



INEXTVIR: High throughput sequencing of the virome of agricultural crops and ecosystems across Europe

Kaoutar Daoud Hiri, Johan Rollin, Sukhorukov Grigorii, Macha Nikolski, Sébastien Massart, Tomaž Curk, Matjaž Hren

► To cite this version:

Kaoutar Daoud Hiri, Johan Rollin, Sukhorukov Grigorii, Macha Nikolski, Sébastien Massart, et al.. INEXTVIR: High throughput sequencing of the virome of agricultural crops and ecosystems across Europe. ESCS-ECCB 2020, Sep 2020, Barcelona, Spain. hal-03271520

HAL Id: hal-03271520

<https://hal.science/hal-03271520>

Submitted on 25 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Kaoutar Daoud Hiri^{a,b}, Johan Rollin^{c,d}, Sukhorukov Grigori^{e,f}, Nikolski Macha^{e,f}, Sébastien Massart^c, Tomaž Curk^b, Matjaž Hren^g

^aJožef Stefan International Postgraduate School, Ljubljana, Slovenia

^bUniversity of Ljubljana, Faculty Of Computer And Information Science, Ljubljana, Slovenia

^cPlant Pathology Laboratory, TERRA, Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium

^dDNAVision, Gosselies, Belgium

^eUniversity of Bordeaux, CNRS, IBGC, UMR 5095, Bordeaux, France

^fUniversity of Bordeaux, Bordeaux Bioinformatics Center, Bordeaux, France

^gBioSistemika, Ljubljana, Slovenia

Presented by
Grigori Sukhorukov
grigori.sukhorukov@u-bordeaux.fr



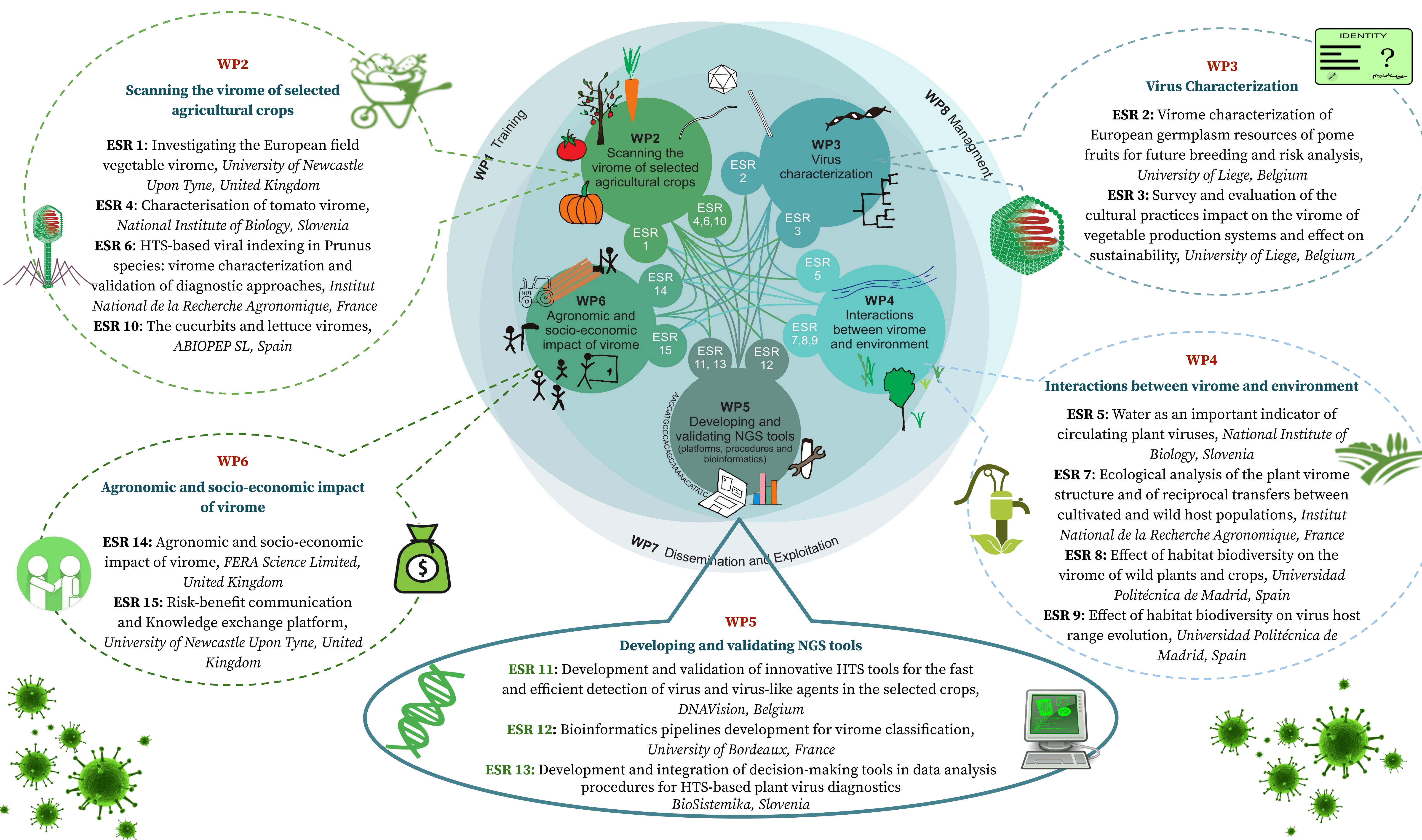
BACKGROUND

Plant Viruses cause 50% of the emerging plant diseases and pose an important threat to many agricultural crops worldwide, they are responsible for production losses estimated at € 15 to 45 billion per year.

INEXTVIR proposes a transdisciplinary research and training approach combining applied and fundamental plant virology, ecology, NGS platform development, computer science (bioinformatics), agronomical, socio-economic and regulatory sciences.

Goal: Better understand, control and exploit the virome of selected agricultural crops and ecosystems across Europe.

Benefits: Political, economic, industrial, social and environmental.



Contribution to bioinformatic and software development segments of the project

Developing of own bioinformatic methods

Finding alternative methods for plant virus detection

- 1) Viral identification in HTS data from agricultural crops and ecosystems
- 2) Generation of artificial HTS datasets for testing purposes
- 3) Assessment of cross-contamination in sequencing dataset
- 4) Obtain validation criteria on plant virus identification

Virus related tools benchmarking

Identify the most relevant HTS methods and tools

- 1) Identify relevant use-cases in the project consortium (e.g. virus detection, virus classification ...)
- 2) Identify the tools that seem relevant for the selected use-cases in our neutral study
- 3) Use Docker images of tools (either existing or created) for testing and results dissemination
- 4) Select datasets (artificial or natural) that meet test criteria
- 5) Share the benchmark result highlighting tools performance

Pipeline design

Pipeline testing

Identify the most relevant **combination** of methods and tools

- 1) Identify relevant steps in HTS data analysis (e.g. reads cleaning, identification ...)
- 2) Create pipelines based on Docker images of the tools and test them
- 3) Compare the pipelines performance
- 4) The most suitable pipelines for each use case will be shared with project consortium in Docker images
- 5) Document the pipeline for users and developers

Decision making

Propose the most appropriate pipeline depending on the use case for the user

- 1) Use performance comparison to make a recommendation
 - a) Store relative performance metrics for each pipeline
 - b) Access those metrics to be able to make a decision
 - c) Update the performance metrics based on new tests and tools
- 2) Explain the recommendation to the users so that they make a final decision

Future consideration: Automated pipeline recommendation using a Machine Learning approach

Partner organizations

INEXTVIR is implemented by a European Consortium of universities, research institutions and companies in Belgium, France, Spain, Slovenia and the UK.

