Netherlands Food and Consumer
Product Safety Authority
Ministry of Economic Affairs

Comparison of NGS pipelines and traditional diagnostics in annual *Daucus carota* surveys

B.T.L.H. van de Vossenberg, M. Botermans, L. Tjou-Tam-Sin, A. Roenhorst, M. Bergsma-Vlami, M. Westenberg

Bari, Italy
November 2017

Overview presentation

- Phytosanitary survey in carrot
- Traditional diagnostics
- Next Generation Sequencing approach
  - Selection of reference genes
  - Reference based detection
  - *De novo* and blast based detection
- Comparison of costs and hands-on time

Phytosanitary survey *Daucus carota* (carrot)

- Annual survey in carrot since 2011

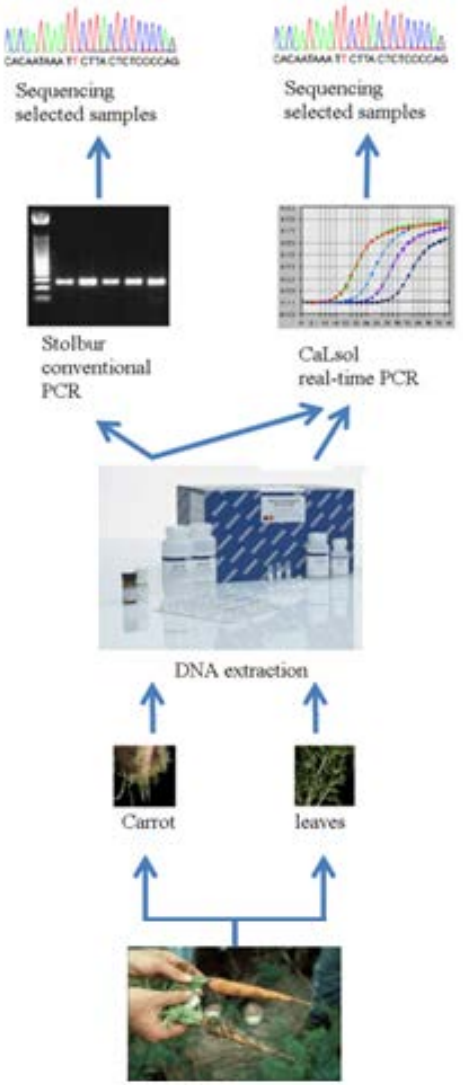
- Presence of
  - *Ca*. Liberibacter solanacearum (CaLsol) (EPPO A1)
  - *Ca*. Phytoplasma solani (CaPsol) (EU II/A2)

Symptomatic material

- ~130 inspections resulting in ~30 samples
  - Presently, both pests have not been detected
- Sampling of symptomatic field-grown carrots
  - Discolored leaves (red, yellow)
  - Stunted growth
  - Formation of side roots
- Symptoms not specific to both pests
  - *Ca* P. asteris suspected causal agent in majority of cases

Diagnostic testing scheme

- Two subsamples (leaves & carrot)
- Detection by:
  - Conventional PCR CaPsol (leaves)
  - Real-time CaLsol (leaves & carrot)
- Verification for selected samples
  - targeted PCR Sanger sequencing

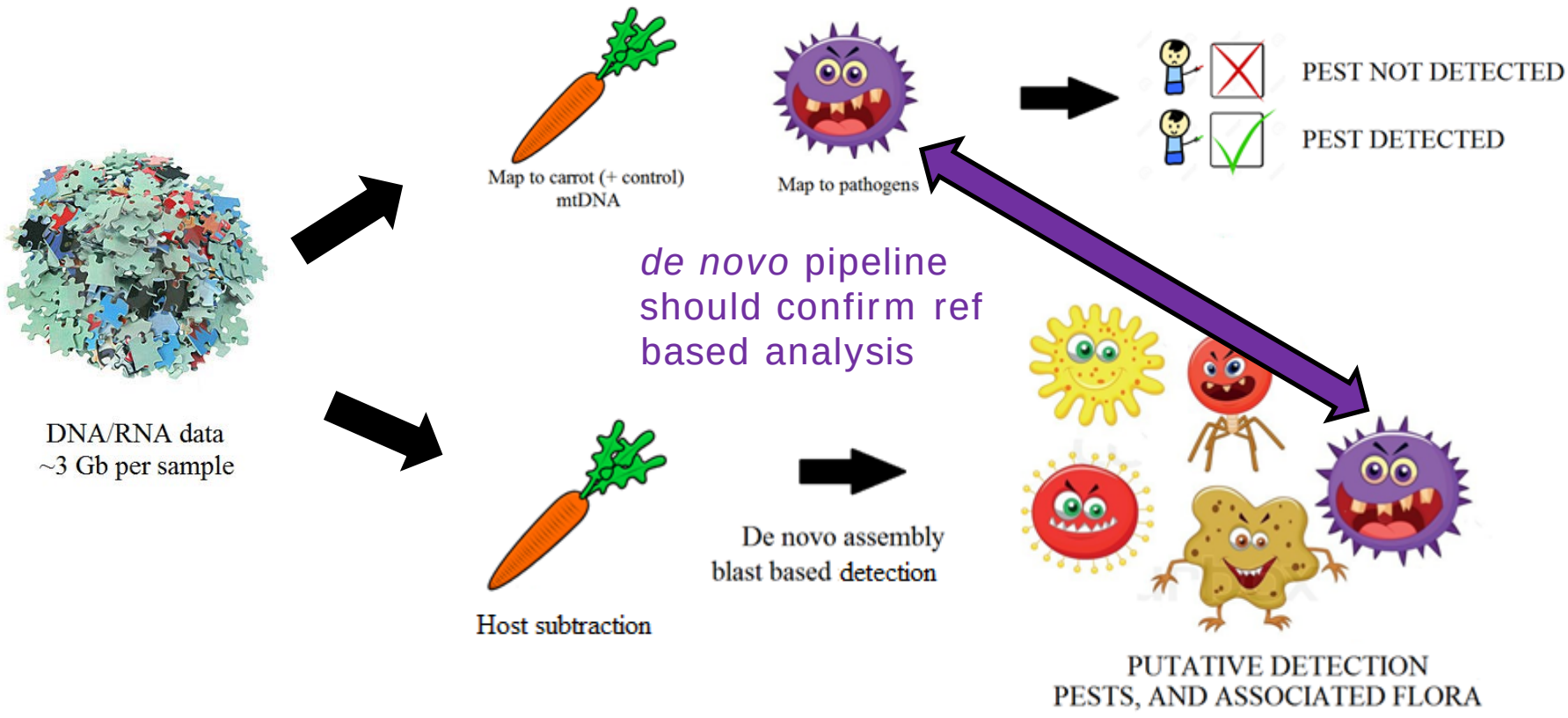


The iterative use of test methods:
1. is time consuming
2. requires lots of hands-on time
3. is therefore costly

Next Generation Sequencing (NGS); *an alternative?*

- Why using the *Daucus carota* survey?
  - Specific scope: 2 pests in symptomatic material (analytical sensitivity)
  - Availability of reference sequences (analytical specificity)
  - Survey shared by multiple disciplines
  - “Long” turn-over time allowed

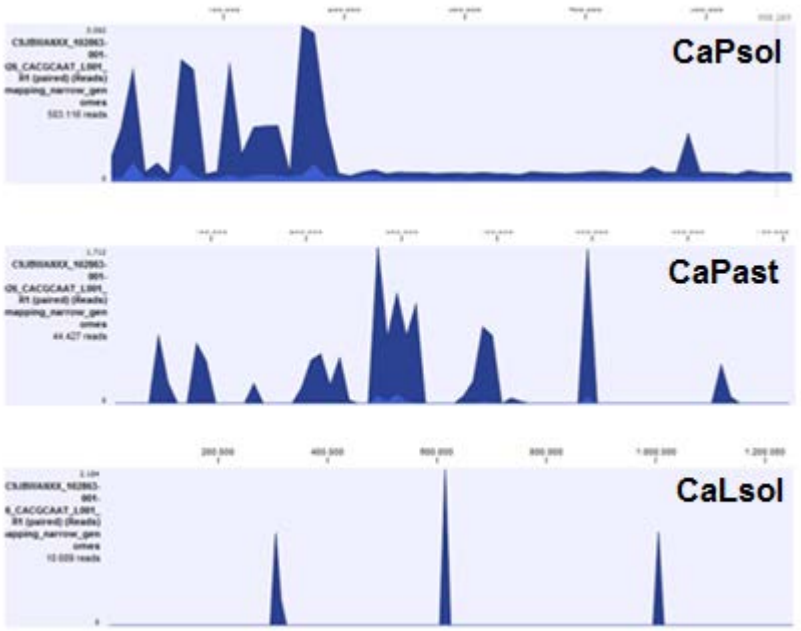
Analysis pipelines – reference vs. *de novo*



Defining suitable reference sequences

Entire genome not suitable for detection

- CaPsol, CaLsol and CaPast genomes share homology (non-specific mapping)
- Regions with variable resolution (non-species level resolution)
- Determining cut-offs for detection using the entire genome is not possible

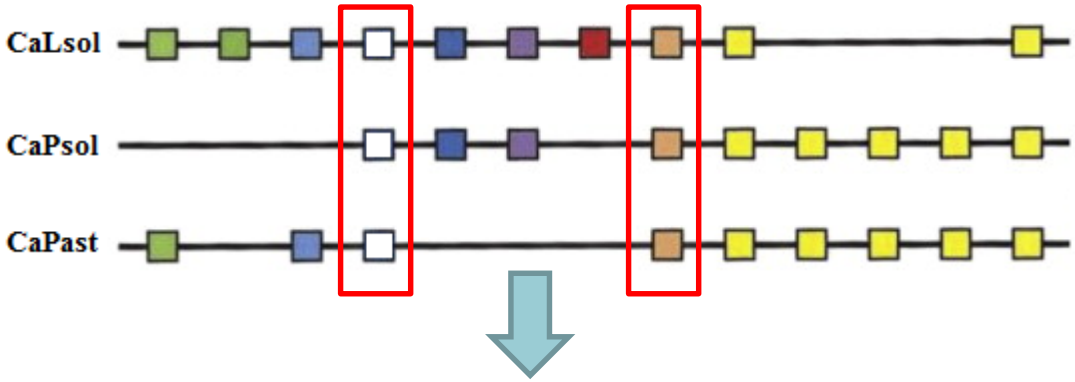


CaPsol sequence data mapped to CaPsol, CaPast and CaLsol

Selection of reference genes

Suitable reference genes are:

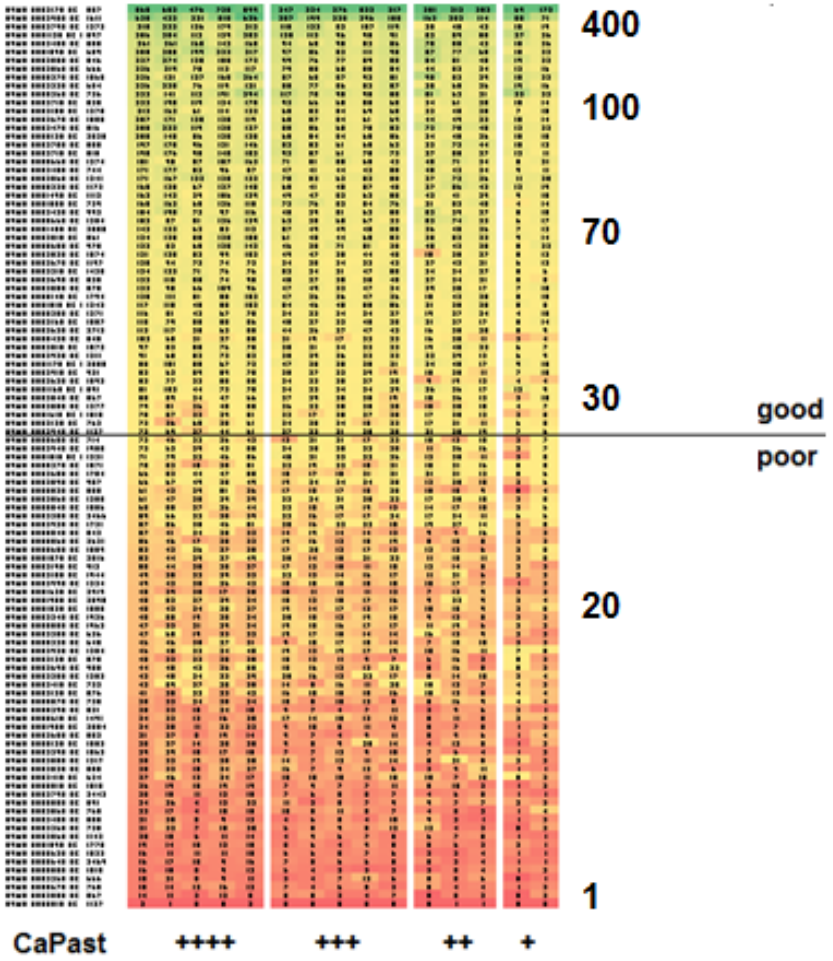
- Single copy orthologs (SCO)
  - Even coverage expected
  - Can be compared over species
- SCO with species level resolution
- SCO >500 nt for reliable mapping



107 reference genes per species

Not all 107 species specific SCOs can be used for RNAseq pipeline

- Selected reference genes for DNA pipeline are not equally transcribed
- Transcription level per SCO is conserved over samples
  - Highly expressed gene in sample 1 = highly expressed gene in sample 2
- 51 species specific SCOs with at least >5x average coverage in individual samples were selected for RNAseq pipeline



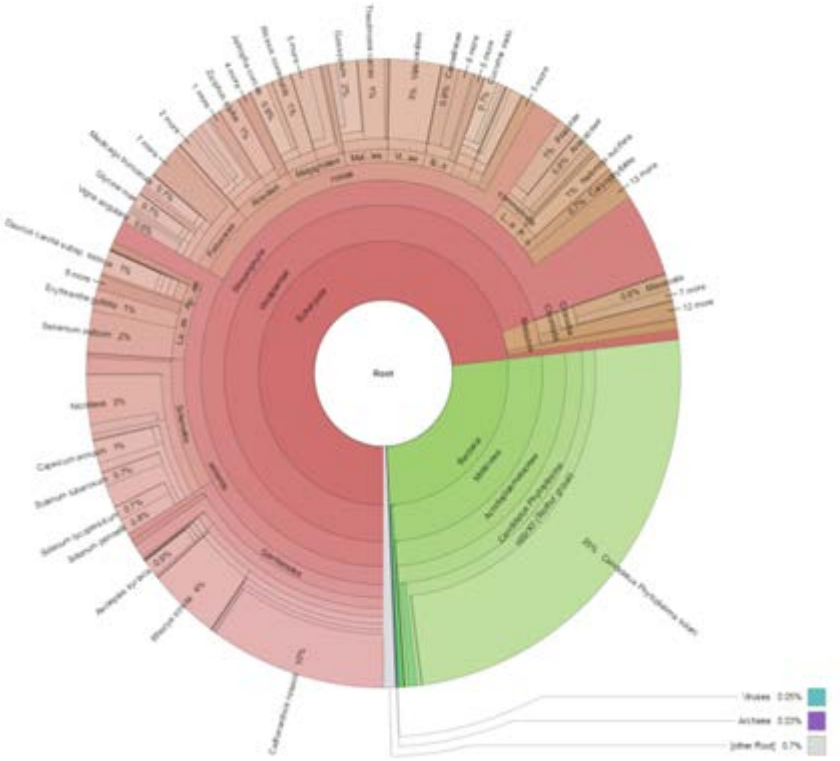
Reference assembly results – DNA and RNA

LIMS number	Traditional diagnostics		DNA NextSeq			RNA NextSeq		
	CaLsol/CaPsol detected	Suspected causal agent	CaLsol	CaPsol	CaPast	CaLsol	CaPsol	CaPast
5727954	NO	putative CaPast	NO	NO	+	NO	NO	+
5727962	NO	CaPast	NO	NO	+	NO	NO	+
5727970	NO	CaPast	NO	NO	+	NO	NO	+
5727989	NO	putative CaPast	NO	NO	+	NO	NO	+
5727997	NO	putative CaPast	NO	NO	+	NO	NO	+
5728009	NO	putative CaPast	NO	NO	+	NO	NO	+
5728228	NO	putative CaPast	NO	NO	+	NO	NO	+
5728244	NO	CaPast	NO	NO	+	NO	NO	+
5728367	NO	unknown	NO	NO	NO	NO	NO	NO
5728375	NO	CaPast	NO	NO	+	NO	NO	+
5728391	NO	CaPast	NO	NO	+	NO	NO	+
5728420	NO	putative CaPast	NO	NO	+	NO	NO	+
5728439	NO	putative CaPast	NO	NO	+	NO	NO	+
5728447	NO	unknown	NO	NO	NO	NO	NO	NO
5728455	NO	putative CaPast	NO	NO	+	NO	NO	+
5728463	NO	unknown	NO	NO	NO	NO	NO	NO
5728471	NO	CaPast	NO	NO	+	NO	NO	+
5974066	NO	putative CaPast	NO	NO	+	NO	NO	+
5974074	NO	unknown	NO	NO	NO	NO	NO	NO
6792299	NO	unknown	NO	NO	NO	NO	NO	NO

- Identical qualitative results were obtained from the DNA and RNA detection pipelines
- Pipeline output could easily be interpreted

de novo assembly + blast-based detection

- Beyond the initial scope of the survey
- When the usual suspects cannot be detected, are there other possible causal agents that could explain the symptoms observed?
- Blast-based detection: indicative and use with caution!
- Interactive visualisation tool: Krona



Possible candidates for observed symptoms

- Carrot viruses detected:
  - Carrot torrado virus 1
  - Carrot cryptic virus
  - Carrot mottle virus
  - Carrot red leaf luteovirus associated RNA
  - Carrot read leaf virus

- Carrot read leaf virus and Carrot mottle virus were detected in all CaPast negative samples (#5) and possibly causing the observed symptoms of Carrot motley dwarf (CMD)

What about the money?

- Direct costs are higher per sample
- Hands-on time is greatly reduced per sample
  - Traditional: 57 min/sample
  - NGS: 21 min/sample
- Net extra costs per sample: €89
  - Saving in hands-on time
  - Re-usable datasets
  - Possibilities for detection beyond initial scope of survey

Traditional diagnostics	costs per sample
DNA extraction (Qiagen, plant mini kit)	€ 8,00
Bio-X-ACT short mix reactions	€ 2,00
TaqMan universal Master Mix	€ 9,00
Qiaquick PCR Purification Kit	€ 2,00
BigDye terminator kit v1.1	€ 21,00
DyeEx kit	€ 6,00
Pop7 polymer	€ 1,00
Capillary	€ 2,00
Total	€ 51,00

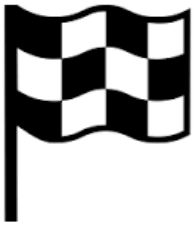
NGS approach (DNA)	costs per sample
DNA extraction (Qiagen, plant mini kit)	€ 4,00
NGS NextSeq, DNA, PE150, 4.5 Gb data	€ 320,00
Total	€ 324,00

Conclusions

- We created a robust and reliable detection pipeline for CaPsol, CaLsol and CaPast detection in symptomatic carrot material

Costs

NGS: +€89 per sample



Turnover time

NGS: similar (~3 weeks)

Hands-on time

NGS: less than half (21 vs 57 minutes)

Results

Data can be used for:

1. specific detection of survey targets
2. detection beyond initial scope of survey
3. Additional analyses (e.g. track & trace)

Future work

- Create scripts for automated generation of result forms including results and conclusions for the different analyses
  - User-friendly, interactive, but stand-alone and write protected (QA)
  - In close collaboration with specialists from different disciplines

- Determine performance criteria following PM7/98(2) for the analysis pipelines and compare those to traditional tests

- Increase computational power and storage for NGS data

Acknowledgements

NPPO-NL

Jeroen van de Bilt (Bacteriology)
Ko Verhoeven (Virology)
Lucas van der Gouw (Molecular biology)
Maud Buimer (Molecular biology)
Joris Voogd (Molecular biology)
Maureen Bruil (Molecular biology)

Valencian Institute for Agricultural Research (IVIA)

Mariano Cambra

Wageningen UR

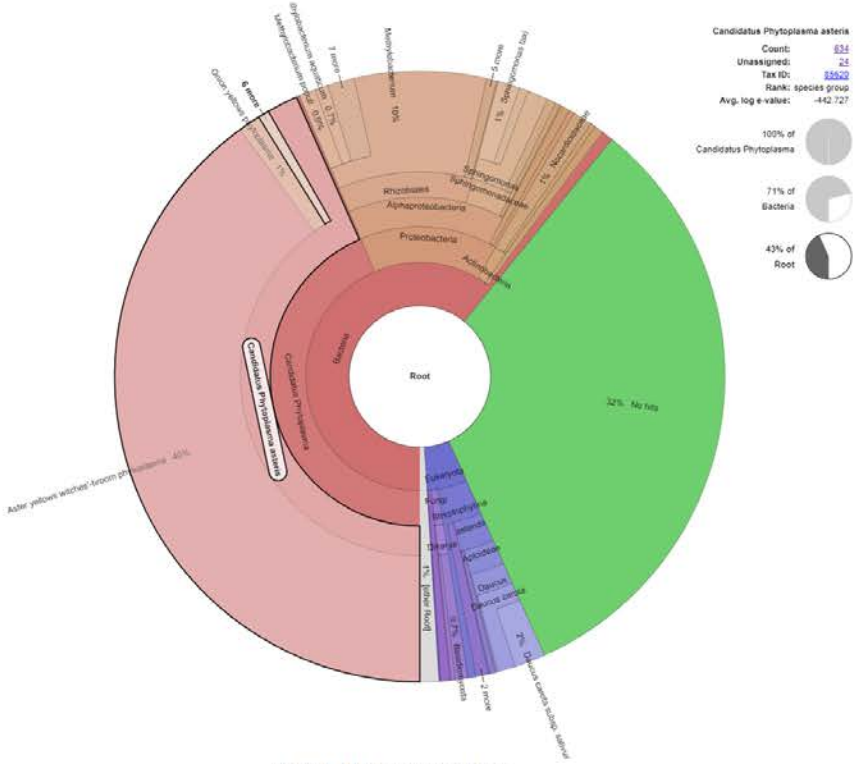
Henri van der Geest, Sven Warris

Thank you for your attention

RNAseq is not properly mapped to the host

***de Novo* assembly and blast-based detection**

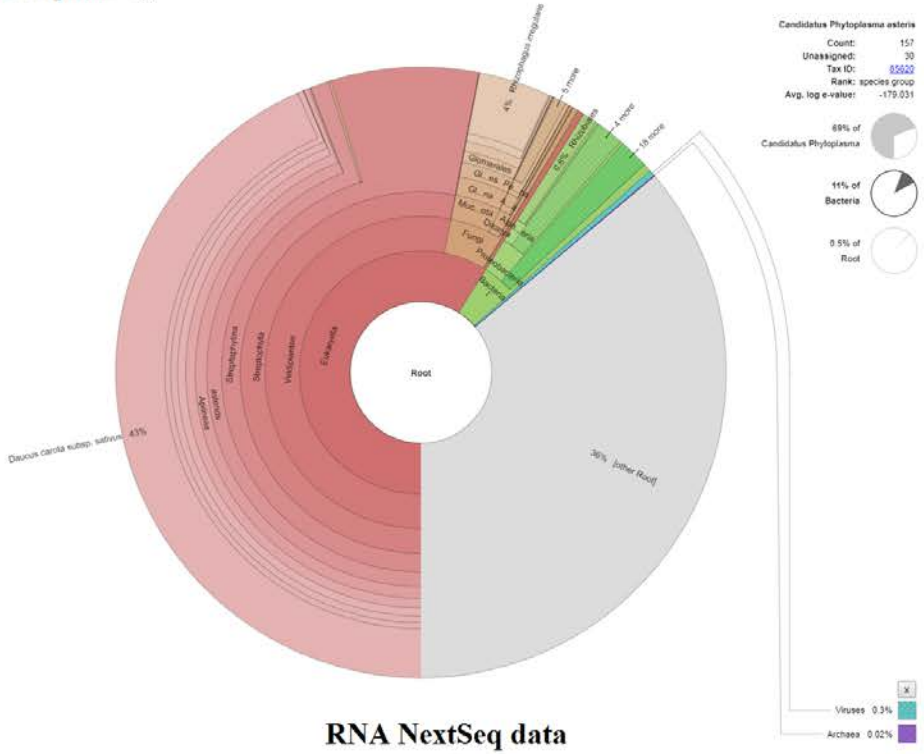
Sample_01 (CaPast positive)



DNA NextSeq data

1463 elements

585 CaPast hits **(40% of total)**



RNA NextSeq data

30,080 elements

157 CaPast hits **(0.5% of total)**

major groups: plant (53%)
unknown (36%)