Effect of template enrichment for RNA-Seq library preparation: a case study of multiple viral infection in red clover

Igor Koloniuk*, Jana Fránová, Tatiana Sarkisova, Jaroslava Přibylová, Karel Petrzik, Ondřej Lenz, Josef Špak

Biology Centre of the Czech Academy of Sciences, Institute of Plant Molecular Biology, Plant Virology Department
Branišovská 31, 370 05 České Budějovice, Czech Republic, * koloniuk@umbr.cas.cz

The plant of red clover (*Trifolium pratense* L.) HZ2 hosted viruses with all kinds of RNA genomes following the Baltimore classification (see Table A 'Virus information'). The analyzed plant had symptoms of severe dwarfism, yellow and necrotic spots on leaf blades, and leaf malformation (Fig. 1). It was selected as a suitable model to compare four different strategies of RNA template enrichment prior to preparation of high throughput sequencing library: dsRNA without modification, dsRNA processed by rRNA depletion, total RNA processed by rRNA depletion, and total RNA with polyA enrichment.

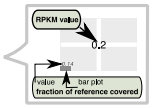
Table A — Virus information; Species established / novel

Suggested taxonomy	Genome	Polyadenylated	Genomic organization (5 kb)	Species established / novel
	+ ssRNA	YES		Red clover RNA virus 1
Luteoviridae	+ ssRNA	NO		Chickpea chlorotic stunt virus
				Red clover enation virus 1
				Soybean dwarf virus
Partitiviridae	dsRNA	NO		Red clover cryptic virus 1
Rhabdoviridae	- ssRNA	Genes		Red clover varicosavirus 1
Satellite	+ ssRNA	NO		Red clover virus-associated RNA
Secoviridae	+ ssRNA	YES		Red clover torradovirus 1
Umbravirus	+ ssRNA	NO		Red clover enation virus 2

Table B — RPKM (Reads per kilo base per million) values for viral contigs by kind of template

Species established / novel	dsRNA	Total RNA > polyA: dsRNA > RiboZero	Total RNA > polyA: Total RNA > RiboZero	Total RNA > RiboZero
Red clover RNA virus 1	5.0	9.2	9.5	0.2
Chickpea chlorotic stunt virus	24.1	1391.6	0.7	145.2
Red clover enation virus 1	647.5	12449.5	10.0	137.0
Soybean dwarf virus	12.5	463.8	0.4	18.2
Red clover cryptic virus 1	6.6	20.7	1.0	2.1
Red clover varicosavirus 1	5.1	57.5	16.1	320.6
Red clover virus-associated RNA	40.6	627.6	2.9	69.1
Red clover torradovirus 1	243.4	2408.1	712.1	123.8
Red clover enation virus 2	44.2	2387.8	1.2	107.3

Fig. 1 Red clover HZ2 plant showing symptoms of dwarfism and irregular yellow stripping



Materials and Methods

The plant was kindly provided by 'Dr. Hana Jakešová Clovers and Grass Plant Breeding' station and originated from Hladké Životice, Czech Republic. Total RNA was extracted from 100 mg of a fresh leave tissue using MagJET Plant RNA Kit (Thermo Scientific, Lithuania). MagJET mRNA enrichment kit was applied for polyA RNA selection. Double-stranded RNA was extracted from 300 mg of fresh plant material as described previously (Morris, T. J., & Dodds, J. A. (1979). Phytopathology, 69(8)) and followed with RQ1 RNase-Free DNase (Promega, Madison, WI, USA) and S1 nuclease treatments according to the manufacturer's recommendations. Ribosomal depletion was performed from 5 μg of a total RNA using Ribo-Zero rRNA Removal Kit (Plant) (Illumina, USA). All RNA preparations were quantified using Qubit HS RNA assay (Invitrogen, USA) and checked with SYBR Gold (Invitrogen, USA) stained agarose gel electrophoresis. Sequencing libraries were prepared from fragmented ds cDNA templates (Maxima H Minus Double-Stranded cDNA Synthesis Kit, Thermo Scientific, Lithuania) following MuSeek Illumina compatible Library Preparation Kit, (Thermo Scientific, Lithuania). The libraries were quantified with NEBNext Library Quant Kit for Illumina (New England BioLabs, Ipswich, MA, USA) and then processed on either HiSeq 2500 (dsRNA) or HiSeq4000 (dsRNA+RiboZero, totalRNA+polyA, totalRNA +RiboZero) in 100 b SE output mode (SEQme s.r.o., Czech Republic).

Random sampling of 15 million short reads was done from each original datasets with CLC Genomic Workbench 9.5.3 (Qiagen, Denmark). The reads were mapped (minimum fraction 0.8, minimum identity 0.95) onto viral references (*de novo* assembled and verified with Sanger sequencing). The mapping were used to estimate values of Reads per kilo base per million mapped reads (RPKM, i.e. how many reads per a million were matched to a 1 kb long region of a viral reference), and fraction of each covered reference (Table B).

Conclusions

Δ all viral targets were detected in each case using mapping approach

Δ low abundant viruses (Red clover RNA virus 1 and Red clover cryptic virus 1) were detected with both polyA selection and rRNA depletion of the total RNA. The numbers and scarsed positions of the reads on their mappings (not shown) indicate that mapping (indexing/ targeted virus detection) might be the only possible way of their detection. Alternatively, the sequencing depth (total number of reads per sample) should be as much as necessary to provide enough reads for *de novo* assembly of the viral contigs and following homology-based virus detection.

Δ rRNA depletion improved total yield of virus-specific reads, especially when applied for dsRNA enrichmed material

Δ polyA selection has not remove all viral RNAs lacking polyadenylated ends

Δ considering time and resources factors, the plain dsRNA enrichment is an optimal choice for the presented case

Acknowledgement This work was performed with institutional support RVO: 60077344, project COST LD15035 "Implementation of NGS in plant virology – unprecedented capacity with a risk of data overflow" from the Ministry of Education, Youth and Sports of the Czech Republic and supported by the COST Action FA1407 (DIVAS), COST (European Cooperation in Science and Technology). We are kindly indebted to Alena Matyášová for excellent technical assistance.