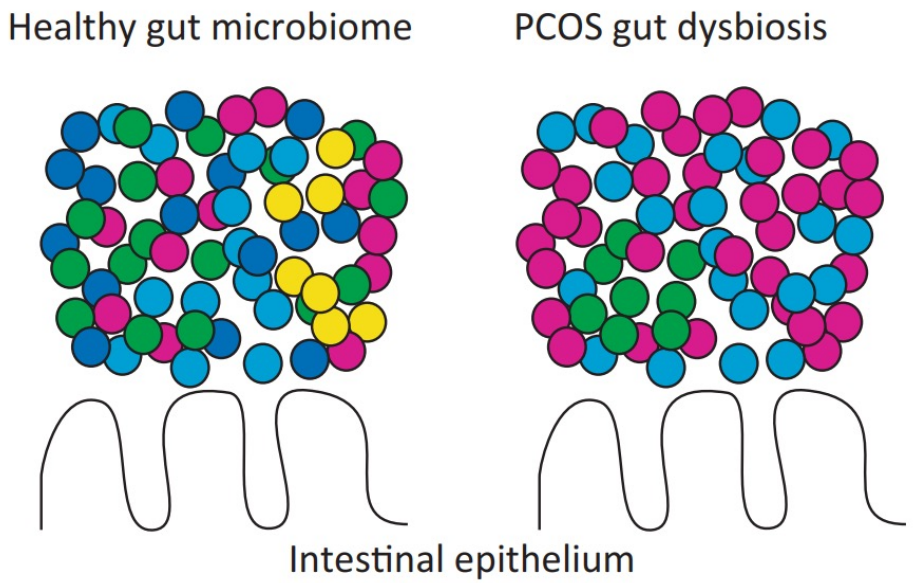
Alpha and Beta diversity

Alpha Diversity:

Refers to the diversity of organisms sharing the same community or habitat. A combination of species richness and eveness is used to represent diversity within a community or habitat

Beta Diversity:

Indicates diversity between communitites. Species frequently change when habitat or community changes



In short, alpha diversity measures diversity at a specific location, while beta diversity measures the variation in diversity between several locations.

Are the microbial communities the same in each compartment?

Which indicators should be used? And

How will we analyse these data?

The four steps, four tools

- Data loading → **readr**: read_csv()
- Cleaning → **dplyr** and **tidyr**: filter(), mutate(), pivot_longer()
- Analysis → **R stats** and **broom**: t.test(), lm(), anova(), TukeyHSD()
- Visualisation → **ggplot2**: ggplot()
- Around everything: an **R project** to organise, a **Quarto** report to share

Where you work today

- LEGcompute → RStudio Server, in your web browser
- **Console**: run one line of R
- **Terminal**: the shell you know from Sessions 1-3 (mkdir, cp, ls)
- **Files** pane: your project folder
- **Render** button: turns your workbook into one HTML report

A clean R project

```
session_04/
 ├── session_04.Rproj
 ├── WORKBOOK.qmd    → your report
 ├── data/           → raw data, never edited by hand
 ├── img/            → images not made by R
 └── results/        → everything your code produces
```

- Raw data is read-only: every change happens in code
- Paths are relative: data/soybean_plants.csv works on any computer
- If it renders, it is reproducible

Quarto: text, code and figures in one file

- Text is written normally, around the code
- Code goes in **chunks**: ```{r} … ```
- Insert an image in one line: ![Caption](img/experimental_design.png){#fig-design width=80%}
- Cite it in your text with @fig-design → "Figure 1"
- **Render** → R runs everything from top to bottom → one HTML file to submit

A tidy table

- One **row** = one observation (one plant, one sample)
- One **column** = one variable (genotype, water, heat, biomass…)
- NA = "not available": nothing was measured there
- Today: two tables → soybean_plants.csv: 240 plants, their place, their job, their biomass → soybean_alpha_diversity.csv: 195 bacterial samples, their diversity

dplyr: verbs for tables

- filter() → keep some rows
- mutate() → add or change a column
- count() → how many rows per group
- group_by() + summarise() → one summary per group
- The pipe %>% reads "and then" (Ctrl/Cmd + Shift + M)

```
physio %>%
  group_by(water, heat) %>%
  summarise(mean_biomass = mean(biomass_total))
```

tidyr: wide → long

- **Wide**: one row per plant, one column per organ → biomass_leaf | biomass_stem | biomass_root
- **Long**: one row per plant *and* organ → organ | biomass
- pivot_longer() goes from wide to long, pivot_wider() goes back
- Ask yourself: "what should one row be?"

ggplot2: a plot in layers

- **Data**: which table → ggplot(physio, …)
- **Aesthetics**: which column goes where → aes(x = water, y = biomass_total, colour = heat)
- **Geometry**: how to draw it → geom_boxplot(), geom_point(), geom_tile()
- **Facets**: one panel per group → facet_wrap(~ genotype)
- Layers are added with +, not %>%

Which test, when?

Assumptions: residuals roughly normal, similar spread in all groups. Check them with plot(model).

Question	Test	In R	If assumptions fail
2 groups differ?	t-test	t.test()	wilcox.test()
3+ groups differ?	ANOVA	anova(lm())	kruskal.test()
Which groups differ?	Tukey HSD	TukeyHSD(aov())	pairwise.wilcox.test()
Several factors, interactions?	Factorial ANOVA	anova(lm(y ~ a * b))	transform y (log)
Two measurements linked?	Correlation, regression	cor.test(), lm(y ~ x)	Spearman

A t-test is a linear model

- t.test(y ~ group) and lm(y ~ group) give **the same p-value**
- In lm(): → the intercept = the mean of the reference group → the estimate = the difference between the two groups
- Add more factors → lm(y ~ water * heat * genotype)
- * = each factor **and** their interaction: does the effect of drought depend on the heatwave?

Significant is not the same as important

- **p-value**: could this difference appear by chance alone?
- **Share of variation**: how much does this factor actually explain?
- Plant biomass: → drought: $p < 0.001$, **78%** of the variation → heatwave: $p = 0.001$, **2%** → cultivar: $p = 0.002$, **2%**
- All three are significant, but only one really matters
- Always look at both, and at the plot

Today's workbook

- **A**: build the project, render once
- **B**: meet the experiment through the plants (map, biomass)
- **C**: the bacteria: first look, then the rhizoplane
- **D**: the statistics, guided
- **E**: the bulk soil, on your own (**graded**)
- Submit WORKBOOK.html on Moodle **by Thursday**