

Supplemental Material

Domain Organization and Evolution of the Highly Divergent 5' Coding Region of Genomes of Arteriviruses, Including the Novel Possum Nidovirus

Anastasia Gulyaeva^{a*}, Magdalena Dunowska^{b*#}, Erik Hoogendoorn^{a\$}, Julia Giles^b, Dmitry Samborskiy^c, Alexander E. Gorbalenya^{a,c,d#}

^a Department of Medical Microbiology, Leiden University Medical Center, Leiden, the Netherlands

^b Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand

^c Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia

^d Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia

* Joint first authors

\$ Present address: Innovatic, Delft, the Netherlands

To whom correspondence should be addressed. Dr. Alexander E. Gorbalenya, Tel: +31-71-5261652; Fax: +31-71-5266761; Email: a.e.gorbalenya@lumc.nl; Dr. Magdalena Dunowska, Tel: +64 (6) 356 9099 ext 7571; Fax: +64 (6) 350 5714; Email: M.Dunowska@massey.ac.nz

Supplementary tables

Table S1. Primer and probe sequences used to sequence the 5'-terminal region of WPDV genome.

Name	Sequence (5' to 3')	Used as	Position (nt)	Round
WPD.S5.R	TGGAGGTGGCGCGTAGGTGT	primer	3,028-3,047	RACE 1
Biotin-WPD.S5.F	Biotin-ATGCAGCTTATGTCCTTGATGGGGT	probe	2,893-2,917	RACE 1
WPD.S7.R	CAGGGCATGTGCGCGGTAGT	primer	2,510-2,530	RACE 1
WPD.S8.R	GCCCACGGTTGCTTCAAAAAGTCT	primer	2,062-2,086	RACE 1
WPD.S10.R	CCCACTCCAGTGCCTTTGTCAT	primer	1,288-1,309	RACE 2
WPD.S13.R	AGGCGCTGCAGTACCGTCGT	primer	1,096-1,115	RACE 2
WPD.S14.R	GATGAACGGCATCCCTGACA	primer	1,003-1,022	RACE 2
Biotin-WPD.S12.F	Biotin-CGGGGCGATCGTGGCTTACAG	probe	887-907	RACE 2
WPD.S15.R	CGTCTCCGGGTATCATGGTC	primer	869-888	RACE 3
WPD.S16.R	AAAATCGGGTGGACGGATGT	primer	545-564	RACE 3
WPD.S18.R	TTGTCGAATCGGGGGTAAGC	primer	150-169	RACE 3

Table S2. Protein domains that are conserved in arteriviruses and were used for phylogeny reconstruction.

Domain ¹	Coordinates in NC_001961.1 genome (nt) ²	
	from	to
nsp3	4,927	5,616
nsp4	5,617	6,228
nsp5	6,229	6,738
nsp7a	6,787	7,233
nsp8-9 ³	7,564	9,617
nsp10_HELcore	10,002	10,775
nsp11	10,941	11,609

¹Domains conserved in all arteriviruses.

²Coordinates of conserved domains in NC_001961.1 genome of PRRSV-2, used to delineate domains in polyprotein MSA of selected arteriviruses (see Figure 3A).

³Translation involves -1 PRF.

Table S3. Lengths (aa) of arteriviral nsp1-2 protein domains.

Virus ¹	Domain ²								
	ZnF	PLP1a	'Nuclease'	PLP1b	PLP1c	Hinge	PLP2	HVR	TM1-CR
PRRSV-2	33	147	69	134	0	45	140	650	331
PRRSV-1	33	147	74	131	0	35	137	557	332
LDV	33	148	65	135	0	0	135	447	324
APRAV	36	146	66	142	0	11	142	436	340
KRCV-2	28	138	61	124	126	107	134	139	329
PBJV	28	137	62	124	128	111	139	127	333
SHFV	28	136	62	124	134	149	136	135	332
DeMAV	28	139	62	123	133	100	137	157	331
KRTGV	28	138	62	123	133	81	137	155	331
KRCV-1	28	136	62	122	131	124	134	133	329
SHEV	28	136	61	123	138	140	133	150	326
MYBV-1	28	137	62	124	134	74	135	164	331
EAV	49	90	0	121	0	0	130	125	317
WPDV	0	98	0	163	142	57	152	143	341

¹One virus-representative from each of the fourteen arterivirus species, delineated by DEmARC, was analysed.

²Domains were delineated based on similarity with domains and cleavage sites of arteriviruses studied experimentally.

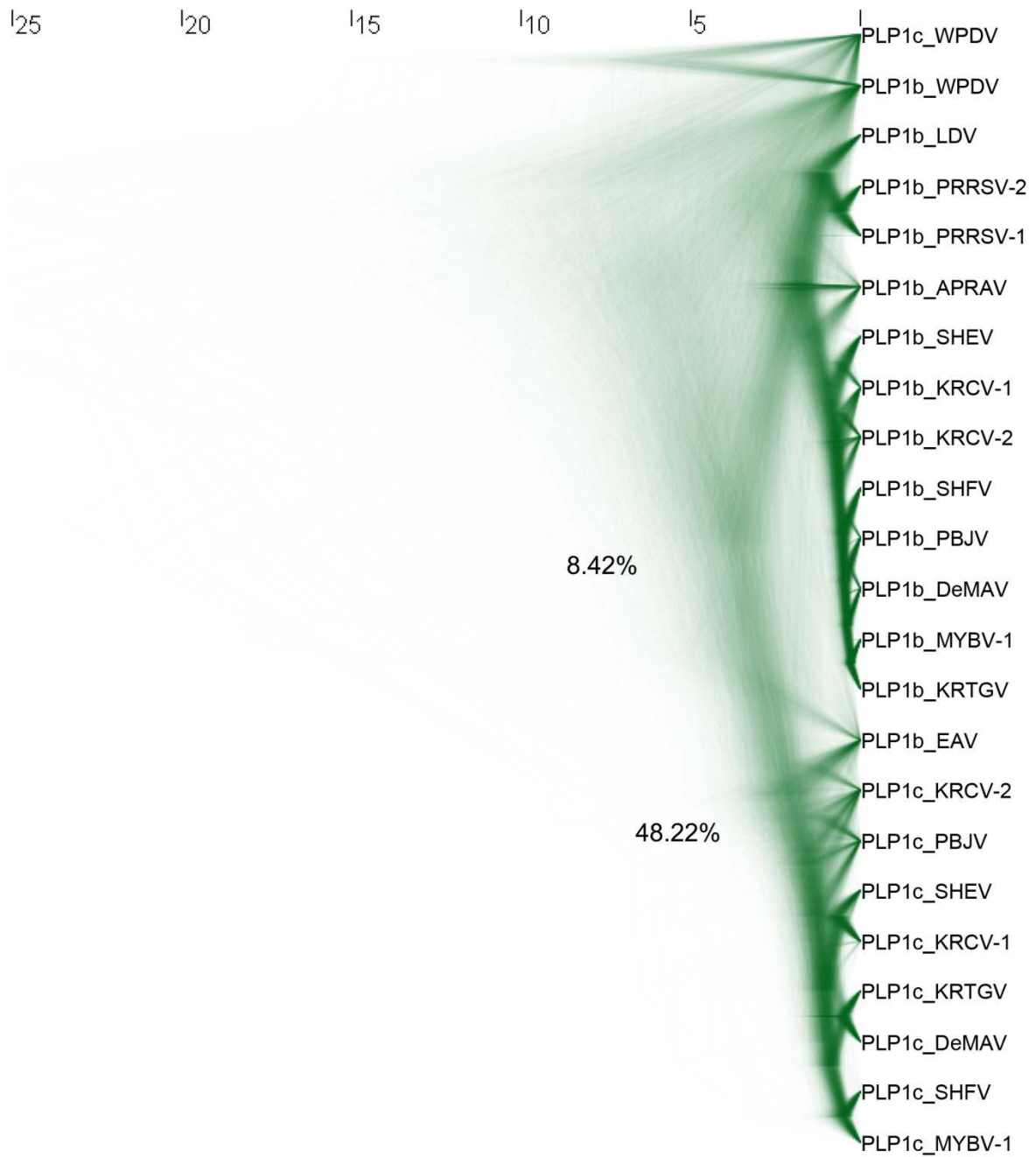


Figure S1. Phylogeny of PLP1b and PLP1c of arteriviruses. Shown is a posterior sample of phylogenetic trees generated by BEAST using pan-arterivirus MSA of PLP1b and PLP1c. Percentages of trees in the sample, in which EAV PLP1b is basal to either non-WPDV PLP1c or non-WPDV PLP1bc clades are indicated near the MRCA of the corresponding clades. For other designations, see Figure 3A legend.