

Assessing the genetic stability of mumps virus preparations for animal oncolytic virotherapy through RNA-seq

Jelena Ivančić Jelečki^{✉,1}, Tanja Košutić Gulija¹, Dragan Jurić², Maja Jagušić¹, Dubravko Forčić¹, Borna Jelečki³, Beata Halassy¹

✉ jivancic@unizg.hr

¹University of Zagreb, Centre for Research and Knowledge Transfer in Biotechnology, Zagreb, Croatia

²Croatian Institute of Public Health, Zagreb, Croatia

³University of Zagreb, Faculty of Science, Department of Physics, Zagreb, Croatia

Introduction

The mumps virus (MV) is an RNA virus from the *Paramyxoviridae* family. Its only natural host is man, in which it causes a systemic infection that primarily manifests as a respiratory disease with painful swelling of the parotid glands. More severe neurological or reproductive system inflammatory complications, particularly in males, may also occur. MV vaccines based on attenuated viral strains have been in use for decades, significantly reducing the incidence of mumps disease.

MV has been documented as a potent oncolytic agent in several preclinical and early clinical studies in humans, but it has a potential to be used as oncolytic agent in veterinary medicine as well: despite strict natural host specificity, MV can successfully replicate in various animal cells and tissues, including canine and feline primary tumour cells and tissue cultures.

MV genome is non-segmented negative-sense single RNA, 15,384 nt long. Particles with defective viral genomes (DVGs) are also present in MV populations; copy-back DVGs may even impact virulence.

Our oncolytic MV preparations are clarified supernatants of infected Vero cells, and we aim to assess the reproducibility of their production. Among other characteristics, we monitor the consistency of viral genetic profile (consensus sequence, intra-population variability and DVG content). In this study, we present the RNA-seq data of four independently produced MV preparation batches (MV1-MV4).

Wet-lab protocol

methods shown in grey were assessed and have been omitted from the final protocol

1. Sample pretreatment

- filtration (0.45 nm)
- centrifugation (10,000 xg, 10 min)
- DENARASE treatment (4.5 h, 60 U/mL, 37 °C, 300 rpm)

2. RNA isolation

- viral RNA extraction kit (chromatography based)
- on-column DNase treatment
- off-column DNase treatment

3. rRNA depletion

- NEBNext rRNA Depletion kit

4. RNA sample QC

- fluorometric RNA concentration determination
- RT-PCR (for 28S rRNA and MV SH gene RNA)

5. Directional RNA library preparation

- NEBNext Ultra II Directional RNA Library Prep
- generation of two libraries for each RNA sample (technical replicates)
- 15 cycles of PCR enrichment

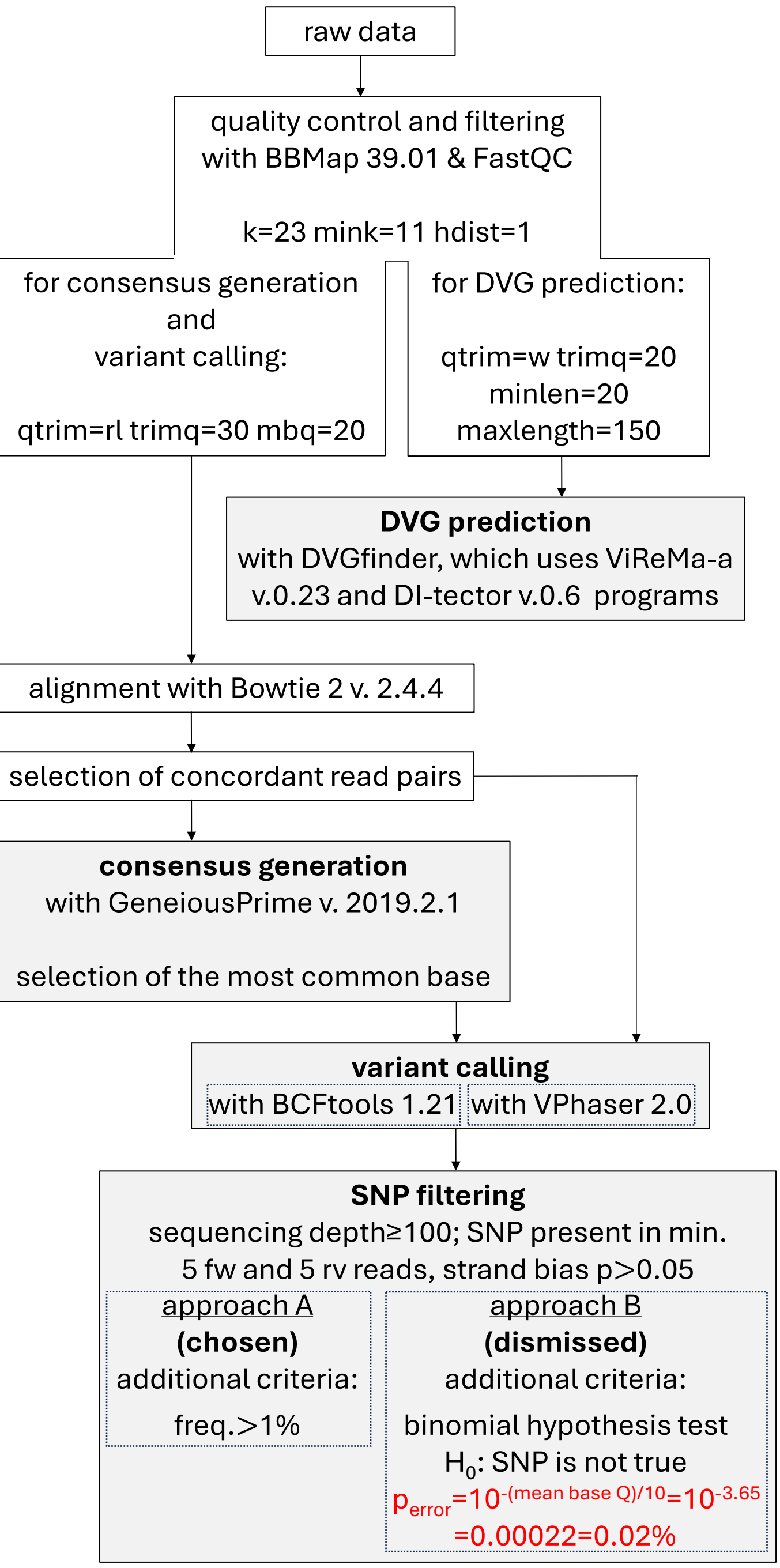
6. Library QC

- Bioanalyzer electrophoresis
- fluorometric DNA concentration determination
- RT-PCR (for 28S rRNA and MV SH gene RNA)

7. Sequencing

- Illumina MiniSeq, 2x150 cycles

Bioinformatic pipeline



Acknowledgements: The research was supported by the European Union - NextGenerationEU mechanism (grant NPOO.C3.2.R3-11.04.0124: "Development of oncolytic virotherapy for the treatment of cancer in animals") and OP Competitiveness and Cohesion (project CERVirVac). The views and opinions expressed are solely those of the authors and do not necessarily reflect the official views of the European Union or the European Commission.



RESULTS

Sequencing statistics and genome coverage distribution

Table 1. Sequencing statistics

sample*	no. of reads after QC	percentage of reads aligned to MV reference	percentage of MV genome length coverage**	mean sequencing depth	percentage of reads originating from the negative sense (genomic) RNA
VarA WS	2,600,982	86.68	99.95	5,735	67.90
MV1 d0	3,242,036	89.21	99.95	7,447	65.14
MV2 d0	2,772,184	91.65	99.96	13,664	65.16
MV3 d0	3,056,154	92.73	99.95	14,742	66.92
MV4 d0	2,698,690	92.38	99.94	9,936	63.91
MV2 d14	2,644,070	92.38	99.93	6,339	66.48
MV2 d14S	2,581,548	93.84	99.94	6,240	65.97
MV3 d14	2,896,478	88.25	99.94	6,582	67.69
MV3 d14S	3,328,026	90.76	99.93	11,842	62.93
LZg113+1pV	1,807,166	86.19	99.95	8,190	80.00

*each sample (except LZg113+1pV) was sequenced in duplicate. Similar results were obtained for both replicates, the data in this table is shown only for one of them

**percentage of positions with sequencing depth ≥ 100×

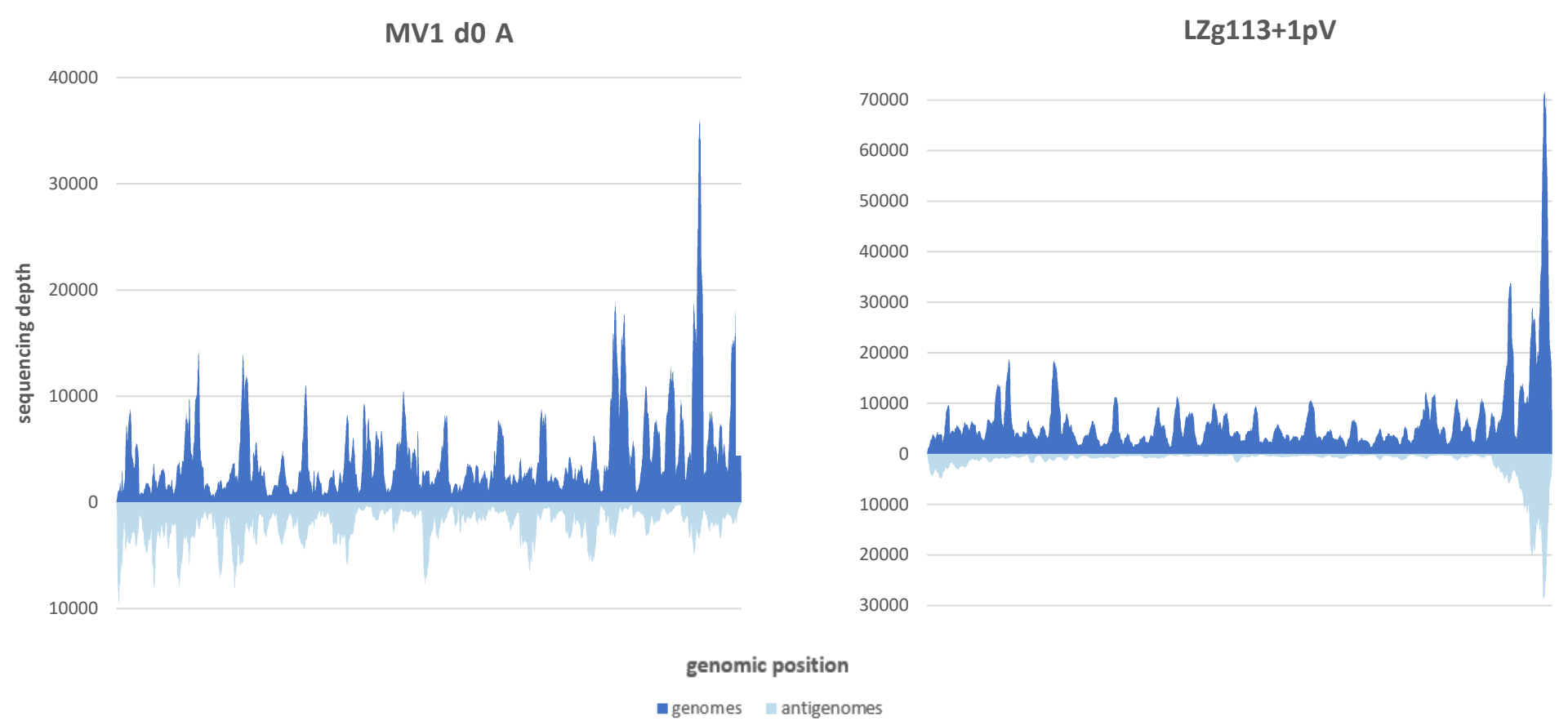


Figure 1. Genome coverage distribution. Coverage of MV1 d0 A is chosen as the representative sample for all OVPs.

R1. Identical consensus sequences and low genomic variability were obtained for all OVPs

Table 2. Intra-population variability, based on VPhaser 2 unfiltered data

sample	π ($\times 10^{-4}$)
VarA WS	1.93
MV1 d0	2.89
MV2 d0	6.94
MV3 d0	7.89
MV4 d0	5.59
MV2 d14	1.87
MV2 d14S	2.07
MV3 d14	2.33
MV3 d14S	4.61
LZg113+1pV	11.86

π , mean nucleotide diversity of two replicates (except for LZg113+1pV)

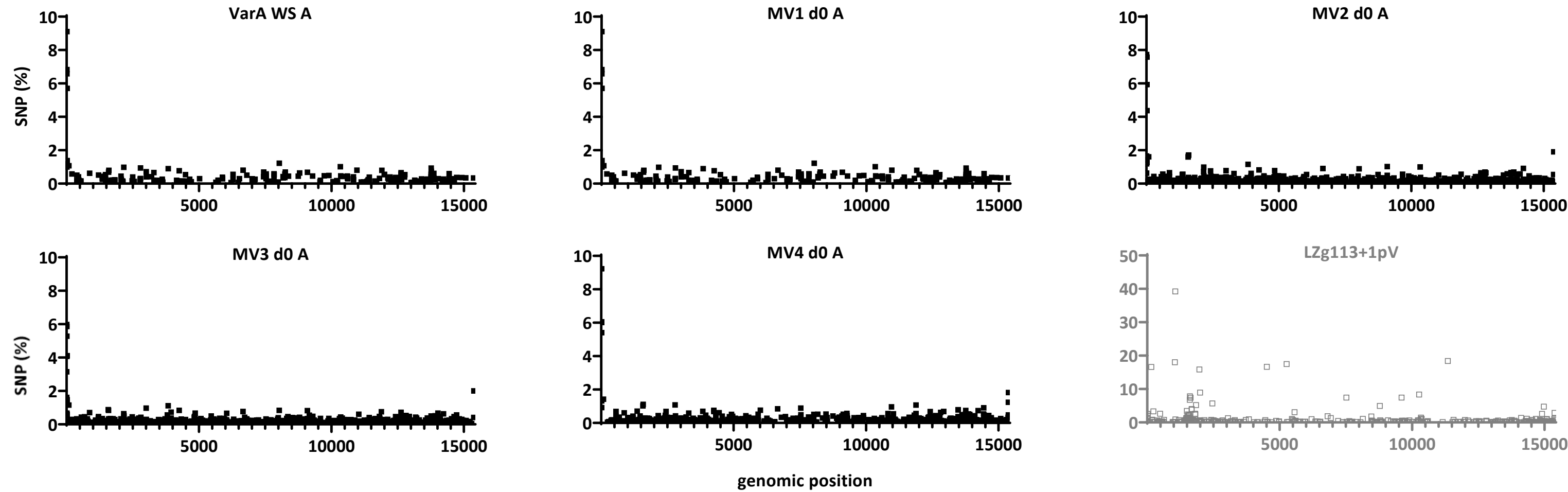


Figure 2. Genomic positions and frequencies of all SNPs inferred by VPhaser 2. Results are shown for only one replicate of each sample. Only SNPs at positions with sequencing depth ≥ 100 are shown.

R2. True SNPs within OVP populations were indistinguishable from technical artifacts

Table 3. Number of SNPs inferred with the two software packages, before and after additional filtering.

sample	no. of SNPs in A B*			
	before filtering		after filtering (additional criteria A)	
	BCFtools	VPhaser 2 (common; % of common)	BCFtools (common)	VPhaser 2 (common)
VarA WS	25,241 21,039	285 216 (59; 23.55)	3 5	0
MV1 d0	26,814 28,195	501 500 (161; 32.17)	6 1	1
MV2 d0	34,455 34,314	1,860 1,957 (714; 37.41)	13 11	7
MV3 d0	35,406 39,586	2,099 2,484 (754; 32.90)	11 9	5
MV4 d0	30,052 28,113	1,212 1,346 (471; 36.83)	7 10	7
MV2 d14	27,054 26,132	280 289 (80; 28.12)	2 2	0
MV2 d14S	27,049 24,891	248 220 (72; 30.77)	1 6	0
MV3 d14	26,695 27,094	376 371 (107; 28.65)	1 2	1
MV3 d14S	32,963 31,219	1,382 1,330 (552; 40.71)	8 5	4

A, B: technical replicates (the two libraries originating from the same RNA)

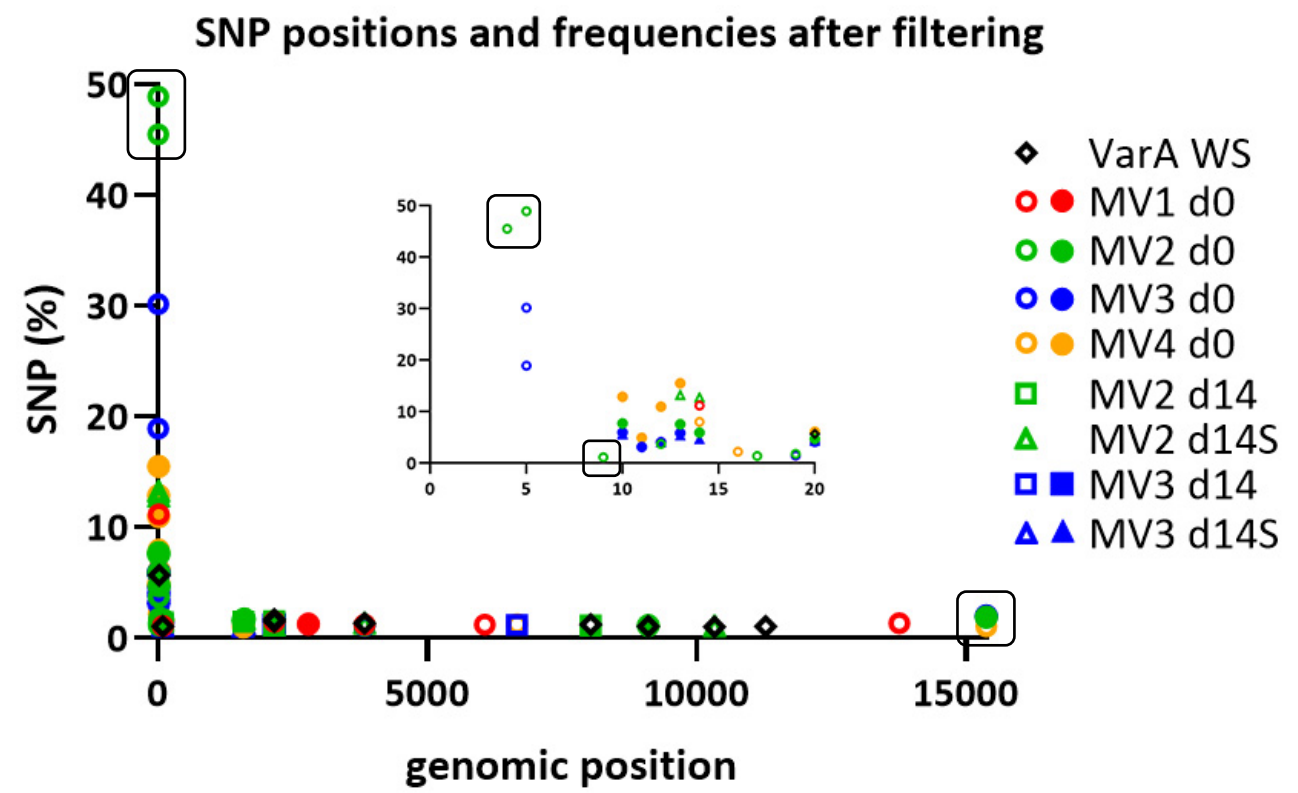


Figure 3. Genomic positions and frequencies of SNPs after filtering. Open and filled symbols represent SNPs detected in only one replicate or in both NGS replicates, respectively. Black rectangles emphasize 8 SNPs detected only with BCFtools. The inset graph displays the same data, only the x-axis range is limited to 20.

Table 4. After SNP filtering, 29 out of 15384 MV genomic positions were flagged as potential true SNPs. All except 1575 and 1599 were in (or next to) GC-rich regions or homopolymeric regions.

SNP position	genomic sequence surrounding SNP
4, 5, 9-14, 16, 17, 19, 20	1-ACCAAGGGGAAAAAGAAATGGG-23
86	83-CGAGGCCGGGCTCG-96
1575	1571-CCAATCAA-1579
1589, 1599	1586-CTTTCATGCCATATAT-1603
2159	2155-CCCCACCGC-2163
2790	2780-GGCGACAG-2800
3836	3828-GCGCGCCCAT-3837
6064	6059-GCTGCTGGGC-6068
6664	6662-CCTGGGCC-6669
8030	8025-GCCCCACTGC-8034
9094	9090-CCCCAG-9095
10331	10330-CTCCGG-10335
11274	11266-GGGGGGCCA-1275
11882	11880-GGTCCC-11885
13760	13757-CCCCCCC-13763
15373, 15375	15371-TTTCCTCCC-15379

R3. Unlike in sample LZg113+1pV, only low amounts of copy-back DVGs were identified in the OVP NGS datasets

Table 5. 5' copy-back DVGs in sample LZg113+1pV supported with the highest read counts. None of them were detected in NGS datasets of OVP samples.

LZg113+1pV	DVG length	breaking point	reinitiation point	read counts	
				ViReMa-a	DI-tector
	648	15290	14832	14657	14775
	648	14831	15291	7603	13041
	360	15287	15123	6689	7201
	360	15121	15289	5844	8725
	828*	15280	14662	3712	3769
	828	14662	15280	2143	3455
	1176*	15182	15512	881	920

*detected by RT-PCR and confirmed by Sanger sequencing

Table 6. 5' copy-back DVGs common for all OVPs. The last row is included for comparison purposes, as an indication that DVG read counts in OVPs were generally low (few RT-PCR amplicons were visible on agarose gels, but they were too weak for Sanger sequencing). The four DVGs listed in this table were also found in the NGS dataset of LZg113+1pV, in which they were supported with similarly low read counts.

all OVPs	DVG length	breaking point	reinitiation point	the highest read count*	
				ViReMa-a	DI-tector
	1469	14626	14675	15	23
	1470	14629	14671	14	16
	1440	14653	14677	17	22
	1365	14689	14716	23	29
	any	not applicable		31	82

*these four DVGs were predicted in all 18 datasets. Only the highest count number in a set is shown.