BIONFORMATIC ANALYSIS OF NGS LIBRARIES: A MODERN TOOL FOR DISCOVERY AND CHARACTERIZATION OF VIRAL DISEASES IN ANCIENT FRUIT CROP

Chiumenti M.[1]*, Morelli M.[1], Giampetruzzi A.[1], Saldarelli P.[1], Minafra A.[1].
1. *CNR Istituto per la Protezione Sostenibile delle Piante (IPSP), UOS Bari, 70126 Bari, Italy.*
**CORRESPONDING AUTHOR: michela.chiumenti@ipsp.cnr.it*

INTRODUCTION

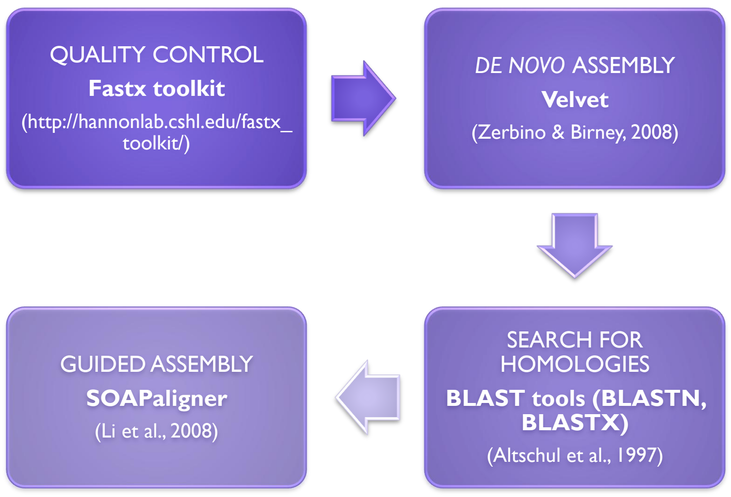
Routinely used bio- and molecular assays for the detection of systemic plant pathogens always hide an *a priori* choice for target, influenced by the evidence of symptoms, diagnostic purposes prone to certification standard enquiries, literature available, etc. NGS analysis, being free from such constraints, widens the target range, and includes also cultivated and wild crops, never before characterized for their sanitary status, in the effort to define the whole 'virome'.

The use of NGS approach allowed the identification and the complete genome characterization of several plant virus species, previously unknown (Persimmon cryptic virus, PeCV; Mulberry badnavirus-1, MBV-1), or poor characterized and never associated to a specific symptomatology and/or host (Apple dimple fruit viroid, ADFVd; Persimmon rhabdovirus A, PeVA; Apple green crinkle associated virus, AGCaV). Significantly these findings arose from crops, respectively fig (Ficus carica) for ADFVd, Japanese persimmon (Diospyros kaki) in the case of PeCV and PeVA, quince (Cydonia oblonga) for AGCV and mulberry (Morus alba) for MBV-1, unconsidered in traditional quest for viral agents and diseases.

MATERIALS AND METHODS

Plant sources. Symptomatic leaves from persimmon, mulberry, fig, and quince were collected between late spring and early autumn from symptomatic plants and used to extract and purify nucleic acids.

Libraries preparation and analysis. Double stranded RNAs purified according to Valverde et al. (1990) and small RNAs isolated using mirVana™ miRNA Isolation Kit (Ambion, Life technologies, USA) from leaf tissues of each accession, were used to construct cDNA libraries according to the Illumina protocol. A 50 base-single read run was done on a HiScan™ SQ apparatus.



RESULTS

PERSIMMON

Symptoms → dsRNA

Library data
- tot reads: 33,632,035
- De novo assembly: 1,829 contigs (kmer 23)

BLAST
- **Persimmon cryptic virus (PeCV)**
- Persimmon rhabdovirus A (PeVA)

NEW!

MULBERRY

Symptoms → sRNA

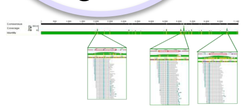
Library data
- tot reads: 4,321,760
- De novo assembly: 1,977 contigs (kmer 17)

BLAST
- Badnavirus contigs

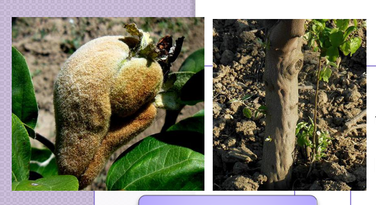
Primer design
- **Mulberry badnvirus 1**

NEW!

Alignment



QUINCE



Symptoms → sRNA

Library data
tot reads: 13,155,692
De novo assembly: 10,013 contigs (kmer 17)

BLAST
- Apple green crinkle-associated virus (AGCaV)

SOAP Alingnment
- final consensus revealed a 100% coverage and an overall nt identity of 92 %.

FIG

Symptoms → sRNA

Library data
- tot reads: 8,567,810
- De novo assembly: 1,655 contigs (kmer 17)

BLAST
- Apple dimple fruit viroid (ADFVd)
- - 310 nt
- - 85–86% identity with the ADFVd reference

CONCLUSIONS

To develop this work, the same bioinformatics tools were used to identify and reconstruct viral and viroid genomes. Despite this, the bioinformatics pipeline adopted in each case was every time adapted to the type of output obtained by *de novo* assembly and sequence homologies with viruses already described. This is due to two factors: first of all, the type of library used (dsRNAs or sRNAs), then the kind of infectious entity found. Being produced by viral enriched fraction randomly cleaved, dsRNAs libraries can give more genomic information about the viruses contained in the sample examined, on the other hand, only viruses producing "long-time" stable dsRNAs during their replication can be better studied with this approach. Instead sRNAs are more informative for what concern plant-pathogens interaction, but, for their biogenesis, they can hardly produce extended contigs to be used for virus genome reconstruction. Otherwise, the kind of infectious entity is important too: for example bioinformatics identification of new viroids can be done only by BLASTN search for homologies in short (few nucleotides) highly conserved regions, rather than viruses that possess wider conserved domains and genomic organization in the same genus, that allows also the use of additional tools, like BLASTX and tBLASTX, for protein identification.

REFERENCES

1. Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J., 1997. Nucleic Acids Research **25: 3389-3402.**
2. Dovas C.I., Katis N.I., 2003. Journal of Virological Methods **107: 99-106**
3. Ito T., Suzaki K., Nakano, M., 2013. Journal of General Virology **94: 1917-1921.**
4. Li R., Li Y., Kristiansen K., Wang J., 2008. Bioinformatics **24(5): 713–714.**
5. Saldarelli P., Rowhani A., Routh G., Minafra A., Digiaro M., 1998. European Journal of Plant Pathology **104(9): 945-950.**
6. Valverde R.A., Nameth S.T. and Jordan R.L., 1990. Plant Disease **74(3): 255–258.**
7. Zerbino D.R., Birney E., 2008. Genome Research **18(5): 821-829.**