Genome-based pest risk assessment for newly discovered plant viruses: Nazca lines or Rosetta stone?

Prof. Sébastien Massart

Plant Pathology Laboratory – TERRA
Gembloux Agro-Bio Tech – University of Liège
Belgium

sebastien.massart@uliege.be

@Be_Phytopath

Introduction

The genome era

Sequencing price ?

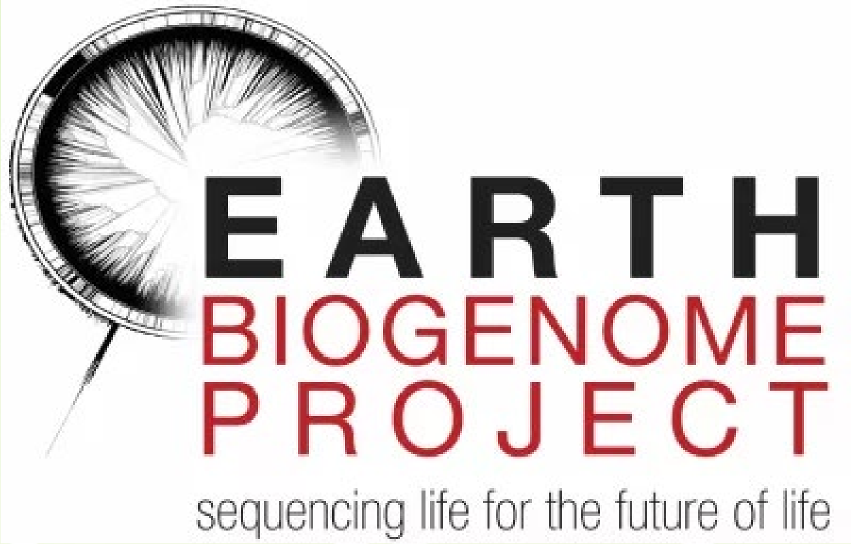
CLEARANCE SALES

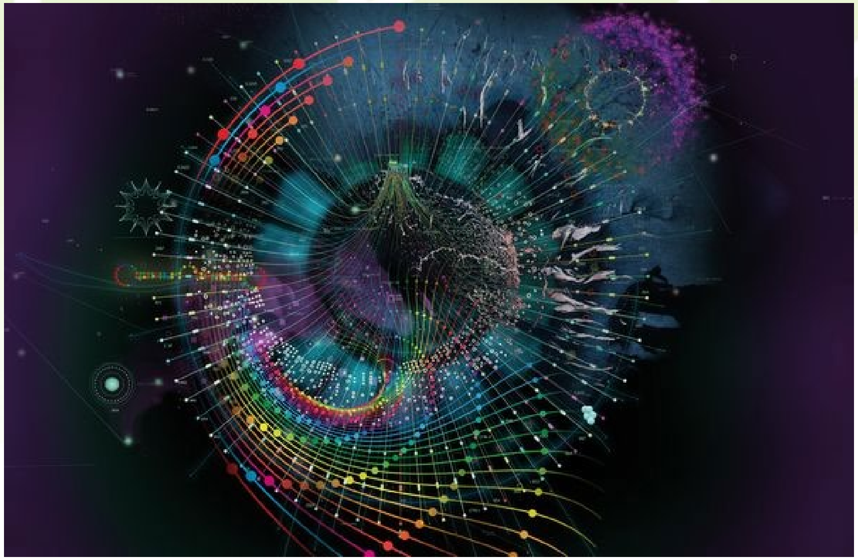
99.99999 % OFF & Better quality

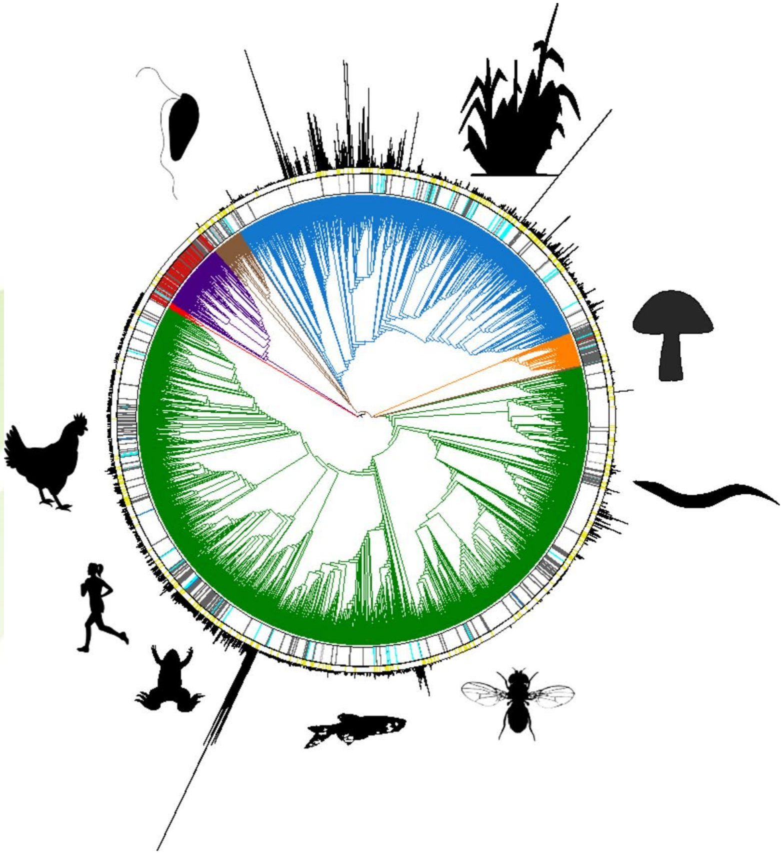
amazon

Pay only 0.00001 % of original genome price

What are doing the scientists ?







LIÈGE université
Gembloux
Agro-Bio Tech

LABORATORY OF
Phytopathology

Metagenome sequencing & Plant Health – Belgian case

service public fédéral
SANTE PUBLIQUE,
SECURITE DE LA CHAINE ALIMENTAIRE
ET ENVIRONNEMENT

Crops: potato, tomato, eggplant, pepper..

⇕

Wild: *black nightshade...*

⇕

Ornamental: petunia, datura…

SEVIPLAN

↓

National project to scan the virome of commodities

↓

17,000 samples of Solanaceae

↓

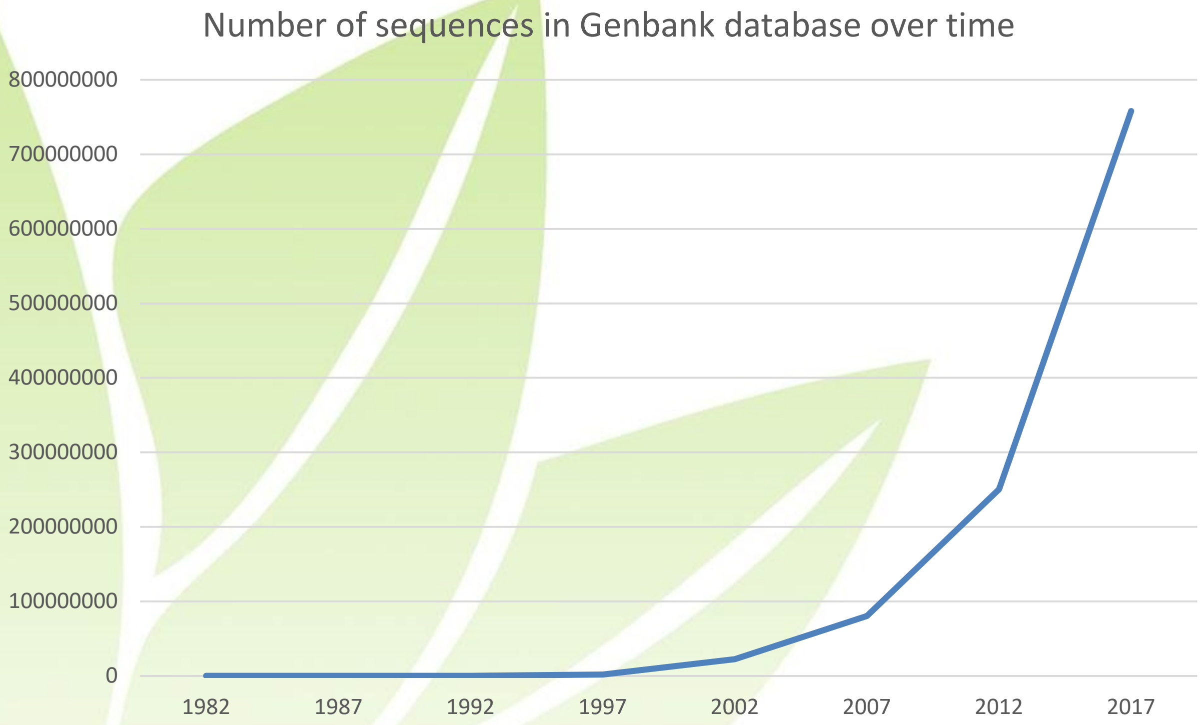
High Throughput Sequencing

↓

“Solanaceae Virus Wikipedia”

LIÈGE université
Gembloux
Agro-Bio Tech

And the databases ?



Bioinformatics

Euphresco - **PHBN project (15 partners + 13 associated partners)**
Coordination: A. Haegemans (ILVO)

The Plant Health Bioinformatic Network

1. Develop training material
2. Bioinformatics challenge on HTS data
3. Data mining

Seed for networking the new stakeholder

Pest genomes ? Known Pests


virus > bacteria > fungi

Plum Pox Virus (A2)

- 271 full genomes
- 2,200 sequences

Potato virus Y : - 443 full genomes
- 3,300 sequences

Ralstonia solanacearum (A2)

- 102 full genomes
- 42,000 sequences

Genomes of known viruses

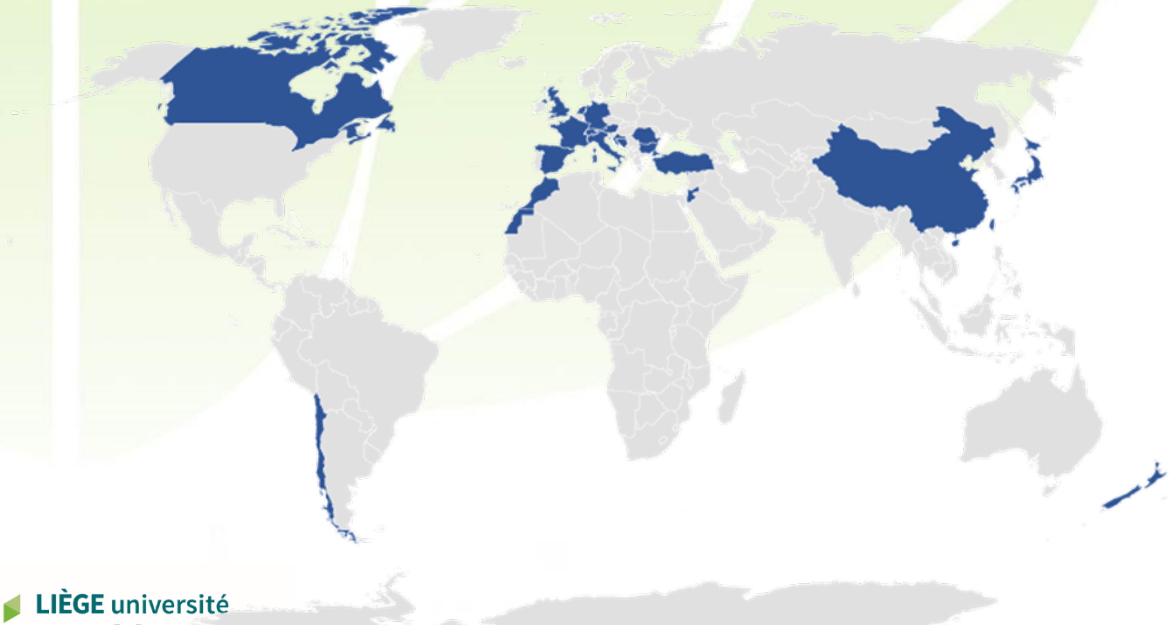
(& also other pests)

Genomes by tens - EURAVELCH

Euphresco - EURAVELCH project (7 partners + 14 collaborating countries)
Coordination: K. De Jonghe (ILVO)

Unravelling the Little Cherry disease complex at European scale to improve transnational diagnostics and management of the disease

Diagnostic Improvement & comparing EU/non EU isolates by genome sequencing



由必应提供支持
© GeoNames, HERE, MSFT, Microsoft, NavInfo, Wikipedia

94 samples
22 countries
5 continents

Genomes by tens - EURAVELCH

BEFORE

Species	# genomes
LChV1	16
LChV2	5

AFTER

Species	# genomes
LChV1	51 **(X3)**
LChV2	33 **(X6)**

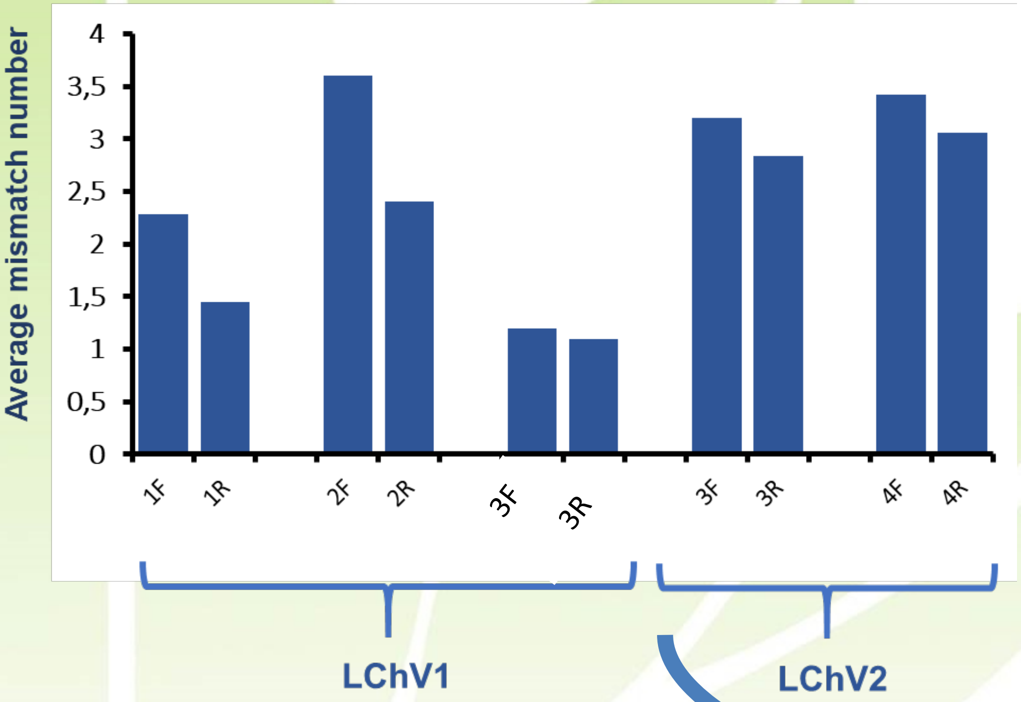
NCBI National Center for Biotechnology Information

LIÈGE université Gembloux Agro-Bio Tech

LABORATORY OF Phytopathology

Genomes by tens - EURAVELCH

Genomes for diagnostic primers



- ✓ Less variable regions genome-wide
- ✓ Design of primers

LChV1 detection primers

LChV1 CP sequences from our Euravelch

LChV1 CP sequences from NCBI

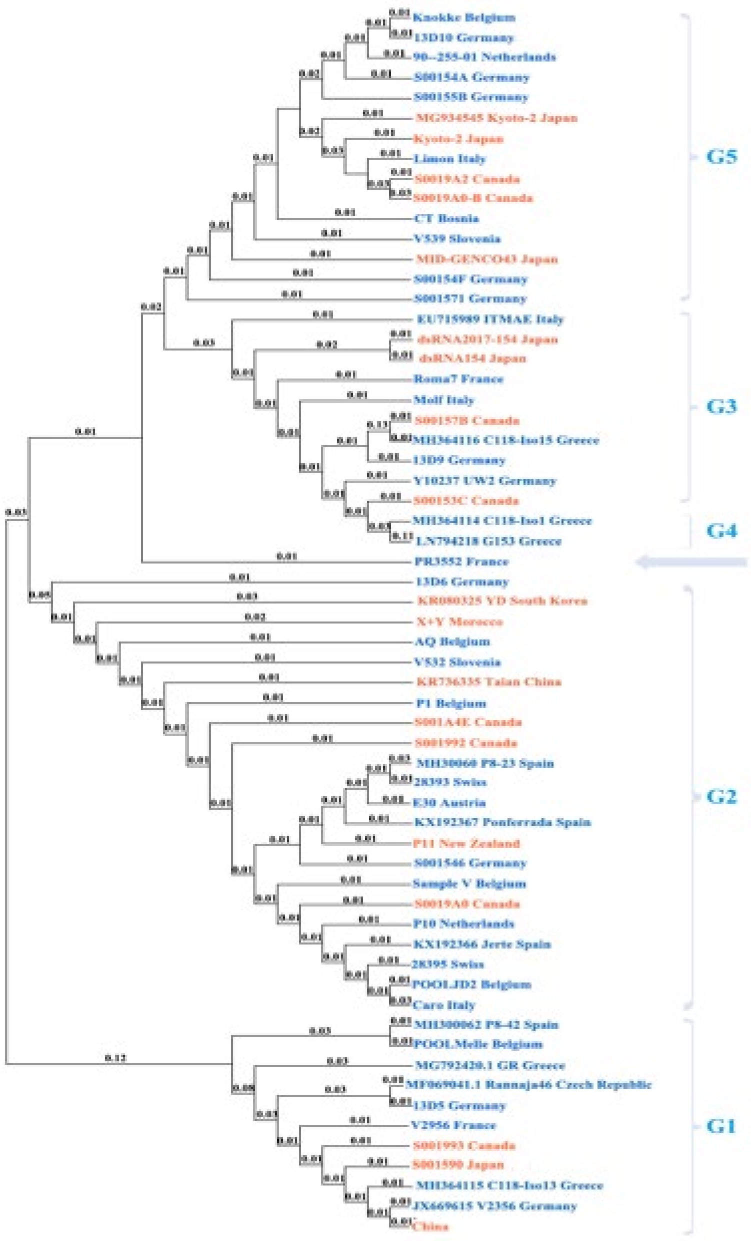
```
LCh-C:       TATTAGGACCACGTCGCGGT
LCh-D:       TATTGGGACCACGTCGAGGT
LCh-DR:      TGTTAGGACCACGTCGTGGT
Sequence1:   TATTGGGACCACGTCGCGGT  1
Sequence2:   TATTGGGACCACGTCGCGGT  1
Sequence3:   TATTAGGACCACGTCGCGGT  0
Sequence4:   TATTAGGACCACGTCGCGGT  0
Sequence5:   TATTAGGACCACGTCRCGKT  0
Sequence6:   TATTGGGACCACGTCGCGGT  1
Sequence7:   TATTGGGACCMCSTCGCGGT  1
Sequence8:   TATTAGGACCACGTCGCGGT  0
Sequence9:   TATTAGGACCACGTCGCGGT  0
Sequence10:  TATTGGKRSMACGWNMCGGT  1
Sequence11:  TATTGGGACCC-GTSAYGCY  4
Sequence12:  TATTGGGACCACGYCGCSGY  2
Sequence13:  TATTGGGACCRCKTCGCGGT  1
Sequence14:  TAYTAGGACCACRTCGTGGT  1
Sequence15:  TATTAGGACCACGTCGAGGT  1
Sequence16:  TATTGGGACCACGTCGCGGT  1
Sequence17:  TATTAGGACCACGTCGTGGT  1
Sequence18:  TATTGGGACCACGTCGAGGT  0
Sequence19:  TATTGGGACCACGTCGCGGT  1
Sequence20:  TATTGGGACCACGTCGCGGT  1
Sequence21:  TATTGGGACCACGTCGCGGT  1
Sequence22:  TATTAGGACCACGTCGTGGT  1
Sequence23:  TATTAGGACCACGTCGTGGT  1
Sequence24:  TATTGGGACCACGTCGCGGT  1
Sequence25:  TATTAGGACCACGTCGCGGT  0
Sequence26:  TATTAGGACCACGTCGTGGT  1
Sequence27:  TATTAGGACCACGTCGTGGT  1
Sequence28:  TATTAGGACCACGTCGCGGT  0
Sequence29:  TATTAGGACCACGTCGCGGT  0
Sequence30:  TATTAGGACCACGTCGCGGT  0
Sequence31:  TATTAGGACCACGTCGCGGT  0
Sequence32:  TATTGGGACCACGTCGCGGT  1
Sequence33:  TATTAGGACCACGTCGCGGT  0
Sequence34:  TATTAGGACCACGTCGCGGT  0
Sequence35:  TATTAGGACCACGTCGAGGT  1
Sequence36:  TATTGGGACCACGTCGCGGT  1
Sequence37:  TATTGGGACCACGTCGAGGT  0
Sequence38:  TATTGGGACCACGTCGAGGT  0
Sequence39:  TATTAGGACCACGTCGAGGT  1
Sequence40:  TATTAGGACCACGTCGAGGT  1
Sequence41:  TATTGGGACCACGTCGCGGT  1
Sequence42:  TATTRGGACCACGTCGCGGT  0
Sequence43:  TATTAGGACCACGTCGCGGT  0
Sequence44:  TATTAGGACCACGTCGAGGT  1
Sequence45:  TATTAGGACCACGTCGAGGT  1
Sequence46:  TATTGGGACCACGTCGCGGT  1
Sequence47:  TATTAGGACCACGTCGCGGT  0
Sequence48:  TATTGGGACCACGTCGCGGT  1
Sequence49:  TATTGGGACCACGTCGCGGT  1
Sequence50:  TGTTAGGACCACGTCGTGGT  0
Sequence51:  TATTAGGACCACGTCGAGGT  1
Sequence52:  TATTAGGACCACGTCGTGGT  1
Sequence53:  TATTAGGACCACGTCGCAGT  1
Sequence54:  TATTGGGACCACGTCGAGGT  0
Sequence55:  TGTTAGGACCACGTCGTAGT  1
Sequence56:  TATTAGGACCACGTCGAGGT  1
Sequence57:  TATTAGGACCACGTCGTGGT  1
Sequence58:  TATTAGGACCACGTCGCGGT  0
```

LIÈGE université
Gembloux
Agro-Bio Tech

Genomes by tens - EURAVELCH

EU & non EU isolates ? LChV1

- Small clusters of EU or non-EU
- Overall: mix of EU and non-EU
- Ongoing analysis



Genomes of **<u>NEW</u>** viruses

Viruses genomes

CONSENSUS STATEMENT

OPEN

CONSENSUS STATEMENT

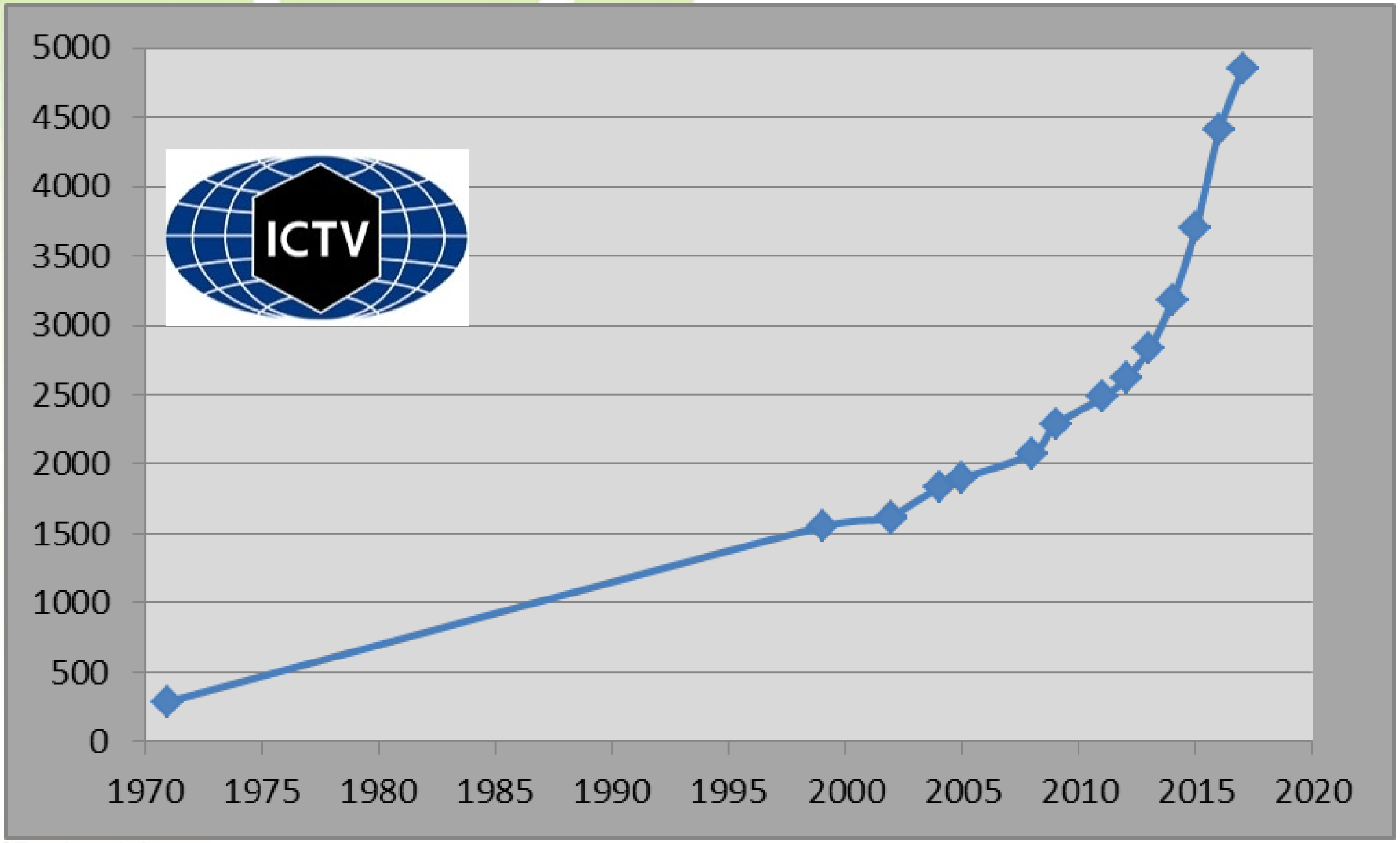
Virus taxonomy in the age of metagenomics

Peter Simmonds[1], Mike J. Adams[2], Mária Benkő[3], Mya Breitbart[4], J. Rodney Brister[5], Eric B. Carstens[6], Andrew J. Davison[7], Eric Delwart[8,9], Alexander E. Gorbalenya[10,11], Balázs Harrach[3], Roger Hull[12], Andrew M.Q. King[13], Eugene V. Koonin[5], Mart Krupovic[14], Jens H. Kuhn[15], Elliot J. Lefkowitz[16], Max L. Nibert[17], Richard Orton[7], Marilyn J. Roossinck[18], Sead Sabanadzovic[19], Matthew B. Sullivan[20], Curtis A. Suttle[21,22], Robert B. Tesh[23], René A. van der Vlugt[24], Arvind Varsani[25] and F. Murilo Zerbini[26]*

New viral sequences = new species

Pest genomes ? New pest candidates

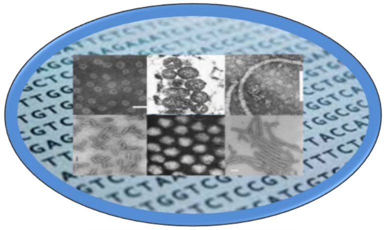

virus >>> bacteria fungi



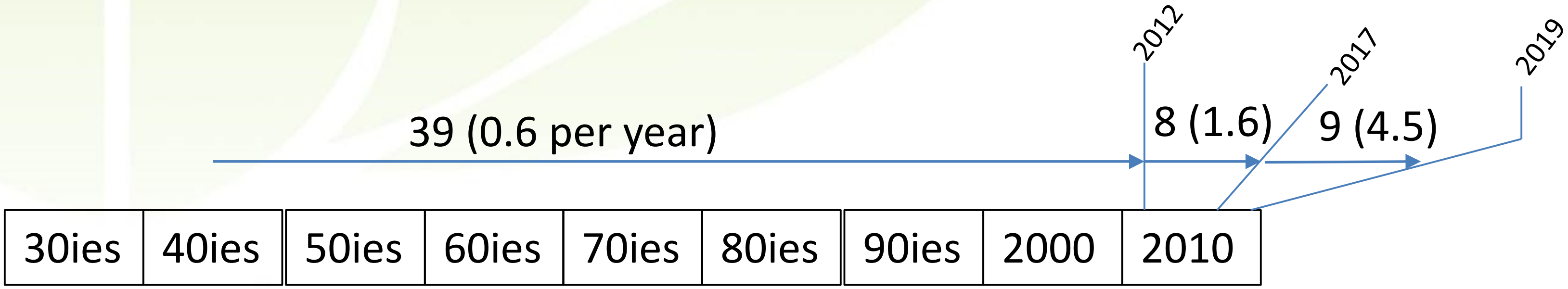
Plant Viruses ?

- ✓ 10 years in R&D
- ✓ Etiology

Hundreds of new viruses discovered



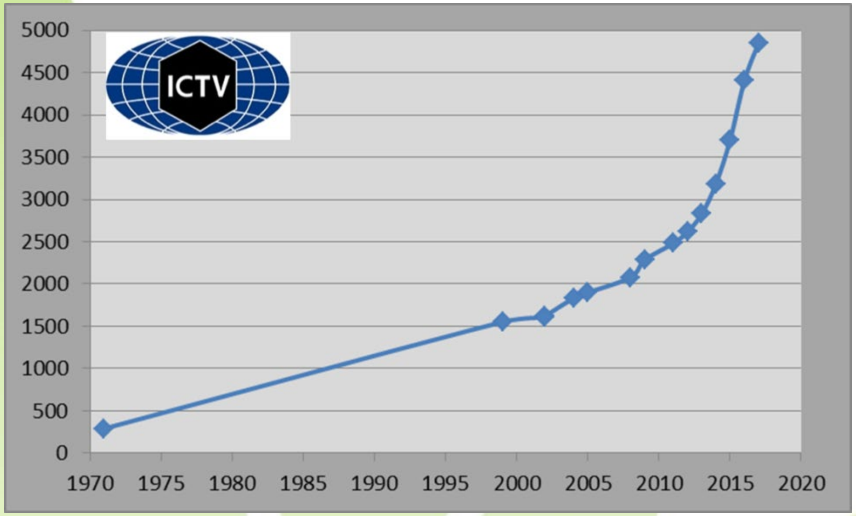
Viruses and viroids on *Prunus* sp.



30ies	40ies	50ies	60ies	70ies	80ies	90ies	2000	2010

LIÈGE université
Gembloux
Agro-Bio Tech

LABORATORY OF
Phytopathology



« ORPHAN VIRUSES »

Which one of the new plant viruses represents a risk for plant trade and production ?

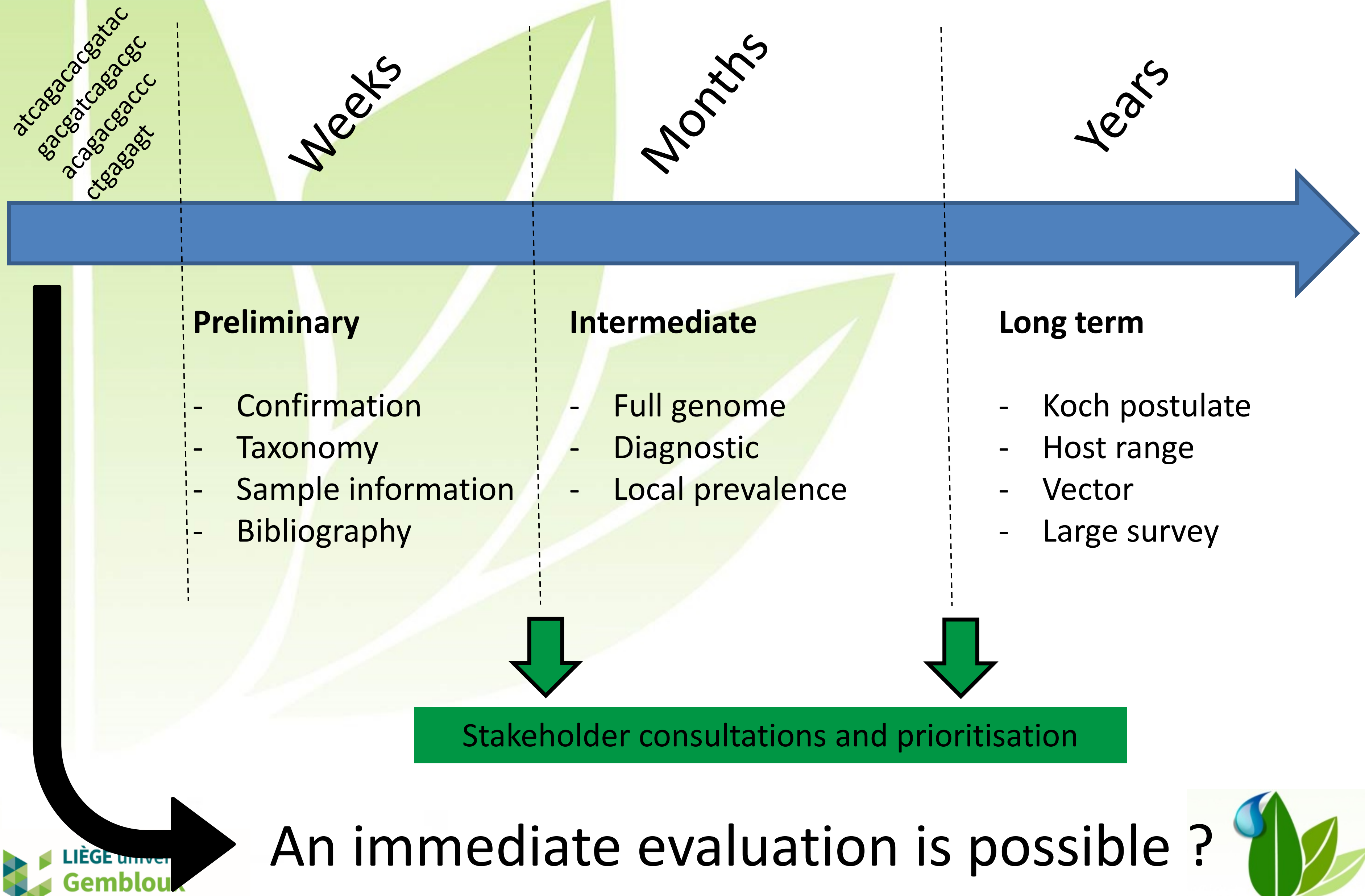
frontiers in Microbiology

PERSPECTIVE
published: 24 January 2017
doi: 10.3389/fmicb.2017.00045

A Framework for the Evaluation of Biosecurity, Commercial, Regulatory, and Scientific Impacts of Plant Viruses and Viroids Identified by NGS Technologies

Sebastien Massart[1], Thierry Candresse[2], José Gil[3], Christophe Lacomme[4], Lukas Predajna[5], Maja Ravnikar[6], Jean-Sébastien Reynard[7], Artemis Rumbou[8], Pasquale Saldarelli[9], Dijana Škorić[10], Eeva J. Vainio[11], Jari P. T. Valkonen[12], Hervé Vanderschuren[13], Christina Varveri[14] and Thierry Wetzel[15]*

Framework ?



atcagacacgatac
gacgatcagacgc
acagacgaccc
ctgagagt

Weeks

Months

Years

Preliminary

- Confirmation
- Taxonomy
- Sample information
- Bibliography

Intermediate

- Full genome
- Diagnostic
- Local prevalence

Long term

- Koch postulate
- Host range
- Vector
- Large survey

Stakeholder consultations and prioritisation

An immediate evaluation is possible ?

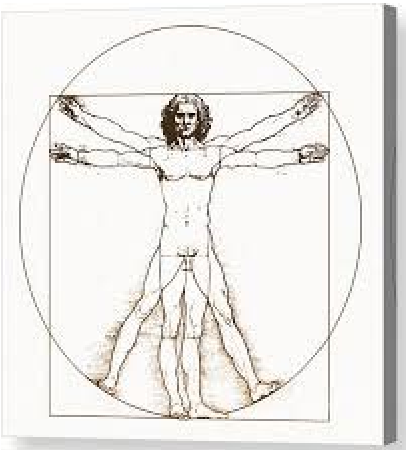
Genome sequence for PRA :

Nazca lines or Rosetta stone ?

EPIDEMIOLOGY

Predicting reservoir hosts and arthropod vectors from evolutionary signatures in RNA virus genomes

Simon A. Babayan[1,2], Richard J. Orton[3], Daniel G. Streicker[1,3]*



- Mamalian ssRNA viruses
- Bibliography:
  - ✓ Host for 437 viruses
  - ✓ Vector group for 527 viruses
  - ✓ Identity of vector for 97 viruses
- How predictable is the host, vector group and identity ?
- Machine learning algorithm (A.I.)

1. Predictive accuracy by Phylogeny:
   - 60% for reservoir host
   - 95% for vector group
   - 67% for vector identity

2. Predictive accuracy by genome mining:
   - 72 % for reservoir host (2nd OK)
   - 99 % for vector group

3. Predicting vector and host of « orphan » viruses

EPIDEMIOLOGY

Predicting reservoir hosts and arthropod vectors from evolutionary signatures in RNA virus genomes

Simon A. Babayan[1,2], Richard J. Orton[3], Daniel G. Streicker[1,3]*

- Taxonomy is already nice (90%)
- Genome-based analysis complete the information (95% = 30 additional species)
- Probability -> error possible
- Orientation of future research and risk evaluation

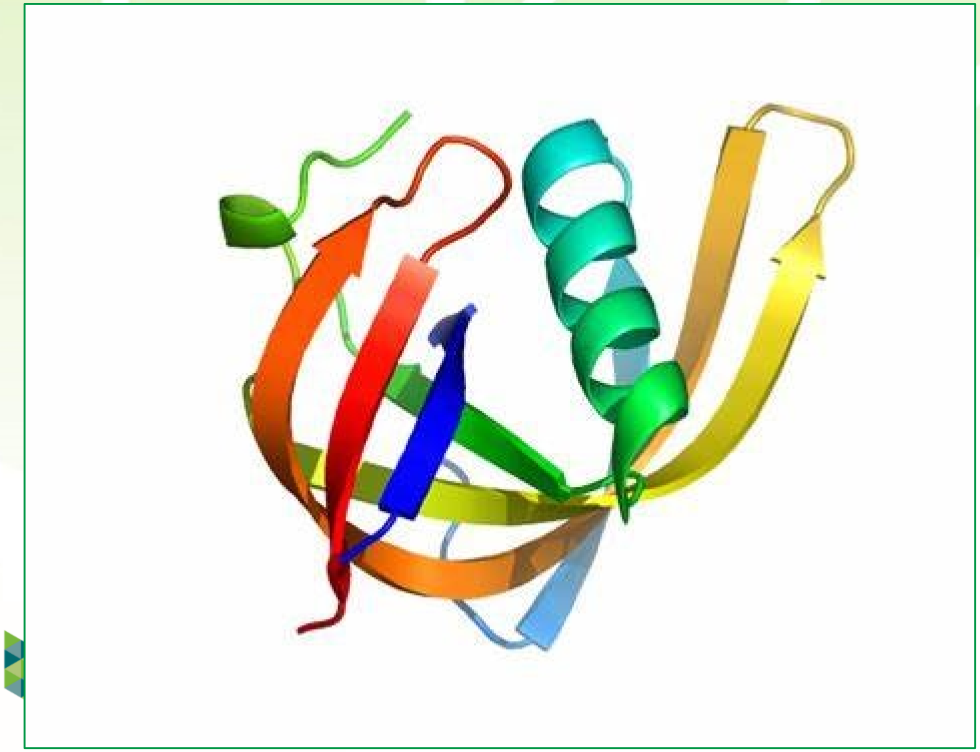
Machine Learning for viruses & plant health

TGAATGAGGATGAGGAAAAATGTCCAT
ACGTGCTATGCCGCTTTCCACTTCTCT
GAGAACCTGCTTCTTGATTTCGTAGAA

Genomes

ILKYVCKTYFPASNREVYMKEFLVTRVNT
WFCKFSRIDTFLLYKGVAHKSVDSEQFY

Proteins

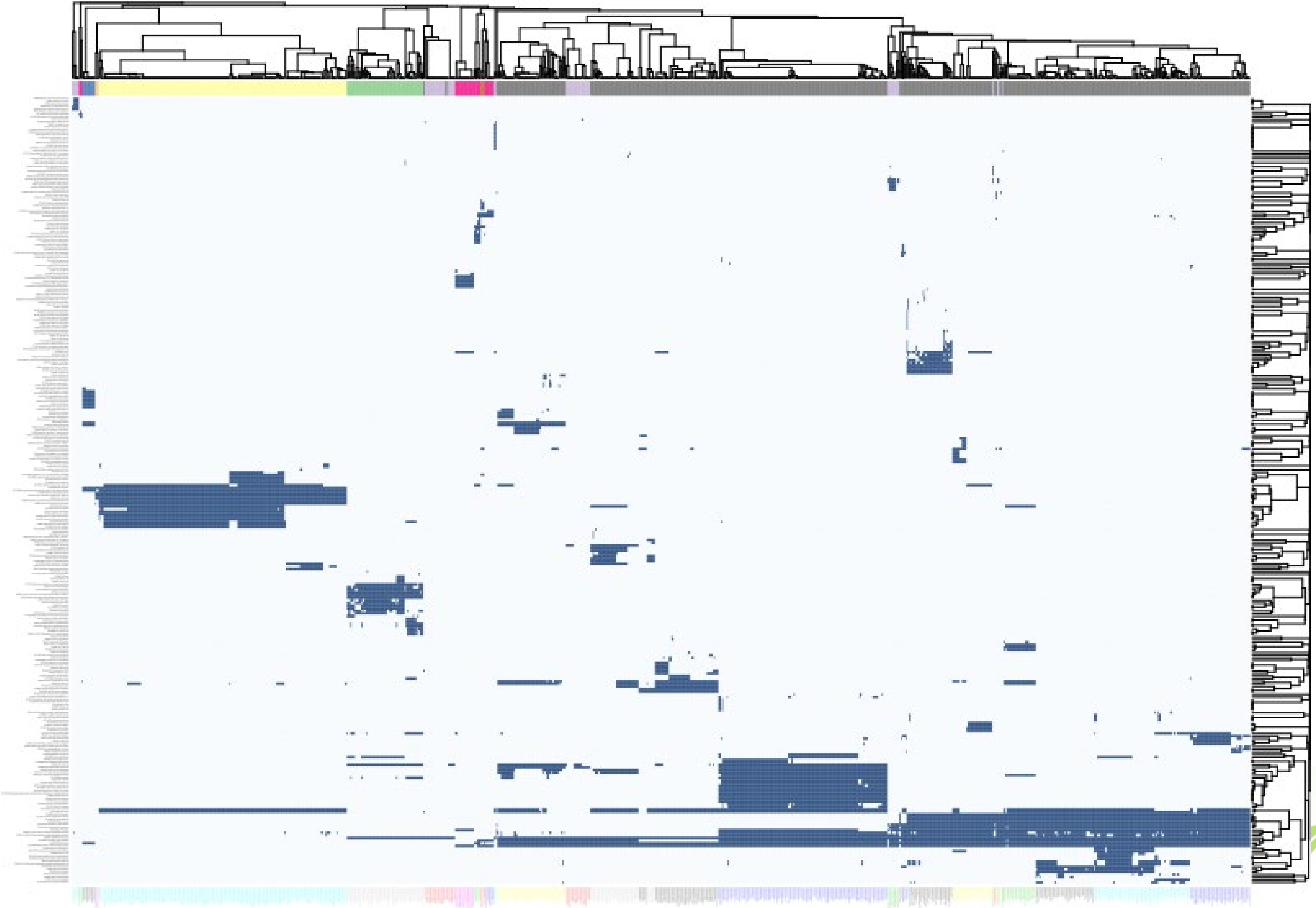


Protein information

1) Function
2) Signature (amino acids)
3) 3-D conserved structures

LABORATORY OF
Phytopathology

Modulome



Modulome & Plant health

- Module & transmission ?
- Taxonomy vs. Modules ?
- Machine learning algorithms

- Similar results but variation
- Sensitivity > specificity (PPV > NPV)

PhD finishing but ongoing story

How accurately can we predict
transmission, host range, symptoms…
from genome sequences ?

Conclusions on Genome era & Plant Health

- ✓ Genome era
- ✓ Bioinformatics
- ✓ Diagnostics and regulation
- ✓ Surveillance, outbreak origin & many others
- ✓ Genome-based P.R.A. ?

➔ **Genome information will improve plant health and promote efficient use of ressources**

LIÈGE université
Gembloux
Agro-Bio Tech

LABORATORY OF
Phytopathology

Many thanks to

SPF Health (NPPO Belgium)

Euphresco projects: EuravelCh, PHBN, Virfast…

EU COST Action FA 1407 – DIVAS

EU H2020 project VALITEST. www.valitest.eu

EU ITN MSC – INEXTVIR. www.inextvir.eu

FNRS (Belgium)

Tens virologists collaborating in the COST Action

Tens scientists from Euphresco projects

K. De Jonghe, R. Tahzima, A. Haegemans (ILVO)

R. Wei, G. Rufflard, A. Locicero (Uliege)

Thank you for your attention

sebastien.massart@uliege.be

@Be_Phytopath