Next Generation Sequencing for virus discovery and large-scale virus screening of the cassava collection at CIAT's genebank


Monica Carvajal-Yepes[1], Maritza Cuervo[2], Cristian Olaya[1+], Angélica Martínez[2], Ivan Lozano[1], Diana Niño[2], Ericson Aranzales[2] Daniel Debouck[2], Peter Wenzl[2], Wilmer Cuellar[1] and Joe Tohme [3]

[1] Virology Laboratory, [2]Genetic Resource Program, [3]Biotechnology research Unit; International Center for Tropical Agriculture (CIAT), AA 6713, Cali, Colombia. [+] Current address: Department of Plant Pathology, Washington State University, USA

m.carvajal@cgiar.org m.cuervo@cgiar.org

BACKGROUND

Analyzing the virome of plants using next generation sequencing (NGS) has accelerated the identification of viruses associated with disease that could not be isolated otherwise. Deep sequencing of small interfering RNAs (siRNAs) is used to diagnose and identify viruses in different crops showing disease symptoms, such as cassava, bean, rice and papaya. By using this approach, novel viruses infecting cassava fields in Colombia, were discovered and reported from plants with Cassava Frogskin disease (CFSD) symptoms (Carvajal-Yepes *et al*. 2014). CFSD is a quarantine disease that causes significant yield losses and is associated to infection of multiple pathogens (Alvarez *et al*., 2009; Calvert *et al*., 2008; Carvajal-Yepes *et al*., 2014). The Genetic Resource Program (GRP) at the International Center for Tropical Agriculture (CIAT) conserves the world's largest and most diverse collection of cassava (*Manihot esculenta*). The health certification of the cassava collection material is done by the Germplasm Health Laboratory (GHL), ensuring the safe distribution of these materials. Following discoveries in cassava with NGS, in addition to screening other viruses such as: *Cassava virus X* (CsXV), *Cassava common mosaic virus* (CsCMV) and *Cassava vein mosaic virus* (CVMV), it was necessary to implement and standardize methods for the detection of Cassava torrado-like virus (CsTLV), Cassava polero-like virus (CsPLV), Cassava frogskin-associated virus (CsFSaV) and Cassava new alphaflexivirus (CsNAV) in the large-scale screening workflow. Up to date, 61% of the *in vitro* cassava collection has been evaluated. The use of NGS for discovering important quarantinable pathogens will contribute to the safe distribution of germplasm around the world.

VIRUS DISCOVERY BY USING NGS

By exploiting the plant immune response system, it is posible to detect viruses infecting plants. During virus replication, host endoribonucleases cleave the viral-derived dsRNAs limiting viral replication. Cleaved small virus-derived RNAs (svRNAs) are used for NGS.

RNA isolation and library preparation

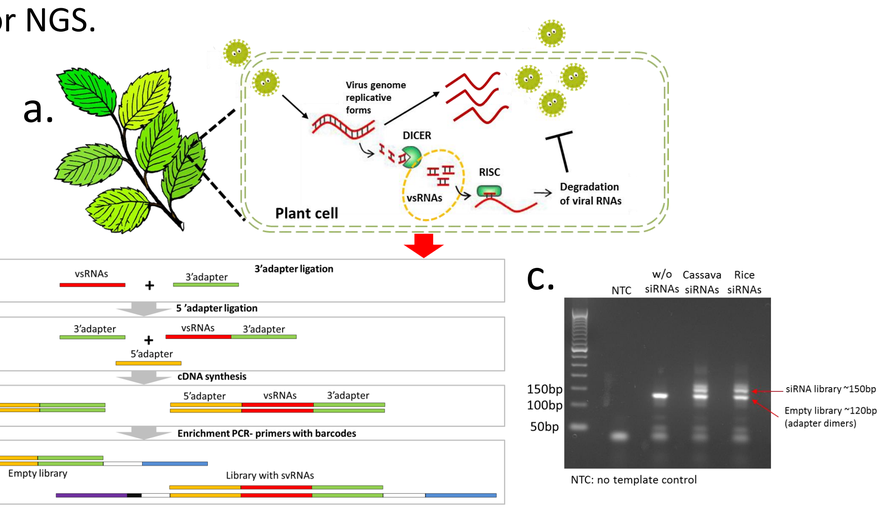


Fig. 1 a) Source of RNAs for library preparation (yellow dashed circle) b) Representation of library preparation c) Library confirmation/Enrichment PCR. (siRNAs: Small interfering RNAs); svRNAs: small virus-derived RNAs)

Sequencing

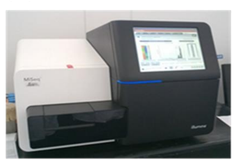
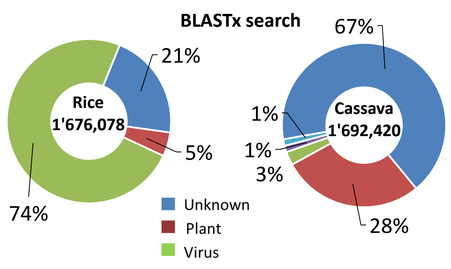


Fig. 2. Deep sequencing of Small interfering RNAs (siRNAs) using the Miseq platform from Illumina (a) and MiSeq Reagent Kit v2 for 50 cycles. Sequencing produces around 15 M reads of 21-24 nt per run (b).

Data analysis

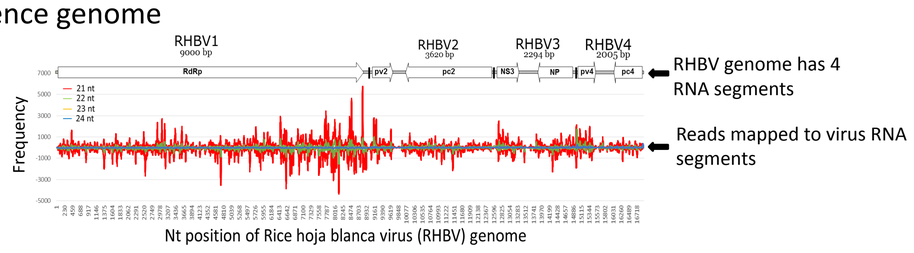
i) *De novo* assembly of 21-24nt reads with VELVET v1.2.03 (Zerbino and Birney, 2008) produces longer assembled sequences (nodes or contigs). BLASTn and BLASTx search are done against the NCBI database using assembled nodes to identify viral sequences.



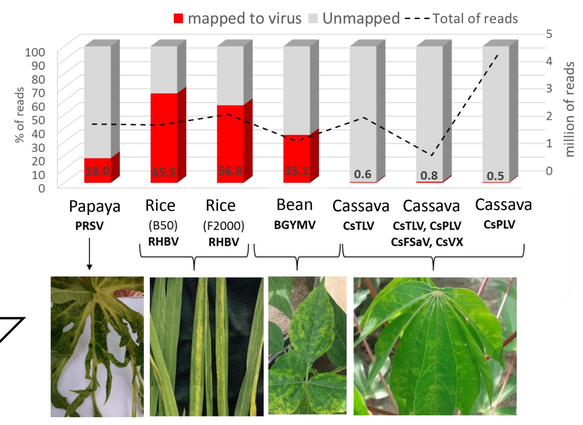
De novo assembly allowed the identification of three novel viruses infecting cassava in Colombia in plants with CFSD (Carvajal-Yepes *et al*., 2014).

Virus name	Classification	GenBank
Cassava torrado-like virus (CsTLV)	*Secoviridae / Torradovirus*	KC 505250, KC 505151
Casava polero-like virus (CsPLV)	*Luteoviridae / Polerovirus*	KC 505249
Cassava new alphaflexivurs (CsNAV)	*Alphaflexiviridae / Potexvirus*	KC 505252

ii) Reads can also be mapped to known viruses (used as reference genomes) by using MAQ 0.7.1 software. This allows multiple detection of viruses simultaneously. **Below**: *Rice hoja blanca virus* (RHBV) derived reads mapped to a RHBV reference genome



Number of reads and percentage of reads mapped to virus references in different plant species

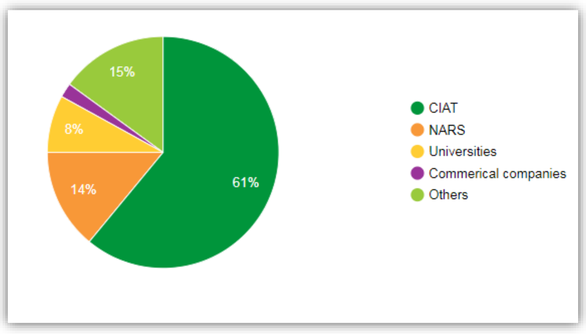


Samples with disease symptoms showing the percentage of reads (siRNAs) that were mapped to known viral sequences (shown in red color). Viral sequences used: PRSV: *Papaya ring spot virus*, RHBV: *Rice hoja blanca virus*, BGYMV: *Bean golden yellow mosaic virus*, Cassava viruses (mentioned above). The (%) of unmapped reads are shown in grey. The number of reads that were obtained per sample are shown by the dashed black line.

LARGE-SCALE VIRUS SCREENING IN THE CASSAVA COLLECTION

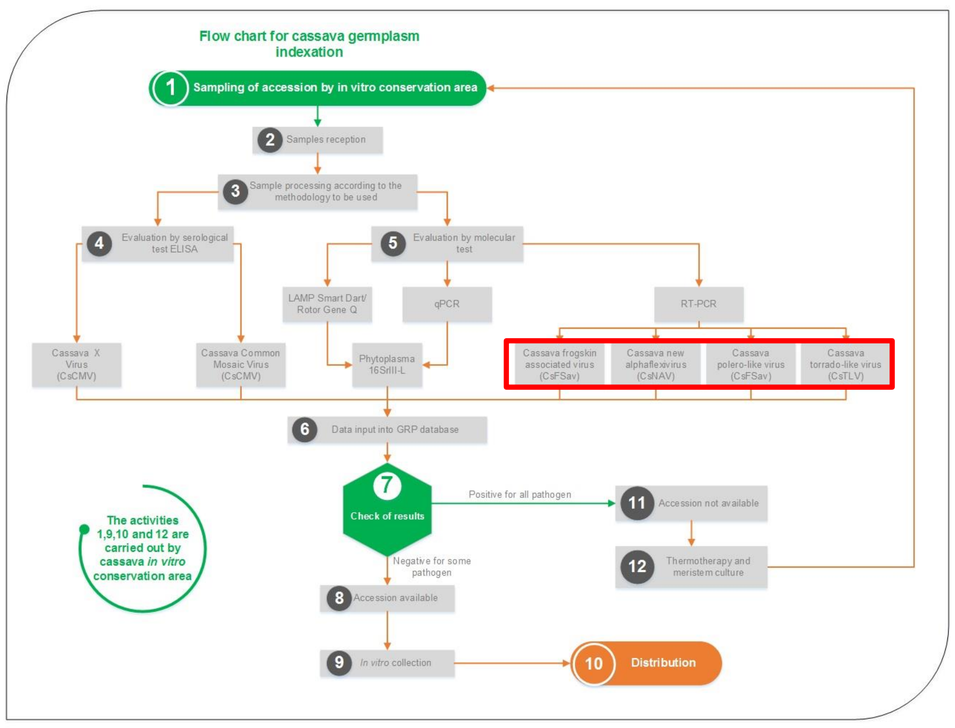
The *in vitro* collection of the genus *Manihot* of the GRP at CIAT is currently represented by 6,643 materials; which have been registered in the Multilateral System of Access and Benefit sharing of the International Treaty on Plant Genetic Resources for Food and Agriculture. These materials must be free of quarantine diseases for distribution to users worldwide.

Cassava *in vitro* collection at CIAT



Cassava germplasm distributed 6,492 accessions since 1979 to 84 countries

Following the discovery of cassava viruses by NGS associated to the quarantine disease CFSD, the large-scale virus screening in the cassava collection updated the number of viruses to be tested and the procedures in the GHL operations workflow (shown within the red rectangle).



Up to date, 61% (4087 accessions) of the *in vitro* cassava collection has been evaluated. The remaining 39% are in this process.

PERSPECTIVES:

- Implementation of NGS for seed borne pathogen detection is highly desirable in genebank collections, especially for collections that need to be regenerated in the field, being exposed to emerging pathogens during regeneration crop seasons.
- NGS is a reliable and sensitive diagnostic option for the evaluation of germplasm with special country phytosanitary requirements, particularly when certain pathogen detection methods are not yet validated in the facility or positive controls are not available.
- The establishment of DNA banks will increase the possibility of optimizing operations for DNA pathogen detection by using NGS.

REFERENCES

- **Alvarez E; Mejía JF; Llano GA; Loke JB; Calari A; Duduk B and Bertaccini A.** 2009. Characterization of a Phytoplasma Associated with Frogskin Disease in Cassava. Plant Disease 93:11, 1139-1145
- **Calvert L; Cuervo M; Lozano I; Villareal N; Arroyave J**. 2008. Identification of three strains of a virus associated with cassava plants affected by frogskin disease. Journal of Phytopathology 156:647-653.
- **Carvajal-Yepes M; Olaya C; Lozano I; Cuervo M; Castaño M; Cuellar WJ**. 2014. Unraveling complex viral infections in cassava (*Manihot esculenta* Crantz) from Colombia. Virus Research 186:76–86.
- **Zerbino, D. R. and Birney, E**. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research*. 18, 821-829.