

NEXT GENERATION SEQUENCING AND PROTEOMICS IN PLANT VIROLOGY: HOW IS COLOMBIA DOING?

Plataformas de secuenciación de nueva generación y proteómica aplicadas a la virología vegetal: ¿Cómo ha avanzado Colombia?

Leidy Johana MADROÑERO^{1*}, Zayda-Lorena CORREDOR-ROZO¹, Javier ESCOBAR-PÉREZ¹, Myriam-Lucía VELANDIA-ROMERO¹

¹ Laboratorio de Genética Molecular Bacteriana, Vicerrectoría de Investigaciones, Universidad El Bosque, Av. Cra 9 n°. 131 A – 02, Edificio Fundadores, Bogotá, Colombia

² Grupo de Virología, Vicerrectoría de Investigaciones, Universidad El Bosque, Av. Cra 9 n°. 131 A – 02, Edificio Fundadores, Bogotá, Colombia

*For correspondence: lmadronero@unbosque.edu.co

Received: 04th May 2019, Returned for revision: 28st June 2019, Accepted: 04th July 2019.

Associate Editor: María Cristina Navas.

Citation/Citar este artículo como: Madroñero LJ, Corredor-Rozo ZL, Escobar-Pérez J, Velandia-Romero ML. Next generation sequencing and proteomics in plant virology: how is Colombia doing? Acta biol. Colomb. 2019;24(3):423-438. DOI: <http://dx.doi.org/10.15446/abc.v24n3.79486>

ABSTRACT

Crop production and trade are two of the most economically important activities in Colombia, and viral diseases cause a high negative impact to agricultural sector. Therefore, the detection, diagnosis, control, and management of viral diseases are crucial. Currently, Next-Generation Sequencing (NGS) and ‘Omic’ technologies constitute a right-hand tool for the discovery of novel viruses and for studying virus-plant interactions. This knowledge allows the development of new viral diagnostic methods and the discovery of key components of infectious processes, which could be used to generate plants resistant to viral infections. Globally, crop sciences are advancing in this direction. In this review, advancements in ‘omic’ technologies and their different applications in plant virology in Colombia are discussed. In addition, bioinformatics pipelines and resources for omics data analyses are presented. Due to their decreasing prices, NGS technologies are becoming an affordable and promising means to explore many phytopathologies affecting a wide variety of Colombian crops so as to improve their trade potential.

Keywords: NGS, plant-virus interactions, plant virology, proteomics, viral genomics.

RESUMEN

La producción y el comercio de cultivos es una de las actividades económicas más importantes para el país. Las enfermedades causadas por virus ocasionan graves pérdidas económicas en el sector, por lo tanto, la detección, diagnóstico y diseño de estrategias para su control y manejo es crucial. Las tecnologías de secuenciación masiva (NGS por sus siglas en inglés) y las ciencias Ómicas constituyen hoy, una herramienta para el descubrimiento de nuevos virus y para el estudio de la interacción entre los virus y su hospedero vegetal. Este conocimiento no solo permite el desarrollo de nuevos métodos de diagnóstico, sino también permite el descubrimiento de componentes claves en la infección, los cuales podrían usarse para obtener plantas resistentes a los virus. En el mundo, el manejo de cultivos se está trabajando con ese enfoque. Por lo tanto, en esta revisión se presentan las diferentes aplicaciones de las tecnologías ómicas en la virología de plantas y el avance que ha alcanzado Colombia. Adicionalmente, se muestran los diferentes recursos y programas usados para el análisis bioinformático de datos ómicos. Debido a su costo cada vez más reducido, las tecnologías NGS son una excelente oportunidad para explorar fitopatologías en una gran diversidad de productos agrícolas y para mejorar su potencial comercial.

Palabras clave: Genómica de virus, interacción planta-virus, NGS, proteómica, virología vegetal.

INTRODUCTION

Next-generation sequencing (NGS) technologies include DNA sequencing and its derivatives, which are based on in-depth, high-throughput, and in-parallel methods. In this review, second-generation sequencing (SGS) and third-generation sequencing (TGS) techniques are treated as NGS. Relative to Sanger-first generation sequencing, NGS technologies are highly efficient, rapid, and cheaper and allow the sequencing of hundreds of gigabases in a few hours (Barba *et al.*, 2014; Kulski, 2016). As a reference, the sequencing of the human genome, accomplished in 2003, took 13 years and cost about USD 2.7 billion (International Human Genome Sequencing, 2004). Today, it is possible to sequence genomes with sizes similar to the human genome in less than a day for less than USD1000 (NIH; Hernandez, 2018) TGS applications have also increased recently; however, these promising technologies are still in the initial phase of the application.

Having been available for approximately 15 years, NGS has become a widely used tool in all life sciences research. Substantial improvements have been seen in costs, sequence accuracy, and yield (Goodwin *et al.*, 2016). Specifically, the increased use of high-throughput sequencing technologies has resulted in an exponential increase in the number of sequenced genomes, metagenomes, and transcriptomes from a wide variety of species. NGS platforms produce an enormous amount of data, the analysis, and interpretation of which has provided new insights and knowledge of molecular bases of human and plant diseases and in plant resistance to both biotic and abiotic stresses. Further, NGS has profoundly impacted the development of medicine, agriculture, and industry. NGS is part of the large scale technologies applied to the study of a whole collection of biomolecules present in a given cell, tissue, organ or organism and these are termed as “omics” and include the study of DNA (genomics), RNA (transcriptomics), proteins (proteomics), lipids (lipidomics), carbohydrates (glycomics), and other metabolites (metabolomics) as well as integrated systems (Kulski, 2016; Stagljar, 2016).

The application of NGS and ‘omic’ technologies in plant virology came into common use around 2009 (Hadidi *et al.*, 2016) and included several approaches. Specifically, in the following text, nucleic acid sequencing and proteomics will be discussed in detail. Although NGS has been widely applied worldwide, in Colombia, its application is still at an early stage. Despite significant advances made in the application of genomics to plant virus identification and characterization, other omics have been little explored. Thus, in addition to introducing the current status and advances in NGS technologies and their main applications, this review will show that the reduced costs and extensive use of these technologies throughout the scientific community have increased their accessibility to everyone, even to

developing countries, such as Colombia. Therefore, we wish to encourage Colombian researchers to implement and increase the use of high-throughput technologies.

NGS PLATFORMS

NGS technologies include several platforms that enable the sequencing of millions to billions of DNA or cDNA fragments. The lengths of the sequenced DNA fragments and sequencing methods vary based on the platform used. At present, there are seven major sequencing platforms (Table 1). According to the sequencing method, these technologies can be grouped as follows: The first group is composed of DNA sequencing technologies that previously required a PCR step for cDNA or DNA amplification. Currently, this group is constituted by both first and second-generation sequencing technologies, such as the GS FLX 454 sequencer (Roche Diagnostics Corp., Branford, CT, USA), which was discontinued, the Illumina platforms (Illumina Inc., San Diego, CA, USA), which are the most widely used, the ABI SOLiD System (Life Technologies Corp., Carlsbad, CA, USA), and the Ion Personal Genome Machine (Life Technologies, South San Francisco, CA, USA). The second group is composed of those technologies that are based on direct single-molecule sequencing (without a preceding PCR amplification step). These technologies are considered third-generation platforms by most authors (Heather and Chain, 2016) and include the HeliScope (HelicosBioScience Corp., Cambridge, MA, USA), which filed for bankruptcy in 2012, the PacBio single-molecule real-time (SMRT) system (Pacific Biosciences, Menlo Park, CA, USA), which is the most widely used third-generation technology, and the GridION, MinION and PromethION Nanopore platforms (ONT) (Oxford Technologies), which are the most promising sequencing technology. The last platform was tested by end-users in a trial in 2014 (Loman and Quinlan, 2014) and is hoped to revolutionize DNA sequencing as it allows the use of small portable devices, production of very long reads in concise times, and low costs; however, this technology has poor-quality profiles (Heather and Chain, 2016; Rang *et al.*, 2018).

Second and third-generation sequencing applied to omics studies includes the complete sequencing of genomes, transcriptomes, metagenome, amplicon sequencing, and other specific categories, such as sequencing for large-scale polymorphism discovery, bisulfite-treated DNA, methylation in genomic DNA, chromatin immunoprecipitation sequencing (ChIP-Seq) to determine protein-DNA interactions, and others. As follows, we focus and describe in detail the application of complete sequencing of genomes, transcriptomes, and metagenomes in plant virology.

Table 1. Summary of Next-generation DNA sequencing technologies platforms and its main characteristics.

| Generation | Sequencing platform | Major Instruments | Amplification method | Sequencing chemistry | Average Read length (bp) | Read per Run | Sequencing Speed/h | Accuracy (%) |
|------------|---------------------------------|--|--|---|--------------------------|-----------------|--------------------|--------------|
| Second | 454 (Roche)* | GS20, 454 GS FLX, GS FLX Titanium+ | | Pyrosequencing | 100-700 | 200-1M | 13 Mbp | 99.9 |
| | SOLiD (Life Technologies) | 5500 W, 5500xl W | Emulsion PCR | Ligation | 75-85 | 3B-6B | 21-28 Mbp | 99.9 |
| | Ion Torrent (Life Technologies) | PGM 314 chip v2, Ion Proton, Ion S5/S5XL 540 | | Detection of released H | 200-400 | 400,000 -80M | 25 Mb - 16 Gbp | 99 |
| | Illumina (Illumina) | MiniSeq, MiSeq, NextSeq, HiSeq, NovaSeq | Bridge PCR | Reversible terminators | 100-300 | 25M-20B | 25 Mbp | 99.9 |
| Third | PacBio (Pacific Biosciences) | RS C1, RS C2, RS II C2 XL, Sequel, RS II P6 C4 | No amplification Single molecule real-time (or SMRT) | Fluorescently labeled nucleotides | 1,300 - 15,000 | 10,000- 140,000 | 50 - 115 Mbp | 95 |
| | Helicos (Helicos Biosciences) | Helicoscope | | Reversible terminators | 25-55 | 12M-20M | 83 Mbp | 97 |
| | Nanopore (Oxford Technologies) | Flongle, MinION, GridION X5 PromethION P24/P48 | No amplification Single molecule | Nanopore, changes in current signals by nucleotides | 2,000- 20,000 | 10,000- 150,000 | 150 Mbp | 96 |

*Discontinued; M: Million; B: Billion

NGS APPLIED TO PLANT VIRAL GENOMICS

Plant viruses are causal agents of several plant diseases; therefore, plant virus is responsible for significant losses in crop production and trade (Nicaise, 2014; Hadidi *et al.*, 2016). As a consequence, its detection and diagnostic is a crucial step to determine a specific disease etiology involving a virus symptomatology and for its respective crop management program. For virus detection and characterization, standard methods have been used such as electron microscopy, serological, and molecular tests (Pecman *et al.*, 2017). However, due to the high variability found among plant viruses, these methods do not always work (Hadidi *et al.*, 2016). The application of High-Throughput Sequencing Technologies to obtain viral genomes has had aid highly to overcome this challenge. Since 2009, when the first viral genomes were obtained using NGS, viral genomics has had an enormous impact on the identification and discovery of novel viruses and viroids and also in refining the characterization and diagnosis methods of previously identified viruses (Barba *et al.*, 2014; Hadidi *et al.*, 2016; Blawid *et al.*, 2017; Pecman *et al.*, 2017).

Conventionally, genomics investigates a complete set of DNA of a given organism, including but not limited to its gene and intragenic sequences, functions, annotations, and structure (Blawid *et al.*, 2017; Pecman *et al.*, 2017). Genomics has progressed in parallel to the development of NGS platforms, and it has been widely applied to all life sciences. However, in viral genomics, this concept should

be redirected to fit specific viral characteristics, including its variable genetic material composition, host-dependent nature, and small genome size (ranging from 2.6 kb to 19.3 kb) (Hull, 2014).

Since nucleic acid from viruses and plant are often in a mixture in which viral sequences are a tiny proportion, enrichment of viral sequences is a crucial step to be considered before sequencing. The protocols will depend on plant virus genetic material, which may be composed by a single-stranded RNA (ssRNA), double-stranded RNA (dsRNA), single-stranded DNA (ssDNA), or double-stranded DNA (dsDNA) (Hull, 2014). The main methods used for plant viral genome sequencing is viral metagenomics or viral meta-transcriptomics using modifications to virus sequence enrichment or using ultra-deep sequencing directly or from purified viral particles.

Viral metagenomics or viromics refers to those methods employed for the detection of all viruses present in a given sample using sequencing technologies (Roossinck *et al.*, 2015). Metagenomics is the sequencing of environmental DNA, comprised of tens to thousands of organisms. In metagenomics analysis is common the use of conserved sequences, usually ribosomal RNA sequences regions; however, in viruses, there are not universally conserved genes or regions. Therefore, in viral metagenomics, the total acid nucleic (DNA or RNA) from virus-infected plants are obtained. Nevertheless, the optimal condition is to minimize (as much as possible) contamination by host genetic

material. Thus, several adjustments in sample preparation protocols have been included.

These approaches improve viral sequence enrichment after sequencing. However, some of these could present disadvantages or are limited to a specific group of viruses. In addition to these approaches, the depletion of ribosomal RNA for plant ribosomal RNA elimination, the enrichment of Double-Stranded (ds) RNA using cellulose or lithium chloride useful for dsRNA viruses sequencing, sequencing from small interfering RNA (sRNA) and RNA or DNA isolation directly from partially or totally purified viral particles, known as Viral Associated Nucleic Acid (VANA), are included (Roossinck *et al.*, 2015; Adams and Fox, 2016; Pecman *et al.*, 2017; Jeske, 2018).

The isolation of nucleic acids and the enrichment of viral sequences are followed by quality assessments, cDNA library construction, sequencing using shotgun sequencing or whole RNA sequencing (RNA-Seq), and Bioinformatics analyses. Sequencing steps, except for cDNA fragmentation, which is the first procedure for SGS, both SGS as TGS have in common the library construction, in which cDNA templates are prepared and modified using procedures according to the requirements of a specific sequencing platform. It is essential to mention that for TGS, DNA isolation is a crucial step; hence the objective is to obtain the longest reads as possible, therefore a high DNA quality and integrity is required (Kchouk *et al.*, 2017). Following, when a genome sequencing project is planned, estimated costs should include both the library construction and sequencing. In developing countries, whole sequencing projects by cost-benefit and time economy usually are led as a service with an external company or in collaboration with foreign universities. Currently, costs quoted by sequencing companies are between USD 500 (for SGS platforms) to USD 1500 (for TGS platforms), including library construction, sequencing, and genome assembly.

Sequencing platforms applied to viral genome sequencing has had a fast evolution, and in a short time, platforms such as GS FLX 454, which was highly used in the past, currently is discontinued. Further, Illumina platforms such as MiSeq and HiSeq series have taken advantage in the market place; however, its main limitation lies in the short length of the reads they generate which hinders the genome assembling of viruses present in similar samples sharing a high degree of sequence identity. In contrast, TGS platforms such as PacBio offer very long read; yet, they have a lower throughput, higher error rates, and higher costs per sequenced base (Rhoads and Au, 2015). Although for TGS, the Throughput has significantly increased in the last years, its major drawback continues to be its high error rates (mainly in ONT) (Villamor *et al.*, 2019). Nevertheless, the use of TGS to obtaining plan-virus genomes is highly promising.

Recently, a new method combining PCR-free, circular DNA enrichment (RCA) and TGS (SMRT-PacBio, Pacific

Biosciences Inc.) named as CIDER-Seq (Circular DNA Enrichment Sequencing) was successfully applied to obtain the full-length sequence of a DNA-viruses without the requirement of an assembling step and therefore reducing the bias that the short length reads assembly generates in the SGS technologies (Mehta *et al.*, 2019). In the last two years, the use of ONT technologies to sequence plant-virus genomes has been increasing. The MinION device was first applied for detecting whole genomes of virus maize streak virus, maize yellow mosaic virus and maize totivirus associated with maize lethal necrosis symptoms (Adams *et al.*, 2017), latter on 2018 was used for detecting plum pox virus in plum plants (Bronzato Badial *et al.*, 2018) and for detecting the yam viruses *Dioscorea bacilliform virus*, *yam mild mosaic virus* and *yam chlorotic necrosis virus* (YCNV) (Filloux *et al.*, 2018). In 2019 MinION also was efficiently used for detecting the *Wheat streak mosaic virus* (WSMV), *Barley yellow dwarf virus* (BYDV) and *Triticum mosaic virus* (TriMV) in wheat plants (Fellers *et al.*, 2019) and for detecting a novel bipartite begomovirus infecting cowpea plants (Naito *et al.*, 2019).

Another crucial variable to consider is the time elapse from sample collection to data collection and analysis. Time taken for DNA isolation could comprise several months, sequencing process including the shipment, take less than a month, and here is necessary to highlight that the step which could be the most arduous and time-demanding is the bioinformatics analysis, which requires computational expertise and full dedication to data curation, annotation, and genomic structures analyses.

Globally, several advances have been achieved, and novel viruses have been identified using NGS. Applications of NGS for detection of plant viruses using different approaches in sample preparation have been well reviewed (Barba *et al.*, 2014; Roossinck *et al.*, 2015; Hadidi *et al.*, 2016). However, in Colombia, the application of NGS in plant virology has been little reviewed. In the country, agro-industry is one of the most important sectors for economic development, and viruses represent a severe problem for crop production (Rodríguez *et al.*, 2016). The economic proposal for agricultural sector 2006-2020 is based on ten groups with exporting potential, which includes mainly exotic and tropical fruits (Rodríguez *et al.*, 2016). Thus, plant virus genomics, which has had an excellent performance in the last years, has been focused on these products.

Several studies (Table 2) involving the use of NGS for genome sequencing (complete or partial), detection, and characterization of RNA viruses infecting potato (*S. tuberosum* and *S. phureja*) (Gutiérrez-Sánchez *et al.*, 2014; Villamil-Garzón *et al.*, 2014; Gutiérrez *et al.*, 2016; Muñoz *et al.*, 2016; Vallejo C *et al.*, 2016), peanut (Gutiérrez Sánchez *et al.*, 2016), rice (Jimenez *et al.*, 2018), cassava (Carvajal-Yepes *et al.*, 2014), exotic fruits, such as *Solanum quitoense* (Gallo *et al.*, 2018), *Solanum betaceum* (Gutiérrez *et al.*, 2014), *Physalis peruviana* (Gutiérrez *et al.*, 2015),

Passiflora edulis (Jaramillo Mesa *et al.*, 2018; Jaramillo Mesa *et al.*, 2019), and horticultural products (Gutiérrez *et al.*, 2017; Muñoz Baena *et al.*, 2017) have been included. The above studies have been conducted using mainly total RNA isolated from symptomatic virus infected plants followed by depletion of rRNA and whole RNA sequencing. Plant viral genome reconstruction has been obtained by combining *de novo* transcriptome assembly for filtering contigs and sequences related to the virus and genome reference based on transcriptome assembly; however, few studies (Carvajal-Yepes *et al.*, 2014; Jimenez *et al.*, 2018) have been conducted using sRNA from virus-infected tissues.

Despite of advances achieved in the detection and identification of RNA viruses infecting plants in Colombia, additional studies are still needed for the discovery of DNA viruses. Colombia possesses a wide diversity of plants with potential for agribusiness. However, viral diseases are also a significant challenge, and many of these diseases have remained uncharacterized. Therefore, the advent of NGS

sequencing is an opportunity to advance the discovery of novel plant viruses in Colombia.

NGS IN TRANSCRIPTOMICS TO ELUCIDATE PLANT-VIRUS INTERACTIONS

Considering that some viruses contain RNA as genetic material and some viruses transcribe their DNA genetic material into RNA, whole transcriptome sequencing, as discussed above, has been widely used to obtain the viral genome from plant virus-infected transcriptomes. Besides, another critical application of transcriptomics in plant virology is the study of a given plant-virus interaction. Plant viruses have small genomes coding just for a small number of pivotal proteins; thus, processes such as virus replication and movement are carried out by hijacking host cell components.

Virus components interact and disturb the plant process through the entry and cell infection. A battle between virus and

Table 2. Viromics and virus genomes characterization in Colombia.

| Plant | Viruses | Covering | Accession Number | Sequencing technology | Source | Reference |
|-----------------------------|--|------------------|---|------------------------|---|--|
| | Cucumber mosaic virus (CMV) | | MG696854 MG696855 MG696856 | | | |
| <i>Solanum quitoense</i> | Potato yellow vein virus (PYVV) | Complete | MG696857 MG696858 MG696859 | | | Gallo <i>et al.</i> , 2018 |
| | Alstroemeria necrotic streak virus (ANSV) | | MG696853 MG696852 MG696851 | | | |
| <i>Capsicum annuum</i> | Bell pepper endornavirus (BPEV) | Partial | KX977569 KX977568 | | | Gutiérrez <i>et al.</i> , 2017 |
| <i>Solanum betaceum</i> | Leaf malformation virus (TaLMV) has been proposed | Complete | KM523548 | | | Gutiérrez <i>et al.</i> , 2014 |
| <i>Physalis peruviana</i> | Potato virus X (PVX). | Complete | KM659859 | Illumina HiSeq-2500 | RNA total with ribosomal RNA depletion | Gutiérrez <i>et al.</i> , 2015 |
| <i>Arachis pintoi</i> | Virome characterization: Peanut mottle virus, Turnip yellows virus and two Allexivirus related variants. | Complete genomes | KU708532 (PeMo-V); KU726090-1 (TuYV); KX058345 (Arachis pintoi virus A); KX058346 (Arachis pintoi virus B). | | | Gutiérrez Sánchez <i>et al.</i> , 2016 |
| <i>Solanum phureja</i> | Potato virus V (PVV) | Complete | KP849483 | | | Gutiérrez <i>et al.</i> , 2016 |
| <i>Passiflora edulis</i> | Passion fruit yellow mosaic virus (PFYMV) | Complete | KY823429 | | | Jaramillo Mesa <i>et al.</i> , 2019 |
| <i>Solanum tuberosum</i> | Potato virus Y (PVY) | Complete | KR149260 KX531041 | | | Muñoz <i>et al.</i> , 2016 |
| <i>Solanum lycopersicum</i> | Potato yellow vein virus (PYVV) | Complete | KX573901 KX573902 KX573903 | | | Muñoz Baena <i>et al.</i> , 2017 |
| <i>Solanum phureja</i> | Potato virus S (PVS) | Complete | KR152654 | | | Vallejo C <i>et al.</i> , 2016 |
| <i>Passiflora edulis</i> | Soybean mosaic virus (SMV) | Complete | KY249371-7 | | | Jaramillo Mesa <i>et al.</i> , 2018 |

(Continued)

Table 2. Viromics and virus genomes characterization in Colombia.

| Plant | Viruses | Covering | Accession Number | Sequencing technology | Source | Reference |
|--------------------------|--|---------------------------|--|---------------------------|-----------|--|
| <i>Solanum phureja</i> | Virome characterization Potato virus S (PVScol and PVS A), Potato virus V (PVV) and Potato leafroll virus (PLRV) | partial sequences | Not provided | 454 GS-FLX | RNA total | Gutiérrez-Sánchez <i>et al.</i> , 2014 |
| <i>Oryza sativa L</i> | Rice hoja blanca tenuivirus (RHBV) | Complete | MG566074 MG566075 MG566076 MG566077 | Illumina not specified | sRNA | Jimenez <i>et al.</i> , 2018 |
| <i>Manihot esculenta</i> | Viruses complex belonging to the families Secoviridae, Alphaflexiviridae and Luteoviridae | partial sequences | KC505249 KC505252 KC 505250 KC 505151 | Illumina platform (GSIII) | sRNA | Carvajal-Yepes <i>et al.</i> , 2014 |
| <i>Solanum tuberosum</i> | Detection of co-infection between Potato yellow vein virus (PVV), The Potato virus Y (PVY), Potato virus V (PVV) | PVY complete, PVV partial | Not provided | Illumina not specified | sRNA | Villamil-Garzón <i>et al.</i> , 2014 |

plant defense systems begins, and the success of viral infection will depend on the plant's ability to resist the virus attack. When the virus overcomes the plant responses, the plant became susceptible, and disease symptoms are triggered. Therefore, in plant-virus interactions, a high transcriptome reprogramming is given, and several sets of genes are modulated in response to the viral infection (Hanley-Bowdoin *et al.*, 2004; Bengyella *et al.*, 2015; Zanardo *et al.*, 2019). The study of the molecular bases of plant-virus interaction is essential for finding genes or pathways involved in plant defense or critical components for virus infection, replication, or movement. Taken together, this knowledge is pivotal for crop improvement programs directed at plant resistance to viral diseases (Hillung *et al.*, 2016; Zanardo *et al.*, 2019).

Among approaches for transcriptome study, currently, whole transcriptome sequencing commonly called RNA-Seq is one of the most widely applied. This technique is aimed to study of all sets of transcripts expressed in a given organism in a given condition (Mosa *et al.*, 2017). Overall, the methodology applied to the plant-virus interaction consists of total RNA isolation from plant-infected or virus-inoculated plants and plant-virus free or mock-inoculated samples, followed by the cDNA library construction and subsequent sequencing using an NGS platform. The most widely used platforms for RNA-Seq are Illumina HiSeq 2000, 2500, and 4000 series. However, with the introduction of the NovaSeq series, also Illumina platforms, this picture probably will change in brief. The current costs for RNA-Seq range from 250 to 450 USDs for library construction and sequencing of approximately 20-40 M reads per sample. Similar to the genome sequencing, after sample preparation, the sequencing process is fast and could take about one month, while bioinformatics analyses could take several months depending on the expertise.

RNA-Seq has been used to find insights in several plant-virus interactions, including model plants such as *Arabidopsis thaliana* (Sun *et al.*, 2016; Wu *et al.*, 2016), plants of agronomical importance such as *Oryza sativa* (Blazquez *et al.*, 2013; Wong *et al.*, 2015), *Zea mays* (Chakrabarty *et al.*, 2018), *Manihot esculenta* (Fang *et al.*, 2014; Amuge *et al.*, 2017; Anjanappa *et al.*, 2018), *Solanum tuberosum* (Goyer *et al.*, 2015; Stare *et al.*, 2017) and tropical fruits such as *C. papaya* (Madroño *et al.*, 2018). However, to date (04-10-2019), a Google Scholar search, the NCBI PMC database, and the Web of Science using the queries "plant", "virus", "interaction", "transcriptome", "RNA-Seq", and "Colombia" with different combinations did not yield any results for plant-virus interactions research developed in Colombia.

PROTEOMICS TO BETTER UNDERSTAND PLANT-VIRUS INTERACTIONS

The term proteomics has been used to denote the "PROtein complement of a genOME". It groups all techniques employed for protein studies in a given condition which involves protein characterization localization, interactions, post-translational modifications and others (Wilkins *et al.*, 1996). Currently, proteomics has allowed identifying a complete set of proteins, present in a particular organism submitted to changes in physiological activities and/or caused by external factors which, is technically termed as "sub-proteomes" (Peck, 2018; Zaynab *et al.*, 2018). Plant proteomics became widely applied around 2000, and its use has increased thanks to the advent of NGS technologies. However, transcriptome, and proteome are not always correlated this lack of correlation is mainly due to the post-transcriptional regulatory processes; however

post-translational modifications also play a crucial role. For that reason, proteome and transcriptome should be taken as complementary tools.

Usually, proteomics could be used to identify proteins from purified virus, virus-enriched samples, or using plant virus-based expression systems. However, in plant-virus interactions, the most common application is using total proteins from infected tissues. Like transcriptomics, in proteomics, the qualitative or quantitative proteome profile of a specific plant previously submitted for virus infection is obtained. Further, an analysis to identify the differentially abundant proteins is conducted in non-infected cells as compared with infected cells. This approach allows identification of host or viral proteins that play critical roles in infection or that evade host defense mechanisms (Xu and D. Nagy, 2010).

For proteomics, it is crucial to know the preparation method and properties of the sample and protein extract. In order to cover a wide range of sample types, improve data coverage, and obtain accurate quantification, different methods have been developed for proteomic analysis, which owing to its high sensitivity can detect changes in the abundance of proteins in different types of samples (Li *et al.*, 2012).

Proteomics is classified into three main groups: I) Expression proteomics studies which are addressed to the investigation of protein expression. These studies can be conducted using conventional (Chromatography, ELISA Western blotting) and/or advanced techniques (Gel-based approaches, Mass spectrometry, Edman sequencing). II) Structural proteomics studies are analyzed using high throughput techniques (X-ray crystallography) or bioinformatics tools for protein structure prediction. These last are very well reviewed in (Roy and Zhang, 2012). Even when this work is from 2012, contains several bioinformatics tools which are still widely used. III) Functional proteomics, which is the focus of this review, comprises the quantitative techniques which allow to understand the protein functions as well as elucidating unknown molecular mechanisms, protein interactions, and associations of an unrevealed protein with partners from a given protein complex (Chandrasekhar *et al.*, 2014).

In plant virology, as described above, it should be taken into account that viruses encode small proteomes (1–2500 proteins) and that the proteomes of viruses and plants are determined by the virus-host protein interaction, wherein the virus focuses on ensuring its infective replication and the plant focuses on blocking the virus infection. Plant responses to virus infections are speedy, and a drastic change in the protein accumulation is triggered in the whole plant, this protein accumulation provides the crucial clues to understand the plant-virus interaction and the resistance mechanisms to the virus infection (Kundu *et al.*, 2013; Varela *et al.*, 2017; Souza *et al.*, 2019). On the other hand, leaves

are a principal organ for studying plant-virus interactions, because these generally exhibit necrotic patches or morphological variations that allows to visually detect the first infection symptoms (Di Carli *et al.*, 2010; Kundu *et al.*, 2013; Varela *et al.*, 2017; Souza *et al.*, 2019). However, not always symptoms are evident, what makes then even more necessary the implementation of reliable and specific techniques, such as proteomics to assess the protein levels and interactions under such conditions (Mochida and Shinozaki, 2011; Mosa *et al.*, 2017).

Not-targeted proteomics is the most commonly used technique. In this technique, proteins are identified through a process known as a shotgun, in which proteins are digested in small peptides using different proteases. The obtained peptides are analyzed by Liquid chromatography-mass spectrometry (LC-MS). However, this approach could produce unreliable and irreproducible results (Li *et al.*, 2012). The most widely applied method for the simultaneous identification and quantification of proteins is the label-free method. Label-free is a low-cost and straightforward method which do not require the use of stable isotopes carried in the sample, instead, the signal intensities (peaks areas) or spectral counts of peptides belonging to a specific protein are correlated with the protein amount present in a sample obtained under different conditions or treatments (Li *et al.*, 2012). Nonetheless, label-free presents poor yields in the quantification of low abundant proteins (Choi *et al.*, 2008; Vowinkel *et al.*, 2014).

Recently, chemical or enzymatic methods for protein or peptides labeling using different isotopes tags *in vitro* or *in vivo* have been developed. These tags include the ICAT (Isotope-coded affinity tag), which, after protein digestion, the C-terminal of peptides is tagged using the isotope H₂18O and ReDi (reductive dimethylation) and the N-terminal and lysine lateral chain of peptides are tagged by reductive demethylation. However, some authors have suggested that methods based on isobaric chemical labeling such as iTRAQ (Isobaric Tags for Relative and Absolute Quantitation), Stable isotope labeling by amino acids in cell culture (SILAC) and TMT (Tandem Mass Tag) have been found to be most successful when applied to samples previously analyzed by LC-MS and IMAC (Immobilized metal affinity chromatography, a strategy for enriching phosphorylated peptides) (de la Fuente van Bentem *et al.*, 2006; Jayaraman *et al.*, 2012).

Around the world, proteomics has been widely applied to study plant-virus interactions. It includes crops of economic importance, such as maize and Soybean (Wu *et al.*, 2013; Pavan Kumar *et al.*, 2016), and also plant models, such as tobacco and tomato (Di Carli *et al.*, 2010; Lin *et al.*, 2015; Alexander and Cilia, 2016; Megias *et al.*, 2018). For those readers interested in deepening, Souza *et al.*, (2019) present an excellent review of proteomics applied to several studies of plant-virus interactions. In Latin-America, proteomics is less

addressed; basically, the country that is leading the research studies on structural, expression and functional proteomics in plant-virus interactions is Brazil. Some examples of researches conducted in Brazil are the study of the chloroplast proteomic profile in Tomato blistering mosaic virus (ToBMV) and *Nicotiana benthamiana* interaction (Megias *et al.*, 2018), the host proteomic response to Citrus tristeza virus (Dória and Pirovani, 2019) and to the Papaya meleira virus complex (PMeV) (Soares *et al.*, 2017). In Colombia, however, similarly to transcriptomics, proteomics has been little explored. To date (16-07-2019), a detailed search on Google Scholar, the NCBI PMC database, and the Web of Science, using the queries “plant”, “virus”, “interaction”, “proteomics”, “functional proteomics” and “Colombia” with different combinations did not yield any results for functional proteomics applied to plant-virus interactions research developed in Colombia.

BIOINFORMATICS RESOURCES AND OMICS TOOLS FOR PLANT-VIROLOGY STUDIES

There are many bioinformatics tools and resources that ease experimental workflows. Each omic has its specific bioinformatics pipeline; however, there are common databases and resources for analysis of data from all omics. Here, the main tools, bioinformatics pipelines, and databases for public data acquisition and data visualization focused on viruses and plants are introduced.

Overall, for plant virus genome discovery, after sequencing, bioinformatics is used for sequence quality analysis and trimming, *de novo* assembly of contigs, and similarity-based searches for finding virus-related scaffolds and contigs. Once viral contigs are obtained, these are used for an extending virus genome contigs using *de novo* assembly or reference genome-based assembly (Blawid *et al.*, 2017). Currently, there are several open-source, free, and fee-based software (Table 3), used for bioinformatics analyses of plant viruses. Among the licensed software, the most commonly used include CLC Genomics Workbench, Geneious, and DNASTAR. Although these tools offer a large number of resources for genome and transcriptome analyses, there are also freely available software with more user-friendly interfaces, such as VirAmp (Wan *et al.*, 2015), which is a tool included in the Galaxy project, and VirusTap (Yamashita *et al.*, 2016), which was released on 2016. For those who are familiar with command lines and prefer to have access to additional sets and parameters, multiple open source assemblers are also available, which are described in detail elsewhere (Blawid *et al.*, 2017).

When the objective is the analyses of plant-virus interactions, similar to the common virus discovery strategy, bioinformatics analyses are carried using the transcriptome data. However, in this approach, the focus is the plant

instead of virus sequences. In comparison to the virus-genome transcriptome-based discovery, the study of plant-virus interactions is directed to find all transcriptome alterations caused by virus infection in the plant. Therefore, quantitative analyses are prioritized, and a higher depth in sequencing is required. For bioinformatics analyses, there are also licensed software and free web interface-based resources such as Galaxy (Goecks *et al.*, 2010) and RobiNa (Lohse *et al.*, 2012) available, which are including pipelines for RNA-Seq. However, because of a large amount of generated data, its processing demands high computational power. Thus, it is recommended to use programs that are run on UNIX-like operative systems and directly using a command line interface that exposes all the parameters of the different packages/tools. Overall, transcriptome analyses of plant-viral interactions involve analysis of sequence quality and trimming, transcriptome assembly, annotation, quantification, and differential expression analyses between infected and control plants.

For reference genome-based transcriptome reconstruction, the most commonly used short reads mappers are TopHat2 (Kim *et al.*, 2013), Star (Dobin *et al.*, 2013), and HiSat2 which is the replacement of TopHat nowadays (Kim *et al.*, 2015). For *de novo* transcriptome assemblers, the commonly used versions included Trinity (Haas *et al.*, 2013), SOAPdenovo-Trans (Xie *et al.*, 2014), Trans-ABYSS (Robertson *et al.*, 2010), IDBA-tran (Peng *et al.*, 2013), and Oases (Schulz *et al.*, 2012). Once transcriptome assembly is completed, before the differential analyses of gene/transcript expression, the abundance levels of genes should be estimated. Unlike genome sequencing, in the transcriptome, sequence coverage is indicative of abundance levels. Therefore, the developed methods for quantification are based on the reads counts or normalized reads counts belonging to a given exon, gene, or transcript.

These methods are grouped into two categories: union exons and transcript-based. While programs such as featureCounts (Liao *et al.*, 2013) and HTSeq-count (Anders *et al.*, 2014) are based on the union exon method, programs such as Cufflinks (Trapnell *et al.*, 2012), RSEM (Li and Dewey, 2011), BitSeq (Glaus *et al.*, 2012), and the recently implemented programs such as Sailfish (Patro *et al.*, 2014), RapMap (Srivastava *et al.*, 2016), Kallisto (Bray *et al.*, 2016), and Salmon (Patro *et al.*, 2017) are transcript-based methods. However, these last are approaches based on algorithms such as pseudo-alignment, lightweight mapping, or quasi-mapping. For these methods, the reads are broken in short-sequences named as k-mers. These characteristics increase the accuracy and efficiency of these quantification methods.

Once the expression levels are estimated, these data are incorporated in programs developed to conduct differential expression analysis. These programs use different statistical methods in their analyses. Programs such as edgeR (Robinson *et al.*, 2009), baySeq (Hardcastle and Kelly,

Table 3. Commonly and friendly-interface omics Tools and web resources used in the study of plant virology.

| | Tool | Source type | Applications | URL |
|------------------------------------|-------------------------------------|-------------|---|---|
| VIROMICS AND TRANSCRIPTOMICS | DNASTAR | Licensed | Genome assembly; Transcriptome, annotation; visualization. | https://www.dnastar.com/ |
| | CLC Genomics Workbench | Licensed | Genome assembly; Transcriptome, annotation; visualization. | https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/ |
| | Geneious | Licensed | Genome assembly; Transcriptome, annotation; visualization; blast and local blast. | https://www.geneious.com/ |
| | VirAmp | Free | Genome assembly and analyses pipeline. | https://galaxyproject.org/use/viramp/ |
| | VirusTap | Free | Viral Genome-Targeted Assembly Pipeline. | https://gph.niid.go.jp/cgi-bin/virustap/index.cgi |
| | ViralZone | Free | Virus web resource. | https://viralzone.expasy.org/ |
| PLANT-VIRUS INTERACTION | RobiNA | Free | RNA-Seq; plant-virus interactions. | http://mapman.gabipd.org/robin |
| | Galaxy | Free | RNA-Seq; plant-virus interactions. | https://galaxyproject.org/tutorials/rb_maseq/ |
| PROTEOMICS | Maxquant | Free | Proteomics analyses. | https://www.maxquant.org/ |
| | Progenesis QI | Licensed | Proteomics analyses. | http://www.nonlinear.com/progenesis/qi/ |
| | Peaks | Licensed | Proteomics analyses. | http://www.bioinform.com/peaks-studio/ |
| | OpenMS | Free | Proteomics analyses. | https://www.openms.de/ |
| | MASCOT | Licensed | Protein databases search engine. | http://www.matrixscience.com/ |
| | SpectrumMill | Licensed | Protein databases search engine. | http://proteomics.broadinstitute.org/ |
| | X! Tandem | Free | Protein databases search engine. | https://www.thegpm.org/tandem/ |
| PATHWAYS AND DATA VISUALIZATION | Perseus | Free | Omics data visualization and statistical graphs. | https://maxquant.net/perseus/ |
| | MeV | Free | Omics data visualization and statistical graphs. | http://mev.tm4.org/#/welcome |
| | MapMan | Free | Whole Metabolic Pathways analyses. | https://mapman.gabipd.org/ |
| | Cytoscape | Free | Whole Metabolic Pathways analyses. | https://cytoscape.org/ |
| | Kegg | Free | Whole Metabolic Pathways analyses. | https://www.genome.jp/kegg/pathway.html |
| DATABASES | Virus-host DB | Free | Virus web resource. | https://www.genome.jp/virushostdb/index/virus/all |
| | Descriptions of Plant Viruses (DPV) | Free | Virus web resource. | http://www.dpvweb.net/seqs/plantviruses.php |
| | Phytozome | Free | Plant genomics database. | https://phytozome.jgi.doe.gov/pz/portal.html |
| | PLAZA | Free | Plant genomics database. | https://bioinformatics.psb.ugent.be/plaza/ |
| | PlantGDB | Free | Plant genomics database. | http://www.plantgdb.org/ |
| | virusSITE | Free | Viral genomics database. | http://www.virusite.org/index.php |

2010), DESeq (Anders and Huber, 2010; Love *et al.*, 2014), and Cuffdiff (Trapnell *et al.*, 2013) are widely used for conducting differential expression analysis. The output of

these programs is spreadsheets or plain text files containing the ID of genes or transcripts and its corresponding statistical data.

For proteomics analyses, data analysis is highly dependent on the used Mass Spectrometry instrument, therefore, the raw files generated can be very variables; however, among the best software for complement quantitative analyses are the MaxQuant (Tyanova *et al.*, 2015) and Progenesis QI (licensed), which are used to analyze, identify, and quantify high-throughput mass spectrometry data. For PEAKS (licensed), this software is used to identify peptides and posttranslational modifications from *de novo* sequence peptides analyses. OpenMS (Röst *et al.*, 2016) is an open source resource based on C++ libraries for LC-MS data analyses. Subsequently, using the obtained mass spectrometry data, proteins can be identified by querying in peptide sequence databases using licensed software such as MASCOT, SEQUEST, Phenyx, SpectrumMill, and IdentityE or open source resources such as X! Tandem (Bjornson *et al.*, 2008) and OMSSA (Geer *et al.*, 2004). The final output of these programs is also filed with the protein ID and corresponding statistical and quantitative data.

Using the output files obtained from proteomics and transcriptomics analyses, data visualization, and pathways analyses for those who are familiar with the R language could be conducted using several R packages. However, there are also free friendly interface tools, which require as input, a file with the gene/transcript or protein IDs and results of the quantitative analyses. This group of free resources include Perseus (Tyanova *et al.*, 2016; Tyanova and Cox, 2018) and MeV (Howe *et al.*, 2011), these programs possess several tools for statistical graphs, Heatmaps, cluster and functional enrichment analyses. For metabolic pathways analyses, among the most widely used programs are MapMan (Thimm *et al.*, 2004), Kegg (Kanehisa and Goto, 2000) and Cytoscape (Shannon, 2003). This last, although is a generic tool for the analysis and visualization of network data, is a powerful resource which works with plug-ins and applications for several tasks including plug-ins for metabolic pathways analyses such as MetScape, MetaNetter 2, GeneMANIA and KEGGscape.

Last but not least, for data storage and acquisition, there are also many public repositories. Among these, there is available the International Nucleotide Sequence Database Collaboration (INSDC) which is a collaboration between the DNA Data Bank of Japan (DDBJ), The European Bioinformatics Institute (EMBL-EBI) and The National Center for Biotechnology Information (NCBI) for making available and gathering the most extensive public repositories of nucleic acid sequence data around the world. NCBI also provides more specific resources such as The Reference Sequence (RefSeq) collection (O'Leary *et al.*, 2016), and the Sequence Read Archive (SRA), which store a large number of NGS bio-projects both transcriptome and genome data. Another useful Database for sequences storing and analyses is the Protein Data Bank (PDB) (Rose *et al.*, 2012) which contain the Protein Data information about the 3D shapes of proteins, nucleic acids, and complex assemblies.

More specifically, databases for the study of plant virology include those focused on virus genomes, such as the International Committee on Taxonomy of Viruses (ICTV). Although this database does not include sequence information, it is a useful tool for determining an officially reported virus and its taxonomy. The ViralZone (Hulo *et al.*, 2011) is a web resource that combines information about the virus host range, replication cycle, and virion structure with genomic and proteomic sequences. The Virus-host DB (Mihara *et al.*, 2016) is a good resource for finding virus sequences by searching the virus name or host taxonomy. Other databases specific for plant viral genomics are the Descriptions of Plant Viruses (DPV) (Adams, 2006) and viruSITE (Stano *et al.*, 2016). Among the databases that contain plant genome sequence information and tools for annotation are Biomart, Blast, Bioextract, Phytozome (Goodstein *et al.*, 2012), PlantGDB (Dong, 2004), and PLAZA (Proost *et al.*, 2015).

Taken together, all these resources highly facilitate work with large amounts of data generated by large-scale omics approaches and help to provide biological context and an explanation for some given phenomena from computational analyses.

FUTURE PERSPECTIVES

NGS technologies and 'omic' technologies have opened the door to immense knowledge regarding the genomics, transcriptomics and proteomics of organisms. Further, these advances have helped discover the high diversity between and within species. The genetic analysis of a unique gene become in an analysis of multiples genes, an analysis of a complete genome. NGS enables the determination of the complete genetic sequence even of various interacting organisms. The challenge, however, lies in establishing the relationship between them. In Colombia, the use of NGS is very scarce, and even more so for bioinformatics analyses. Currently, bioinformatics analyses are a bottleneck for progress in Colombian research because these are time-demanding and require qualified personnel; for instance, the comparative genome analyses are increasingly delayed regards to the massive amounts of data that are deposited in databases globally. To facilitate workflow, Figure 1 summarizes the general process to follow for plant virology studies using NGS technologies. Based on the reports of Colombian research teams that have used NGS technologies and our own experience, it is necessary to design strategies to extensive the use the NGS and 'omic' technologies in the country. However, it is more urgent to increase the number of researchers trained to perform bioinformatics studies and to create and strengthen collaborations for genome sequencing projects.

The use of NGS and "omics" technologies in the study of plant-virus could allow to the Colombian researchers

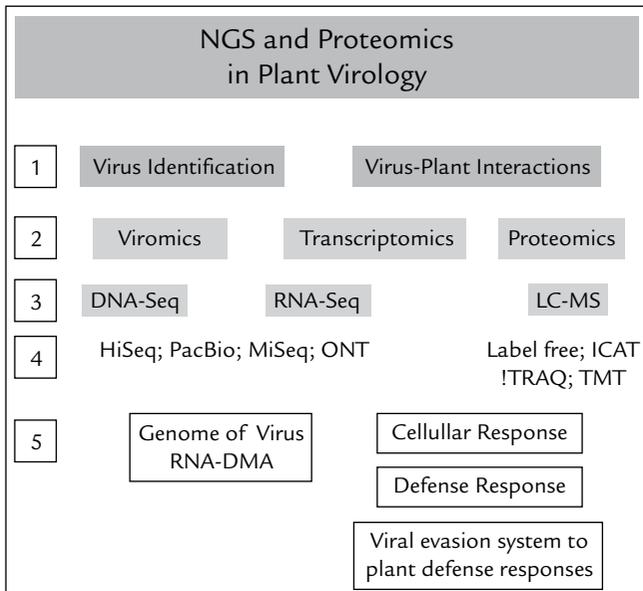


Figure 1. Strategies (NGS and proteomics) for the study of viruses in plants. First, the objectives must be established: identify viruses or evaluate how they interact with plants (1). According to this, you must choose which is the most relevant strategy between Genomics, Transcriptomics and / or Proteomics (2) and which is the technique (3) and platform (4) most suitable in terms of costs, throughput and quality of information collected and evaluated. Finally, with the data analysis you can identify the viruses (DNA or RNA) present in your sample, or establish the cellular responses, the defense mechanisms established by the plants against virus, and vice versa (5).

to analyze the complete view of any biomolecules into the organism, a complete information of genome, how many, which and where are located the virus genes in the host, which and how many of them are being transcribed and translated. Viral genomics allows to identify new viruses and to establish the relationship among them. The comparative genomic analysis will reveal the genetics mechanisms of entrance and movement of virus into the cell. Also, the proteomic allow us to know the changes in the host protein profile induced by the expression of the virus proteins. Taken together, omics used as complementary tools are crucial for revealing the molecular mechanisms determining resistance or susceptibility, as well as finding genes with critical roles in the viral infection processes which can be used in crop improvement programs.

ACKNOWLEDGMENTS

We gratefully acknowledge to Vice Chancellery for Research of Universidad El Bosque. This study was funded by Vice Chancellery for Research of Universidad El Bosque. LJM was partially funded by the Departamento Administrativo de Ciencia, Tecnología e Innovación, Colciencias (grant number: 1308-71250819). The funders had no role in study

design, data collection and interpretation, or the decision to submit the work for publication.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

- Adams MJ. DPVweb: a comprehensive database of plant and fungal virus genes and genomes. *Nucleic Acids Res.* 2006;34(90001):D382-D385. Doi: <http://dx.doi.org/10.1093/nar/gkj023>
- Adams I, Fox A. Diagnosis of Plant Viruses Using Next-Generation Sequencing and Metagenomic Analysis. In: Wang A, Zhou X, editors. *Current Research Topics in Plant Virology*. Cham: Springer International Publishing, 2016. p. 323-335.
- Adams IP, Braidwood LA, Stomeo F, Phiri N, Uwumukiza B, Feyissa B, *et al.* Characterising maize viruses associated with maize lethal necrosis symptoms in sub Saharan Africa. *bioRxiv*. 2017:161489. Doi: <http://dx.doi.org/10.1101/161489>
- Alexander MM, Cilia M. A molecular tug-of-war: Global plant proteome changes during viral infection. *Curr. Plant Biol.* 2016;5:13-24. Doi: <http://dx.doi.org/10.1016/j.cpb.2015.10.003>
- Amuge T, Berger DK, Katari MS, Myburg AA, Goldman SL, Ferguson ME. A time series transcriptome analysis of cassava (*Manihot esculenta* Crantz) varieties challenged with Ugandan cassava brown streak virus. *Sci. Rep.* 2017;7(1). Doi: <http://dx.doi.org/10.1038/s41598-017-09617-z>
- Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010;11(10):R106. Doi: <http://dx.doi.org/10.1186/gb-2010-11-10-r106>
- Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2014;31(2):166-169. Doi: <http://dx.doi.org/10.1093/bioinformatics/btu638>
- Anjanappa RB, Mehta D, Okoniewski MJ, Szabelska-Bereżewicz A, Gruissem W, Vanderschuren H. Molecular insights into Cassava brown streak virus susceptibility and resistance by profiling of the early host response. *Mol. Plant Pathol.* 2018;19(2):476-489. Doi: <http://dx.doi.org/10.1111/mpp.12565>
- Barba M, Czosnek H, Hadidi A. Historical Perspective, Development and Applications of Next-Generation Sequencing in Plant Virology. *Viruses.* 2014;6(1):106-136. Doi: <http://dx.doi.org/10.3390/v6010106>
- Bengyella L, Waikhom SD, Allie F, Rey C. Virus tolerance and recovery from viral induced-symptoms in plants are associated with transcriptome reprogramming. *Plant Mol. Biol.* 2015;89(3):243-252. Doi: <http://dx.doi.org/10.1007/s11103-015-0362-6>

- Bjornson RD, Carriero NJ, Colangelo C, Shifman M, Cheung K-H, Miller PL, *et al.* X!Tandem, an Improved Method for Running X!Tandem in Parallel on Collections of Commodity Computers. *J. Proteome Res.* 2008;7(1):293-299. Doi: <http://dx.doi.org/10.1021/pr0701198>
- Blawid R, Silva JMF, Nagata T. Discovering and sequencing new plant viral genomes by next-generation sequencing: description of a practical pipeline. *Ann. Appl. Biol.* 2017;170(3):301-314. Doi: <http://dx.doi.org/10.1111/aaab.12345>
- Blazquez MA, Zheng W, Ma L, Zhao J, Li Z, Sun F, *et al.* Comparative Transcriptome Analysis of Two Rice Varieties in Response to Rice Stripe Virus and Small Brown Planthoppers during Early Interaction. *PLoS One.* 2013;8(12):e82126. Doi: <http://dx.doi.org/10.1371/journal.pone.0082126>
- Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 2016;34(5):525-527. Doi: <http://dx.doi.org/10.1038/nbt.3519>
- Bronzato Badial A, Sherman D, Stone A, Gopakumar A, Wilson V, Schneider W, *et al.* Nanopore sequencing as a surveillance tool for plant pathogens in plant and insect tissues. *Plant Dis.* 2018;102(8):1648-1652. Doi: <http://dx.doi.org/10.1094/pdis-04-17-0488-re>
- Carvajal-Yepes M, Olaya C, Lozano I, Cuervo M, Castaño M, Cuellar WJ. Unraveling complex viral infections in cassava (*Manihot esculenta* Crantz) from Colombia. *Virus Res.* 2014;186:76-86. Doi: <http://dx.doi.org/10.1016/j.virusres.2013.12.011>
- Chakrabarty D, Ghorbani A, Izadpanah K, Dietzgen RG. Changes in maize transcriptome in response to maize Iranian mosaic virus infection. *PLoS One.* 2018;13(4):e0194592. Doi: <http://dx.doi.org/10.1371/journal.pone.0194592>
- Chandrasekhar K, Dileep A, Lebonah DE, Kumari JP. A Short Review on Proteomics and its Applications. *Int. Lett. Nat. Sci.* 2014;17:77-84. Doi: <http://dx.doi.org/10.18052/www.scipress.com/ILNS.17.77>
- Choi H, Fermin D, Nesvizhskii AI. Significance Analysis of Spectral Count Data in Label-free Shotgun Proteomics. *Mol. Cell. Proteom.* 2008;7(12):2373-2385. Doi: <http://dx.doi.org/10.1074/mcp.M800203-MCP200>
- De La Fuente Van Bentem S, Roitinger E, Anrather D, Cszasz E, Hirt H. Phosphoproteomics as a tool to unravel plant regulatory mechanisms. *Physiol. Plant.* 2006;126(1):110-119. Doi: <http://dx.doi.org/10.1111/j.1399-3054.2006.00615.x>
- Di Carli M, Villani ME, Bianco L, Lombardi R, Perrotta G, Benvenuto E, *et al.* Proteomic Analysis of the Plant-Virus Interaction in Cucumber Mosaic Virus (CMV) Resistant Transgenic Tomato. *J. Proteome Res.* 2010;9(11):5684-5697. Doi: <http://dx.doi.org/10.1021/pr100487x>
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21. Doi: <http://dx.doi.org/10.1093/bioinformatics/bts635>
- Dong Q. PlantGDB, plant genome database and analysis tools. *Nucleic Acids Res.* 2004;32(90001):354D-359. Doi: <http://dx.doi.org/10.1093/nar/gkh046>
- Dória MS, Pirovani CP. Proteomic Response of Host Plants to Citrus tristeza virus. 2019;2015:209-218. Doi: http://dx.doi.org/10.1007/978-1-4939-9558-5_15
- Fang DD, Maruthi MN, Bouvaine S, Tufan HA, Mohammed IU, Hillocks RJ. Transcriptional Response of Virus-Infected Cassava and Identification of Putative Sources of Resistance for Cassava Brown Streak Disease. *PLoS One.* 2014;9(5):e96642. Doi: <http://dx.doi.org/10.1371/journal.pone.0096642>
- Fellers J, Webb C, Fellers M, Shoup Rupp J, De Wolf E. Wheat virus identification within infected tissue using nanopore sequencing technology. *Plant Dis.* 2019. Doi: <http://dx.doi.org/10.1094/pdis-09-18-1700-re>
- Filloux D, Fernandez E, Loire E, Claude L, Galzi S, Candresse T, *et al.* Nanopore-based detection and characterization of yam viruses. *Sci. Rep.* 2018;8(1). Doi: <http://dx.doi.org/10.1038/s41598-018-36042-7>
- Gallo Y, Toro LF, Jaramillo H, Gutiérrez PA, Marín M. Identificación y caracterización molecular del genoma completo de tres virus en cultivos de lulo (*Solanum quitoense*) de Antioquia (Colombia). *rev.colomb.cienc.hortic.* 2018;12(2):281-292. Doi: <http://dx.doi.org/10.17584/rcch.2018v12i2.7692>
- Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, *et al.* Open Mass Spectrometry Search Algorithm. *J. Proteome Res.* 2004;3(5):958-964. Doi: <http://dx.doi.org/10.1021/pr0499491>
- Glaus P, Honkela A, Rattray M. Identifying differentially expressed transcripts from RNA-seq data with biological variation. *Bioinformatics.* 2012;28(13):1721-1728. Doi: <http://dx.doi.org/10.1093/bioinformatics/bts260>
- Goecks J, Nekrutenko A, Taylor J, Galaxy Team T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 2010;11(8):R86. Doi: <http://dx.doi.org/10.1186/gb-2010-11-8-r86>
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 2012;40(D1):D1178-D1186. Doi: <http://dx.doi.org/10.1093/nar/gkr944>
- Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 2016;17(6):333-351. Doi: <http://dx.doi.org/10.1038/nrg.2016.49>
- Goyer A, Hamlin L, Crosslin JM, Buchanan A, Chang JH. RNA-Seq analysis of resistant and susceptible potato varieties during the early stages of potato virus Y infection. *BMC genomics.* 2015;16(1). Doi: <http://dx.doi.org/10.1186/s12864-015-1666-2>
- Gutiérrez-Sánchez P, Alzate-Restrepo J, Marín-Montoya M. Caracterización del viroma de ARN en tejido radical de *Solanum phureja* mediante pirosecuenciación 454 GS-FLX. *Bioagro.* 2014;26(2):89-98.

- Gutiérrez P, Mesa HJ, Marín Montoya M. Genome sequence of a divergent Colombian isolate of potato virus V (PVV) infecting *Solanum phureja*. *Acta Virol.* 2016;60(01):49-54. Doi: http://dx.doi.org/10.4149/av_2016_01_49
- Gutiérrez PA, Alzate JF, Marín Montoya M. Genome sequence of a virus isolate from tamarillo (*Solanum betaceum*) in Colombia: evidence for a new potyvirus. *Arch. Virol.* 2014;160(2):557-560. Doi: <http://dx.doi.org/10.1007/s00705-014-2296-8>
- Gutiérrez PA, Alzate JF, Montoya MM. Complete genome sequence of an isolate of Potato virus X (PVX) infecting Cape gooseberry (*Physalis peruviana*) in Colombia. *Virus Genes.* 2015;50(3):518-522. Doi: <http://dx.doi.org/10.1007/s11262-015-1181-1>
- Gutiérrez PA, Marín-Montoya M, Muñoz-Baena L. Genome sequencing of two Bell pepper endornavirus (BPEV) variants infecting *Capsicum annuum* in Colombia. *Agron. Colomb.* 2017;35(1):44. Doi: <http://dx.doi.org/10.15446/agron.colomb.v35n1.60626>
- Gutiérrez Sánchez PA, Jaramillo Mesa H, Marín Montoya M. Next generation sequence analysis of the forage peanut (*Arachis pintoi*) virome. *Rev Fac Nac Agron.* 2016;69(2):7881. Doi: <http://dx.doi.org/10.15446/rfna.v69n2.59133>
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 2013;8(8):1494-1512. Doi: <http://dx.doi.org/10.1038/nprot.2013.084>
- Hadidi A, Flores R, Candresse T, Barba M. Next-Generation Sequencing and Genome Editing in Plant Virology. *Front Microbiol.* 2016;7. Doi: <http://dx.doi.org/10.3389/fmicb.2016.01325>
- Hanley-Bowdoin L, Settlege SB, Robertson D. Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. *Mol. Plant Pathol.* 2004;5(2):149-156. Doi: <http://dx.doi.org/10.1111/j.1364-3703.2004.00214.x>
- Hardcastle TJ, Kelly KA. baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinform.* 2010;11(1):422. Doi: <http://dx.doi.org/10.1186/1471-2105-11-422>
- Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics.* 2016;107(1):1-8. Doi: <http://dx.doi.org/10.1016/j.ygeno.2015.11.003>
- Hernandez R. Sequencing Within Reach. *Genet Eng Biotechnol N.* 2018;38(4):1,22,25. Doi: <http://dx.doi.org/10.1089/gen.38.04.01>
- Hillung J, García-García F, Dopazo J, Cuevas JM, Elena SF. The transcriptomics of an experimentally evolved plant-virus interaction. *Sci. Rep.* 2016;6(1). Doi: <http://dx.doi.org/10.1038/srep24901>
- Howe EA, Sinha R, Schlauch D, Quackenbush J. RNA-Seq analysis in MeV. *Bioinformatics.* 2011;27(22):3209-3210. Doi: <http://dx.doi.org/10.1093/bioinformatics/btr490>
- Hull R. Plant Virus Viromics. In: *Plant Virology.* Ciudad: Academic Press, 2014. p. 929-971.
- Hull R. Plant Viruses and Their Classification. In: Hull R, editor. *Plant Virology.* Ciudad: Academic Press, 2014. p. 15-68.
- Hulo C, De Castro E, Masson P, Bougueleret L, Bairoch A, Xenarios I, *et al.* ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.* 2011;39(suppl_1):D576-D582. Doi: <http://dx.doi.org/10.1093/nar/gkq901>
- International Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. *Nature.* 2004;431(7011):931-45. Doi: <http://dx.doi.org/10.1038/nature03001>
- Jaramillo Mesa H, Marín Montoya M, Gutiérrez PA. Molecular characterization of Soybean mosaic virus (SMV) infecting Purple passion fruit (*Passiflora edulis* f. *edulis*) in Antioquia, Colombia. *Arch Phytopathology Plant Protect.* 2018;51(11-12):617-636. Doi: <http://dx.doi.org/10.1080/03235408.2018.1505411>
- Jaramillo Mesa H, Marín Montoya MA, Gutiérrez Sánchez P. Complete genome sequence of a Passion fruit yellow mosaic virus (PFYMV) isolate infecting purple passionfruit (*Passiflora edulis* f. *edulis*). *Rev Fac Nac Agron Medellín.* 2019;72(1):8643-8654. Doi: <http://dx.doi.org/10.15446/rfnam.v72n1.69438>
- Jayaraman D, Forshey KL, Grimsrud PA, Ané J-M. Leveraging Proteomics to Understand Plant-Microbe Interactions. *Front. Plant Sci.* 2012;3. Doi: <http://dx.doi.org/10.3389/fpls.2012.00044>
- Jeske H. Barcoding of Plant Viruses with Circular Single-Stranded DNA Based on Rolling Circle Amplification. *Viruses.* 2018;10(9):469. Doi: <http://dx.doi.org/10.3390/v10090469>
- Jimenez J, Carvajal-Yepes M, Leiva AM, Cruz M, Romero LE, Bolaños CA, *et al.* Complete Genome Sequence of Rice hoja blanca tenuivirus Isolated from a Susceptible Rice Cultivar in Colombia. *Genome Announc.* 2018;6(7):e01490-17. Doi: <http://dx.doi.org/10.1128/genomeA.01490-17>
- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000;28(1):27-30. Doi: <http://dx.doi.org/10.1093/nar/28.1.27>
- Kchouk M, Gibrat JF, Elloumi M. Generations of Sequencing Technologies: From First to Next Generation. *Biol. Med.* 2017;09(03). Doi: <http://dx.doi.org/10.4172/0974-8369.1000395>
- Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods.* 2015;12(4):357-360. Doi: <http://dx.doi.org/10.1038/nmeth.3317>
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013;14(4):R36. Doi: <http://dx.doi.org/10.1186/gb-2013-14-4-r36>

- Kulski JK. Next-generation sequencing—an overview of the history, tools, and “Omic” applications. In: Next Generation Sequencing—Advances, Applications and Challenges. IntechOpen, 2016.
- Kundu S, Chakraborty D, Kundu A, Pal A. Proteomics approach combined with biochemical attributes to elucidate compatible and incompatible plant-virus interactions between *Vigna mungo* and Mungbean Yellow Mosaic India Virus. *Proteome Sci.* 2013;11(1):15. Doi: <http://dx.doi.org/10.1186/1477-5956-11-15>
- Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* 2011;12(1):323. Doi: <http://dx.doi.org/10.1186/1471-2105-12-323>
- Li Z, Adams RM, Chourey K, Hurst GB, Hettich RL, Pan C. Systematic Comparison of Label-Free, Metabolic Labeling, and Isobaric Chemical Labeling for Quantitative Proteomics on LTQ Orbitrap Velos. *J. Proteome Res.* 2012;11(3):1582-1590. Doi: <http://dx.doi.org/10.1021/pr200748h>
- Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2013;30(7):923-930. Doi: <http://dx.doi.org/10.1093/bioinformatics/btt656>
- Lin P-C, Hu W-C, Lee S-C, Chen Y-L, Lee C-Y, Chen Y-R, *et al.* Application of an Integrated Omics Approach for Identifying Host Proteins That Interact With *Odontoglossum* ringspot virus Capsid Protein. *Mol Plant Microbe In.* 2015;28(6):711-726. Doi: <http://dx.doi.org/10.1094/mpmi-08-14-0246-r>
- Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, *et al.* RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* 2012;40(W1):W622-W627. Doi: <http://dx.doi.org/10.1093/nar/gks540>
- Loman NJ, Quinlan AR. Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics.* 2014;30(23):3399-3401. Doi: <http://dx.doi.org/10.1093/bioinformatics/btu555>
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12). Doi: <http://dx.doi.org/10.1186/s13059-014-0550-8>
- Madroñero J, Rodrigues SP, Antunes TFS, Abreu PMV, Ventura JA, Fernandes AaR, *et al.* Transcriptome analysis provides insights into the delayed sticky disease symptoms in *Carica papaya*. *Plant cell reports.* 2018;37(7):967-980. Doi: <http://dx.doi.org/10.1007/s00299-018-2281-x>
- Megias E, Do Carmo LST, Nicolini C, Silva LP, Blawid R, Nagata T, *et al.* Chloroplast proteome of *Nicotiana glauca* infected by Tomato Blistering mosaic virus. *Protein J.* 2018;37(3):290-299. Doi: <http://dx.doi.org/10.1007/s10930-018-9775-9>
- Mehta D, Hirsch-Hoffmann M, Were M, Patrignani A, Zaidi SS-E-A, Were H, *et al.* A new full-length circular DNA sequencing method for viral-sized genomes reveals that RNAi transgenic plants provoke a shift in geminivirus populations in the field. *Nucleic Acids Res.* 2019;47(2):e9-e9. Doi: <http://dx.doi.org/10.1093/nar/gky914>
- Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, *et al.* Linking Virus Genomes with Host Taxonomy. *Viruses.* 2016;8(3):66. Doi: <http://dx.doi.org/10.3390/v8030066>
- Mochida K, Shinozaki K. Advances in Omics and Bioinformatics Tools for Systems Analyses of Plant Functions. *Plant Cell Physiol.* 2011;52(12):2017-2038. Doi: <http://dx.doi.org/10.1093/pcp/pcr153>
- Mosa KA, Ismail A, Helmy M. Omics and System Biology Approaches in Plant Stress Research. In: *Plant Stress Tolerance. SpringerBriefs in Systems Biology.* Ciudad: Springer, Cham, 2017. p. 21-34.
- Muñoz Baena L, Gutiérrez Sánchez PA, Marín Montoya M. Secuenciación del genoma completo del Potato yellow vein virus (PYV) en tomate (*Solanum lycopersicum*) en Colombia. *Acta biol. colomb.* 2017;22(1):5-17. Doi: <http://dx.doi.org/10.15446/abc.v22n1.59211>
- Muñoz E, Gutiérrez S, Marín M. Detection and genome characterization of Potato virus Y isolates infecting potato (*Solanum tuberosum* L.) in La Union (Antioquia, Colombia). *Agron. Colomb.* 2016;34(3):317-328. Doi: <http://dx.doi.org/http://dx.doi.org/10.15446/agron.colomb.v34n3.59973>
- Naito FYB, Melo FL, Fonseca MEN, Santos CaF, Chanes CR, Ribeiro BM, *et al.* Nanopore sequencing of a novel bipartite New World begomovirus infecting cowpea. *Arch. Virol.* 2019;164(7):1907-1910. Doi: <http://dx.doi.org/10.1007/s00705-019-04254-5>
- Nicaise VR. Crop immunity against viruses: outcomes and future challenges. *Front. Plant Sci.* 2014;5. Doi: <http://dx.doi.org/10.3389/fpls.2014.00660>
- NIH. The Cost of Sequencing a Human Genome [cited 2019 24/04/2019]. Available from: <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>.
- O’leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, Mcveigh R, *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016;44(D1):D733-D745. Doi: <http://dx.doi.org/10.1093/nar/gkv1189>
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods.* 2017;14(4):417-419. Doi: <http://dx.doi.org/10.1038/nmeth.4197>
- Patro R, Mount SM, Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat. Biotechnol.* 2014;32(5):462-464. Doi: <http://dx.doi.org/10.1038/nbt.2862>
- Pavan Kumar BK, Kanakala S, Malathi VG, Gopal P, Usha R. Transcriptomic and proteomic analysis of yellow mosaic diseased soybean. *J Plant Biochem Biot.* 2016;26(2):224-234. Doi: <http://dx.doi.org/10.1007/s13562-016-0385-3>
- Peck SC. Proteomics: Setting the Stage for Systems Biology. 2018;35:243-257. Doi: <http://dx.doi.org/10.1002/9781119312994.apr0379>

- Pecman A, Kutnjak D, Gutiérrez-Aguirre I, Adams I, Fox A, Boonham N, *et al.* Next Generation Sequencing for Detection and Discovery of Plant Viruses and Viroids: Comparison of Two Approaches. *Front Microbiol.* 2017;8. Doi: <http://dx.doi.org/10.3389/fmicb.2017.01998>
- Peng Y, Leung HCM, Yiu S-M, Lv M-J, Zhu X-G, Chin FYL. IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. *Bioinformatics.* 2013;29(13):i326-i334. Doi: <http://dx.doi.org/10.1093/bioinformatics/btt219>
- Proost S, Van bel M, Vanechoutte D, Van de peer Y, Inzé D, Mueller-Roeber B, *et al.* PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Res.* 2015;43(D1):D974-D981. Doi: <http://dx.doi.org/10.1093/nar/gku986>
- Rang FJ, Kloosterman WP, De Ridder J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* 2018;19(1):90. Doi: <http://dx.doi.org/10.1186/s13059-018-1462-9>
- Rhoads A, Au KF. PacBio Sequencing and Its Applications. *Genom Proteom Bioinf.* 2015;13(5):278-289. Doi: <http://dx.doi.org/10.1016/j.gpb.2015.08.002>
- Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, *et al.* De novo assembly and analysis of RNA-seq data. *Nat. Methods.* 2010;7(11):909-912. Doi: <http://dx.doi.org/10.1038/nmeth.1517>
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2009;26(1):139-140. Doi: <http://dx.doi.org/10.1093/bioinformatics/btp616>
- Rodríguez MH, Niño NE, Cutler J, Langer J, Casierra-Posada F, Miranda D, *et al.* Certificación de material vegetal sano en Colombia: Un análisis crítico de oportunidades y retos para controlar enfermedades ocasionadas por virus. *rev. colomb.cienc.hortic.* 2016;10(1):164-175. Doi: <http://dx.doi.org/10.17584/rcch.2016v10i1.4921>
- Roossinck MJ, Martin DP, Roumagnac P. Plant Virus Metagenomics: Advances in Virus Discovery. *Phytopathology.* 2015;105(6):716-727. Doi: <http://dx.doi.org/10.1094/phyto-12-14-0356-rww>
- Rose PW, Bi C, Bluhm WF, Christie CH, Dimitropoulos D, Dutta S, *et al.* The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Res.* 2012;41(D1):D475-D482. Doi: <http://dx.doi.org/10.1093/nar/gks1200>
- Röst HL, Sachsenberg T, Aiche S, Bielow C, Weisser H, Aicheler F, *et al.* OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat. Methods.* 2016;13(9):741-748. Doi: <http://dx.doi.org/10.1038/nmeth.3959>
- Roy A, Zhang Y. Protein Structure Prediction. 2012. Doi: <http://dx.doi.org/10.1002/9780470015902.a0003031.pub2>
- Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics.* 2012;28(8):1086-1092. Doi: <http://dx.doi.org/10.1093/bioinformatics/bts094>
- Shannon P. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-2504. Doi: <http://dx.doi.org/10.1101/gr.1239303>
- Soares EDA, Werth EG, Madroñero LJ, Ventura JA, Rodrigues SP, Hicks LM, *et al.* Label-free quantitative proteomic analysis of pre-flowering PMeV-infected *Carica papaya* L. J. *Proteom.* 2017;151:275-283. Doi: <http://dx.doi.org/10.1016/j.jprot.2016.06.025>
- Souza PFN, Garcia-Ruiz H, Carvalho FEL. What proteomics can reveal about plant-virus interactions? Photosynthesis-related proteins on the spotlight. *Theor Exp Plant Phys.* 2019;31(1):227-248. Doi: <http://dx.doi.org/10.1007/s40626-019-00142-0>
- Srivastava A, Sarkar H, Gupta N, Patro R. RapMap: a rapid, sensitive and accurate tool for mapping RNA-seq reads to transcriptomes. *Bioinformatics.* 2016;32(12):i192-i200. Doi: <http://dx.doi.org/10.1093/bioinformatics/btw277>
- Stagljari I. The power of OMICs. *Biochemical and biophysical research communications.* 2016;479(4):607-609. Doi: <http://dx.doi.org/10.1016/j.bbrc.2016.09.095>
- Stano M, Beke G, Klucar L. viruSITE—integrated database for viral genomics. *Database.* 2016;2016:baw162. Doi: <http://dx.doi.org/10.1093/database/baw162>
- Stare T, Stare K, Weckwerth W, Wienkoop S, Gruden K. Comparison between Proteome and Transcriptome Response in Potato (*Solanum tuberosum* L.) Leaves Following Potato Virus Y (PVY) Infection. *Proteomes.* 2017;5(4):14. Doi: <http://dx.doi.org/10.3390/proteomes5030014>
- Sun F, Fang P, Li J, Du L, Lan Y, Zhou T, *et al.* RNA-seq-based digital gene expression analysis reveals modification of host defense responses by rice stripe virus during disease symptom development in *Arabidopsis*. *Virol. J.* 2016;13(1):202. Doi: <http://dx.doi.org/10.1186/s12985-016-0663-7>
- Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, *et al.* Mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* 2004;37(6):914-939. Doi: <http://dx.doi.org/10.1111/j.1365-313X.2004.02016.x>
- Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology.* 2013;31:46-53. Doi: <http://dx.doi.org/10.1038/nbt.2450>
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 2012;7:562-578. Doi: <http://dx.doi.org/10.1038/nprot.2012.016>
- Tyanova S, Cox J. Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. 2018;1711:133-148. Doi: http://dx.doi.org/10.1007/978-1-4939-7493-1_7
- Tyanova S, Temu T, Carlson A, Sinitcyn P, Mann M, Cox J. Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics.* 2015;15(8):1453-1456. Doi: <http://dx.doi.org/10.1002/pmic.201400449>

- Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods.* 2016;13(9):731-740. Doi: <http://dx.doi.org/10.1038/nmeth.3901>
- Vallejo C D, Gutiérrez S PA, Marín M M. Genome characterization of a Potato virus S (PVS) variant from tuber sprouts of *Solanum phureja* Juz. et Buk. *Agron. Colomb.* 2016;34(1):51-60. Doi: <http://dx.doi.org/10.15446/agron.colomb.v34n1.53161>
- Varela ALN, Komatsu S, Wang X, Silva RGG, Souza PFN, Lobo AKM, *et al.* Gel-free/label-free proteomic, photosynthetic, and biochemical analysis of cowpea (*Vigna unguiculata* [L.] Walp.) resistance against Cowpea severe mosaic virus (CPSMV). *J. Proteom.* 2017;163:76-91. Doi: <http://dx.doi.org/10.1016/j.jprot.2017.05.003>
- Villamil-Garzón A, Cuellar WJ, Guzmán-Barney M. Natural coinfection of *Solanum tuberosum* crops by the Potato yellow vein virus and potyvirus in Colombia. *Agron. Colomb.* 2014;32(2):213-223. Doi: <http://dx.doi.org/10.15446/agron.colomb.v32n2.43968>
- Villamor DEV, Ho T, Al Rwahnih M, Martin RR, Tzanetakos IE. High Throughput Sequencing For Plant Virus Detection and Discovery. *Phytopathology.* 2019;109(5):716-725. Doi: <http://dx.doi.org/10.1094/phyto-07-18-0257-rww>
- Vowinckel J, Capuano F, Campbell K, Deery MJ, Lilley KS, Ralser M. The beauty of being (label)-free: sample preparation methods for SWATH-MS and next-generation targeted proteomics. *F1000.* 2014;2:272. Doi: <http://dx.doi.org/10.12688/f1000research.2-272.v2>
- Wan Y, Renner DW, Albert I, Szpara ML. VirAmp: a galaxy-based viral genome assembly pipeline. *GigaScience.* 2015;4(1). Doi: <http://dx.doi.org/10.1186/s13742-015-0060-y>
- Wilkins MR, Sanchez J-C, Gooley AA, Appel RD, Humphrey-Smith I, Hochstrasser DF, *et al.* Progress with Proteome Projects: Why all Proteins Expressed by a Genome Should be Identified and How To Do It. *Biotechnol Genet Eng.* 1996;13(1):19-50. Doi: <http://dx.doi.org/10.1080/02648725.1996.10647923>
- Wong S-M, Cho WK, Lian S, Kim S-M, Seo BY, Jung JK, *et al.* Time-Course RNA-Seq Analysis Reveals Transcriptional Changes in Rice Plants Triggered by Rice stripe virus Infection. *PLoS One.* 2015;10(8):e0136736. Doi: <http://dx.doi.org/10.1371/journal.pone.0136736>
- Wu C, Li X, Guo S, Wong S-M. Analyses of RNA-Seq and sRNA-Seq data reveal a complex network of anti-viral defense in TCV-infected *Arabidopsis thaliana*. *Sci. Rep.* 2016;6(1). Doi: <http://dx.doi.org/10.1038/srep36007>
- Wu L, Han Z, Wang S, Wang X, Sun A, Zu X, *et al.* Comparative proteomic analysis of the plant-virus interaction in resistant and susceptible ecotypes of maize infected with sugarcane mosaic virus. *J. Proteom.* 2013;89:124-140. Doi: <http://dx.doi.org/10.1016/j.jprot.2013.06.005>
- Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, *et al.* SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics.* 2014;30(12):1660-1666. Doi: <http://dx.doi.org/10.1093/bioinformatics/btu077>
- Xu K, D. Nagy P. Dissecting Virus-Plant Interactions Through Proteomics Approaches. *Curr Proteomics.* 2010;7(4):316-327. Doi: <http://dx.doi.org/10.2174/157016410793611792>
- Yamashita A, Sekizuka T, Kuroda M. VirusTAP: Viral Genome-Targeted Assembly Pipeline. *Front Microbiol.* 2016;7. Doi: <http://dx.doi.org/10.3389/fmicb.2016.00032>
- Zanardo LG, De Souza GB, Alves MS. Transcriptomics of plant-virus interactions: a review. *Theor Exp Plant Phys.* 2019;31(1):103-125. Doi: <http://dx.doi.org/10.1007/s40626-019-00143-z>
- Zaynab M, Fatima M, Abbas S, Sharif Y, Jamil K, Ashraf A, *et al.* Proteomics Approach Reveals Importance of Herbal Plants in Curing Diseases. *Mol. Microbiol.* 2018;1(1):23-28.