Raising awareness among plant virologists on the richness of their high-throughput sequencing data
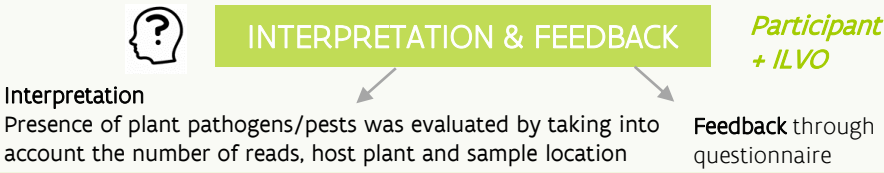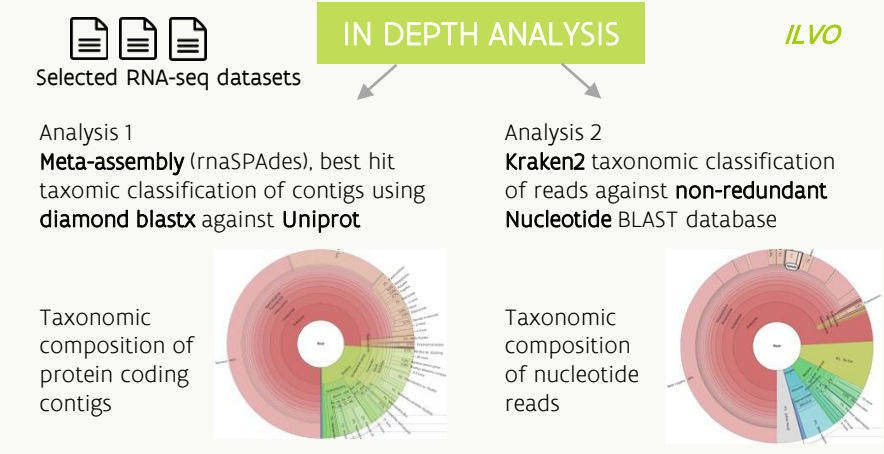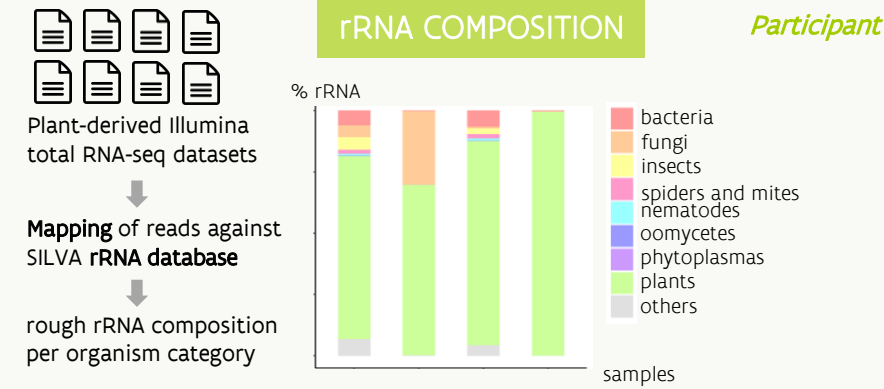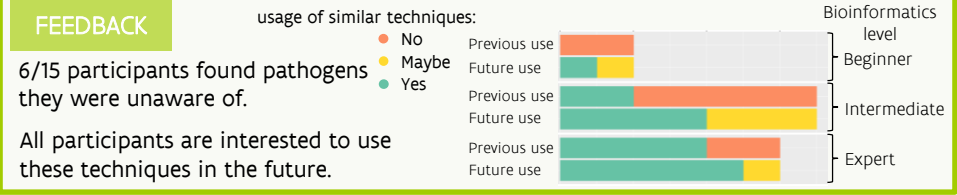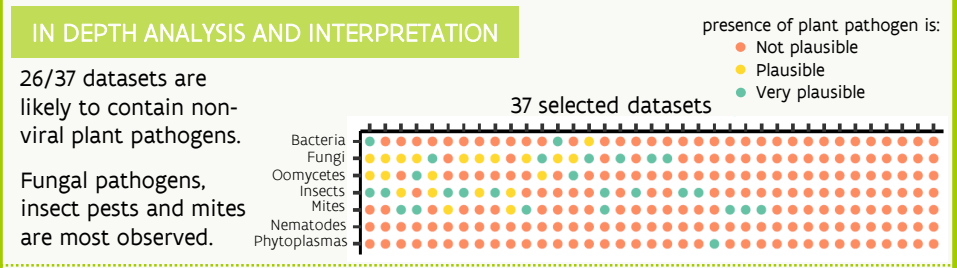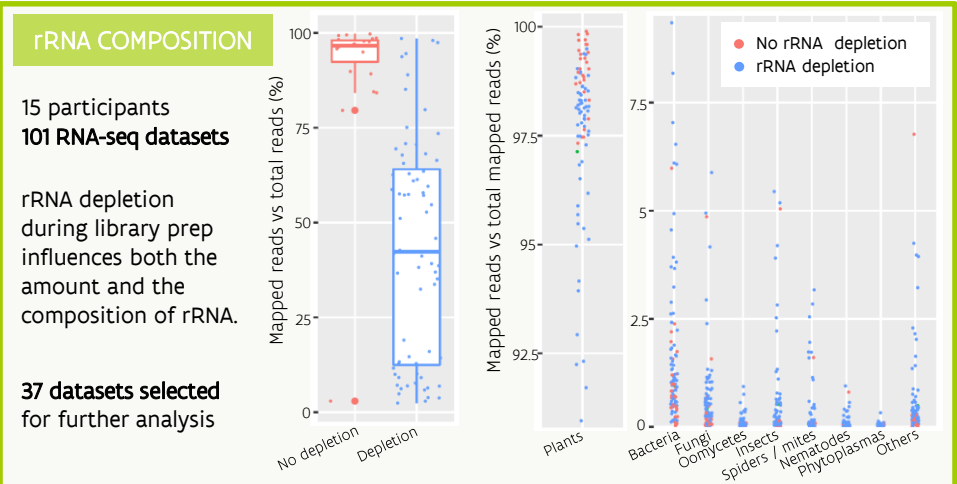
Intro

? Virologists typically compare high throughput sequencing data with plant virus databases. But do they ever check the presence of other possible pathogens or pests?

💡 Let's organize an "**RNA-seq community effort**" to see what other plant pathogens we can find in existing datasets!

Approach

rRNA COMPOSITION *Participant*

Plant-derived Illumina total RNA-seq datasets

⬇

Mapping of reads against SILVA **rRNA database**

⬇

rough rRNA composition per organism category

% rRNA

- bacteria
- fungi
- insects
- spiders and mites
- nematodes
- oomycetes
- phytoplasmas
- plants
- others

samples

IN DEPTH ANALYSIS *ILVO*

Selected RNA-seq datasets

Analysis 1
Meta-assembly (rnaSPAdes), best hit taxomic classification of contigs using **diamond blastx** against **Uniprot**

Analysis 2
Kraken2 taxonomic classification of reads against **non-redundant** Nucleotide BLAST database

Taxonomic composition of protein coding contigs

Taxonomic composition of nucleotide reads

INTERPRETATION & FEEDBACK *Participant + ILVO*

Interpretation
Presence of plant pathogens/pests was evaluated by taking into account the number of reads, host plant and sample location

Feedback through questionnaire

Results

rRNA COMPOSITION

15 participants
101 RNA-seq datasets

rRNA depletion during library prep influences both the amount and the composition of rRNA.

37 datasets selected for further analysis

- No rRNA depletion
- rRNA depletion

Mapped reads vs total reads (%)

Mapped reads vs total mapped reads (%)

No depletion / Depletion

Plants, Bacteria, Fungi, Oomycetes, Insects, Spiders / mites, Nematodes, Phytoplasmas, Others

IN DEPTH ANALYSIS AND INTERPRETATION

26/37 datasets are likely to contain non-viral plant pathogens.

Fungal pathogens, insect pests and mites are most observed.

presence of plant pathogen is:
- Not plausible
- Plausible
- Very plausible

37 selected datasets

Bacteria
Fungi
Oomycetes
Insects
Mites
Nematodes
Phytoplasmas

FEEDBACK

6/15 participants found pathogens they were unaware of.

All participants are interested to use these techniques in the future.

usage of similar techniques:
- No
- Maybe
- Yes

Bioinformatics level

Beginner — Previous use / Future use
Intermediate — Previous use / Future use
Expert — Previous use / Future use

Annelies Haegeman[1], Yoika Foucart[1], Thomas Goedefroit[1], Kris De Jonghe[1], Maher Al Rwahnih[2], Thierry Candresse[3], Yahya Gaafar[4], Oscar Hurtado-Gonzales[5], Zala Kogej[6,7], Denis Kutnjak[6], Janja Lamovšek[8], Marie Lefebvre[3], Martha Malapi-Wight[9], Irena Mavrič Pleško[8], Serkan Onder[10], Jean-Sébastien Reynard[11], Ferran Salavert Pamblanco[12], Olivier Schumpp[11], Kristian Stevens[2], Chandan Pal[13], Lucie Tamisier[14], Çiğdem Ulubaş Serçe[15], Inge Van Duivenbode[16], David Waite[13], Hu Xiaojun[5], Heiko Ziebell[4] & Sébastien Massart[14]

[1] ILVO, Belgium; [2] UCDavis, FPS, USA; [3] INRAE, UMR1332, France; [4] JKI, Germany; [5] USDA-APHIS, PGQP, USA; [6] NIB, Slovenia; [7] MPS, Slovenia; [8] KIS, Slovenia; [9] USDA-APHIS, BRS, USA; [10] ESOGU, Turkey; [11] Agroscope, Switzerland; [12] NU, UK; [13] MPI, New Zealand; [14] ULiège, Belgium, [15] NOHU, Turkey, [16] NAK, Netherlands

ILVO
Institute for Agricultural and Fisheries Research

federal public service
HEALTH, FOOD CHAIN SAFETY AND ENVIRONMENT

Euphresco

annelies.haegeman@ilvo.vlaanderen.be