



Plant Health at the Age of Metagenomics

A scientific Colloquium organised by the
European and Mediterranean Plant Protection Organization
and the
Euphresco network for phytosanitary research coordination
and funding

UNESCO, Paris, France

26th of September 2019



Foreword

New techniques allowing processing of large numbers of samples and generating huge volumes of genomic and protein data offer new opportunities to study plant pests. Rather than isolating individual molecules, in order to understand their role in the biology of an organism, it is now possible to investigate the genome as a whole, or the interactions of proteins and other metabolites as a holistic approach. These new techniques allow known pests to be studied in more detail as well as to investigate the genomic diversity within species. The information produced will improve diagnostics and may help in the management of pests, for instance to develop resistant plant species. These new techniques can also detect new species that may have been present for many years unnoticed, as well as emerging pests. They may also help identify the causal agents of diseases which were previously unknown. Genomic data may show similarity between new species found and known pests, which may give indications on the relevance of these newly identified organisms as potential pests.

Data needs to be stored in a way that it is findable, accessible, interoperable and reusable (FAIR). Bioinformatics is an important tool to handle and analyse the ever-growing amount of data produced for example by comparing genomic sequences, predicting the activities of proteins or modelling metabolic pathways. What is the biological significance of the data? In a plant health context, one may ask which genes, proteins or other molecules cause organisms to be pathogenic? Which molecules increase virulence?

The EPPO/Euphresco Colloquium 'Plant Health at the age of metagenomics' will bring together scientists and regulators who are involved in plant health. By sharing experiences and information, together we can better understand the possibilities that metagenomics offers and how to use the vast amounts of information that will become available to improve protection of plant health. A crucial question is what makes a new species that is detected a threat, and warrants phytosanitary action? What in the genome, the proteins and other molecules of a pest makes it potentially invasive in a new region or a risk to a new host? If we shift the focus from pests only to the interaction between plants and pests one can ask what in a plant makes it a host, or more (or less) susceptible to damage by the pest?

This colloquium is a very good opportunity to discuss the challenges associated with the study of the genome of pests, populations, and plants, and the opportunities offered to obtain information about which plants are hosts, the extent of damage the pest could cause in a host and the way in which we could mitigate the effect of the pest. In the same way that I see the benefits of the rapid developments in sequencing and bioinformatic analysis capacity, I am looking forward to the fast development of data interpretation, in particular concerning information relevant to the interaction between pests and plants and thereby to the management of pests and diseases.

Nico Horn

Colloquium agenda

Chair: Marcel Westenberg, National Plant Protection Organization, NL

- 09:00-09:10** **Introduction** (Nico Horn, European and Mediterranean Plant Protection Organization and Marcel Westenberg, National Plant Protection Organization)
- 09:10-09:30** **Genome-based pest risk assessment for newly discovered plant viruses: Nazca lines or Rosetta stone?** (Sébastien Massart, University of Liège, BE)
- 09:30-10:00** **Using collections to support pest risk assessment of novel and unusual pests discovered by high throughput sequencing** (Adrian Fox, Fera Science Ltd., GB)
- 10:00-10:30** **Practical use of metagenomics: the case study of the *Ralstonia solanacearum* species complex** (Gilles Cellier, French Agency for Food, Environmental and Occupational Health & Safety, FR)
- 10:30-11:00** **Coffee break**
- 11:00-11:30** **Genomics as a tool for the diagnosis of bacterial diseases: the case of *Xanthomonas arboricola* pv. *pruni*, the causal agent of bacterial spot of stone fruits** (Jaime Cubero, National Institute for Agricultural Research and Food Technology, ES)
- 11:30-12:00** **Genome-enhanced detection and identification of plant pathogens** (Guillaume Bilodeau, Canadian Food Inspection Agency, CA)
- 12:00-12:30** **Discussions**
- 12:30-13:30** **Lunch break (lunch not provided)**
- 13:30-14:00** **The wheat-rust conflict: shifty enemies and the long reach of genomics** (Diane Saunders, John Innes Centre, GB)
- 14:00-14:30** **Using a FAIR database and bioinformatics analyses to improve plant, human, animal and ecosystem health** (Kim Hammond-Kosack, Rothamsted Research, GB)
- 14:30-15:15** **Discussions and conclusions**
- 15:15** **End of the meeting**

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

Sébastien Massart, University of Liège, Belgium

sebastien.massart@uliege.be

High-Throughput Sequencing (HTS) technologies have allowed the discovery of more than one hundred new plant viruses in recent years and the pace of discovery is expected to be maintained in the near future. HTS and genome data are accumulating and plant virology is therefore progressively entering the “big data” era with its complexity and its opportunities. For example, HTS has impacts on the taxonomy of viruses as the International Committee on Taxonomy of Viruses (ICTV) is now accepting new viral species based on HTS data only (Simmonds *et al.*, 2017).

Plant virologists and risk managers are now facing a new challenge: they need to do the work backwards, namely to characterize the biology (and phytosanitary risks) of new viruses, with genome sequence information as the unique starting point.

A framework of biological characterisation has been proposed recently (Massart *et al.*, 2017). It relies on a succession of biological experiments to progressively and efficiently characterize the risk posed by a new virus. The accumulation of genome data provides new opportunities and challenges: is it possible to determine some biological properties of a new virus based on its genome sequence and if so what is the level of certainty?

Recent approaches carried out for mammalian viruses will be presented. In addition, the presentation will discuss the situation for plant virology and describe ongoing work based on big data management and machine learning algorithms to exploit genome information for *in silico* biological characterization of plant viruses.

Using collections to support pest risk assessment of novel and unusual pests discovered by High Throughput Sequencing

Adrian Fox, Fera Science Ltd., United Kingdom

Adrian.Fox@fera.co.uk

Since the first reported use of High Throughput Sequencing (HTS) a decade ago, the technology has moved from being a research support tool into the virus diagnostic laboratory. The technology can be used: to investigate plants expressing symptoms of unknown aetiology; to screen plant material in support of phytosanitary certification; and for crop surveillance. The use of HTS in academic and official activities has led to an ever-increasing catalogue of previously uncharacterised viruses being detected. Alongside this, there have been some limited studies looking at historic samples for the presence of pathogens. Some reports have arisen from incidental findings, such as using HTS to study ancient nucleic acids in archaeological samples such as the discovery of a novel chrysovirus from a 1000-year-old maize cob. Studies have helped to shed light on the emergence and spread of key pathogens such as late blight of potato (*Phytophthora infestans*). However there have been few, if any, attempts to exploit historic collections (such as herbaria) as a resource for better understanding pathogens of plant health importance.

In the framework of Euphresco it was proposed to start an international collaboration (entitled VirusCurate) to expand work being conducted to systematically sequence viruses from collections. Currently, this has largely focussed on isolates of viruses collected from tuber crops of Andean origin, due to the recent outbreak of previously unknown viruses in the novel crop *Ullucus tuberosus*. A combination of ELISA, PCR and HTS was used to detect and identify a suite of non-native plant viruses including both characterised and novel viruses. In many cases the 'novel' viruses detected appeared to be similar to viruses which had previously been biologically characterised, and some of these appeared to be similar to quarantine listed viruses. There are no reference sequences to definitively link the new findings and old viruses. This case highlighted the apparent similarity between 'novel' species detected through HTS, and previous reports which predate sequencing technologies. Unfortunately, this case is not isolated and there are multiple examples of viruses described in the literature for which there are no reference sequence data. Linking these data would allow risk assessments on novel findings to be informed through historic literature. Previous detections of viruses from historic samples and the scope of the Euphresco topic 'VirusCurate' will be discussed.

Practical use of metagenomics: the case study of the *Ralstonia solanacearum* species complex

Gilles Cellier, French Agency for Food, Environmental and Occupational Health & Safety, France

gilles.cellier@anses.fr

How do some pathogenic bacteria adapt to new host plants? How can we quickly detect and respond to new pathogen variants? The advent of genomics, defined as the analysis of the structure and function of genomes, has led to important advances in many fields of phytobacteria research. Comparative genomics has made it possible to compare the different sequenced genomes to better understand their biological functions through their commonalities and differences, and to study evolutionary history of pests.

The *Ralstonia solanacearum* species complex provides a unique model for studying plant-pathogen interactions, including basic biology of pathogenesis and non-host resistance in the context of an unusually broad host range and latent (asymptomatic) infection. This species complex has a high genetic diversity associated with a high phenotypic diversity that results in a host range comprising about 250 plant species, distributed in 50 botanical families (Hayward, 1991). Extensive knowledge on species and subspecies assignment is not only of interest in the world of taxonomy but is particularly important when one considers key strategic areas such as plant breeding in agronomy, development of diagnostic methods, and epidemiology.

The use of genomic data, mainly through comparative genomics, contributed to: i) taxonomically reviewing species, mainly through High Throughput Sequencing and mass spectrometry ii) unravelling the genetic background of well-known characterized phylogenetic lineages and identify lineage-specific features potentially involved in host range variation iii) redefining plant-host relationship for specific lineages through genome sequencing and RNA sequencing (expression profiling) iv) producing massive amounts of data on gene content to study epidemiology and for diagnostic applications. These four main thematic areas will be presented through practical cases to demonstrate the potential of metagenomic data to derive information on pest biology, and to provide information for pest risk analyses.

Genomics as a tool for the diagnosis of bacterial diseases: the case of *Xanthomonas arboricola* pv. *pruni*, the causal agent of bacterial spot of stone fruits

Jaime Cubero, National Institute for Agricultural Research and Food Technology, Spain

cubero@inia.es

The omics era has provided the opportunity to study in-depth the mechanisms that underline infection processes and those involved in host colonization by plant pathogens. Only by understanding these processes will it be possible to develop effective pest management strategies. In addition, detection and identification are essential to any phytosanitary measures. Advances in genomics can contribute to a better understanding of the biology of plant pathogenic bacteria and genomic tools have become crucial for accurate classification of these pests. The presentation will provide examples of how genotypic analysis can be used to develop diagnostic tests based on bacterial virulence factors.

Xanthomonas arboricola pv. *pruni* (Xap), the causal agent of bacterial spot of stone fruits and almond, is a quarantine pest already present in several EPPO countries. As in other pathosystems, Xap coexists in *Prunus* spp. with non-pathogenic *Xanthomonas* that show distinct genotypic and phenotypic characters. Genomic analysis confirmed differences in the genes involved in strain virulence and population establishment in plants between these two intraspecific groups. Moreover, comparison with other *Xanthomonas* isolated from non-*Prunus* spp. allowed identification of virulence genes associated with infection in specific hosts. Taking advantage of gene variation, PCR tests can be developed to differentiate pathogenic and non-pathogenic strains or bacterial strains that are pathogenic to different plant species. A real-time PCR test for Xap was developed that showed high analytical specificity and high analytical sensitivity. The test is a valuable molecular tool for the diagnosis of bacterial spot of stone fruit and represents a good example of how genomics can support plant pathology. This work was supported financially by Instituto Nacional de Investigación Agraria y Alimentaria (INIA) projects RTA2014-0018 and RTA 2011-0040 (cofounded by FEDER).

Genome-enhanced detection and identification of plant pathogens

Guillaume Bilodeau, Canadian Food Inspection Agency, Canada

guillaume.bilodeau@canada.ca

High Throughput Sequencing (HTS) technologies are capable of processing large numbers of samples and producing even larger volumes of genomic data. By mining the genomes of closely-related pests and comparing them to one another, it is possible to design molecular markers that can be used to distinguish pests. These technologies can also be used to obtain genetic information from environmental samples thereby allowing the trajectory of plant diseases to be traced. Genome-Enhanced Detection and Identification (GEDI) techniques involve using DNA-based tools such as real-time PCR, SNP markers, and AmpliSeq technologies for the purpose of detecting and assessing the genetic diversity of pathogens in plant samples from the species to intra-lineage levels. A proof-of-concept integrated method and bioinformatic pipeline was used to design some *Fusarium* and *Phytophthora* spp. and *Lachnellula willkommii* - specific markers to demonstrate the utility and applicability of HTS for the detection of targeted species. Metagenomics tools and a custom bioinformatics pipeline can also be used to focus pest monitoring efforts. We have demonstrated that it is possible to use these genomic tools associated with bioinformatic pipelines to quickly develop molecular markers that can be used to facilitate detection and identification of fungal pathogens and improve regulatory activities.

The wheat-rust conflict and the long reach of genomics

Diane Saunders, John Innes Centre, United Kingdom

Diane.Saunders@jic.ac.uk

Wheat rusts have been associated with crop failures and famine throughout history. Recent outbreaks of stem and yellow (stripe) rust in Europe have been linked to expansions in pathogen geographic distribution, exotic incursions and increased virulence. To achieve effective disease management, timely and accurate diagnosis are required to guide control measures. Furthermore, the capacity to distinguish between individuals in a pathogen population with specific properties such as fungicide resistance, toxin production and virulence profiles is often essential for the choice of control methods. The genomics revolution has provided technologies that can rapidly generate genome-scale genetic information to define individual variants of a pathogen species. However, their application to fungal plant pathogens has, to date, remained limited due to their frequent obligate biotrophy and large genome sizes. We developed a portable, genomics-based, point-of-care approach that utilizes the nanopore sequencer and is specifically tailored for identifying individual strains of the wheat yellow rust pathogen. Using targeted sequencing, we overcame the limitations associated with applying such strategies to fungal pathogens. This new method called Mobile And Real-time PLant disEase (MARPLE) diagnostics enables rapid identification of both newly emergent strains and those with specific properties such as fungicide resistance. Generating results within 48 h of field sampling, this new strategy has the potential to revolutionise plant disease diagnostics, changing how plant health threats are identified and tracked in the future.

Using a FAIR database and bioinformatics analyses to improve plant, human, animal and ecosystem health

Kim Hammond-Kosack, Rothamsted Research, United Kingdom

kim.hammond-kosack@rothamsted.ac.uk

The amount of genomic data is doubling every seven months. In 2020, data will be generated a million times faster than it was less than 10 years earlier. As the biosciences become an increasingly data-intensive discipline and mega-scale data analyses become the new norm, building and maintaining community resources that ensure the Findability, Accessibility, Interoperability, and Reusability of data (i.e. are FAIR) will benefit many different bioscience disciplines.

In 2005, the pathogen-host interactions database PHI-base (www.phi-base.org) was publicly launched. PHI-base is a gold-standard database storing phenotypes on genes implicated in virulence and also the 1st host targets of pathogen effectors (Urban *et al.*, 2017). It is a primary information source for researchers studying plant-pathogen interactions as well as human pathogens of fungal and bacterial origin. Manually curated, information from more than 3 000 research articles are made accessible for researchers to easily familiarise themselves with relevant molecular and biological facts on pathogenicity, virulence and effector genes, on 1st host target and *in vitro* / *in host* phenotypes. High-level phenotypes are used to describe the overall experimental interaction outcomes enabling comparative analysis of different pathosystems. Since 2011, the phenotypic data in PHI-base has been directly connected to the individual gene entries within the genomes of more than 300 plant pathogenic species available within ENSEMBL invertebrates (Pedro *et al.*, 2016; Kersey *et al.*, 2019). PHI-base adheres to the FAIR principles and in 2016 joined the UK node of the ELIXIR 'Data for Life' project supplying agri-genomics data.

The many different ways in which the curated phenotypic and genotypic information on pathogens can be used in conjunction with metagenomics data to increase the repertoire of approaches deployed to enhance plant health will be described and discussed. These include tracking pathogen genes which, when mutated, lead to an enhanced disease-causing abilities (Brown *et al.*, 2016), targeting the removal of host susceptibility loci in plant breeding programmes, undertaking effectoromics directed screening of plant germplasm to maintain the widest possible repertoire of disease resistance genes, using new information about 'essential for life' genes to nominate novel intervention targets to be explored by bioscientists and the agrochemistry/agrobiotechnology industries; undertake pan-genome analyses between historic and current strains to explore recent evolutionary changes within and across regions caused as a result of previous and existing control practices, new types of interspecies comparative analyses that take advantage of taxonomic and lifestyle similarities or differences and finally graphical visualisation analyses of protein-protein network analyses to predict potential Achilles' heels for particularly problematic species. These types of analyses can be undertaken at multiple scales to permit disease risk managers, plant breeders and bioscientists to take full advantage of the deluge of metagenomic and other types of 'omics data'. This research and the PHI-base resources are supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC) within the grants BB/I/001077/1, BB/K020056/1 and as a National Capability BB/J/004383/.