Comparative genomics of nidoviruses: towards understanding the biology and evolution of the largest RNA viruses

Gulyaeva, A.

LIST OF ABBREVIATIONS

(-)ssRNA	negative-sense single-stranded RNA
(+)ssRNA	positive-sense single-stranded RNA
2'-PDE	2',5'-phosphodiesterase
3CLpro (3CL^{pro})	3C-like protease
aa	amino acid
AAbV	aplysia abyssovirus 1
AIC	Akaike information criterion
AMP, ADP, ATP	adenosine mono-, di-, triphosphate
ANK	ankyrin domain
APRAV	African pouched rat arterivirus
AsD	arterivirus-specific domain
BIC	Bayesian information criterion
BNV1	Beihai nido-like virus 1
BPNV	ball python nidovirus
BRV	Breda virus
BSA	bovine serum albumin
CAVV	Cavally virus
CIP	calf intestine alkaline phosphatase
CMP, CDP, CTP	cytidine mono-, di-, triphosphate
CoV	coronavirus
CPD	cyclic phosphodiesterase
CPE	cytopathic effect
CPU	central processing unit
CR domain	cysteine-rich domain
DdCoV	duck-dominant coronavirus
DEmARC	DivErsity pArtitioning by hieRarchical Clustering

DeMAV	De Brazza's monkey arterivirus
DNA	deoxyribonucleic acid
dsRNA	double-stranded RNA
E	nidovirus envelope protein
EAV	equine arteritis virus
EM	electron microscopy
ER	endoplasmic reticulum
EToV	equine torovirus
EVD	extreme value distribution
ExoN	DEDDh subfamily exoribonuclease
FN2	fibronectin type II domain
FSBG	5'-(4-fluorosulfonylbenzoyl)guanosine
GAV	gill-associated virus
GMP, GDP, GTP	guanosine mono-, di-, triphosphate
GTase	guanylyltransferase
HE	hemagglutinin-esterase
HEL1	superfamily 1 helicase
HGT	horizontal gene transfer
HMM	hidden Markov model
HVR	hypervariable region
IBV	infectious bronchitis virus
ICTV	International Committee on Taxonomy of Viruses
InfV	influenza virus
ISH	*in situ* hybridization
kb	kilobase
KRCV	Kibale red colobus virus
KRTGV	Kibale red-tailed guenon virus

LAMPA	LArge Multidomain Protein Annotator
LDV	lactate dehydrogenase-elevating virus
LGT	lateral gene transfer
M	nidovirus matrix protein
MAR	mono-ADP-ribose
MCMC	Markov chain Monte Carlo
MERS	Middle East respiratory syndrome
MHV	mouse hepatitis virus
ML	maximum likelihood
MMP-2	matrix metalloproteinase-2
Mpro (M^{pro})	main protease
MRCA	most recent common ancestor
mRNA	messenger RNA
MSA	multiple sequence alignment
MTase	methyltransferase
N	nidovirus nucleocapsid protein
n.a.	not applicable
n.d.	not done
NAD	nicotinamide adenine dinucleotide
NDiV	Nam Dinh virus
NendoU	uridylate-specific endonuclease
NGS	next generation sequencing
NiRAN	nidovirus RdRp-associated nucleotidyltransferase
NMP, NDP, NTP	nucleoside mono-, di-, triphosphate
N-MT	SAM-dependent N7-methyltransferase
nsp	non-structural protein
nt	nucleotide

O-MT	SAM-dependent 2'-O-methyltransferase
ORF	open reading frame
p.i.	post infection
p.t.	post transfection
PAR	poly-ADP-ribose
PBJV	Pebjah virus
PCBP	poly(C) binding protein
PDB	Protein Data Bank
Pkinase	protein kinase
PLP	papain-like protease
polyA	polyadenylate
pp	polyprotein
PPD	pairwise patristic distance
PRF	programmed ribosomal frameshifting
PRRSV	porcine reproductive and respiratory syndrome virus
PSCNV	planarian secretory cell nidovirus
PSSM	position-specific scoring matrix
PV	poliovirus
RdRp	RNA-dependent RNA polymerase
RHD	Rel homology domain
(RLM) RACE	(RNA ligase-mediated) rapid amplification of cDNA ends
RMSD	root mean square deviation
RNA	ribonucleic acid
RNase T2	ribonuclease T2
RNP	RNA-protein
RsD	ronivirus-specific domain
RTC	replication-transcription complex

RTPase	RNA 5'-triphosphotase
S	nidovirus spike protein
SAM	S-adenosyl methionine
SARS	severe acute respiratory syndrome
SD	standard deviation
sg	subgenomic
SH3 domain	Src homology 3 domain
SHEV	simian hemorrhagic encephalitis virus
SHFV	simian hemorrhagic fever virus
SI	standard inoculum
SPase	signal peptidase
SPR	subtree pruning and regrafting
SUD	"SARS-unique" domain
TAP	tobacco acid pyrophosphatase
TGEV	transmissible gastroenteritis virus
TM	transmembrane
tRNA	transfer RNA
TRS	transcription-regulating sequence
Ub	ubiquitin
UMP, UDP, UTP	uridine mono-, di-, triphosphate
UTR	untranslated region
WBV	white bream virus
WJHAV	Wuhan Japanese halfbeak arterivirus
WPDV	wobbly possum disease virus
wt	wild-type
ZBD	zinc-binding domain
ZnF	zinc finger

CURRICULUM VITAE

Anastasia Gulyaeva was born on November 6, 1991 in Moscow, Russia (USSR at the time). In June 2009 she graduated from the physico-mathematical lyceum № 1580 in Moscow. In September 2009 Anastasia enrolled in studies at the Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia. In the course of her studies she conducted rotation projects in the research groups of Prof. dr. V.I. Muronetz, Prof. dr. A.A. Mironov and Prof. dr. A.V. Alexeevsky. In July 2012 Anastasia participated in the MoBiLe Bioinformatics Summer School, where she was working on a scientific assignment in the research group of Prof. dr. P.A.C. 't Hoen in the Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands. In 2014 Anastasia graduated from the Lomonosov Moscow State University after defending her MSc. thesis dedicated to the usage of sequence weights in the hierarchical classification of viral genomes, and supervised by Dr. A.M. Leontovich, Dr. I.A. Sidorov and Prof. dr. A.E. Gorbalenya. In the same year, she started her doctoral research in the Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands under supervision of Dr. I.A. Sidorov and Prof. dr. A.E. Gorbalenya. Her doctoral research resulted in the present thesis entitled “Comparative genomics of nidoviruses: towards understanding the biology and evolution of the largest RNA viruses”.

LIST OF PUBLICATIONS

Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, **Gulyaeva AA**, Haagmans BL, Lauber C, Leontovich AM, Neuman BW, Penzar D, Perlman S, Poon LLM, Samborskiy DV, Sidorov IA, Sola I, Ziebuhr J: The species *Severe acute respiratory syndrome-related coronavirus*: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 2020, 5:536–544.

Gulyaeva AA, Sigorskih AI[#], Ocheredko ES[#], Samborskiy DV, Gorbalenya AE: LAMPA, LArge Multidomain Protein Annotator, and its application to RNA virus polyproteins. *Bioinformatics* 2020.

Kanitz M, Blanck S, Heine A, **Gulyaeva AA**, Gorbalenya AE, Ziebuhr J, Diederich WE: Structural basis for catalysis and substrate specificity of a 3C-like cysteine protease from a mosquito mesonivirus. *Virology* 2019, 533:21-33.

Nijhuis RHT[#], Sidorov IA[#], Chung PK, Wessels E, **Gulyaeva AA**, de Vries JJ, Claas ECJ, Gorbalenya AE: PCR assays for detection of human astroviruses: In silico evaluation and design, and in vitro application to samples collected from patients in the Netherlands. *J Clin Virol* 2018, 108:83-89.

Bukhari K, Mulley G, **Gulyaeva AA**, Zhao L, Shu G, Jiang J, Neuman BW: Description and initial characterization of metatranscriptomic nidovirus-like genomes from the proposed new family Abyssoviridae, and from a sister group to the *Coronavirinae*, the proposed genus Alphaletovirus. *Virology* 2018, 524:160-171.

Saberi A[#], **Gulyaeva AA**[#], Brubacher JL, Newmark PA, Gorbalenya AE: A planarian nidovirus expands the limits of RNA genome size. *PLoS Pathog* 2018, 14(11):e1007314.

Gulyaeva AA[#], Dunowska M[#], Hoogendoorn E, Giles J, Samborskiy D, Gorbalenya AE: Domain Organization and Evolution of the Highly Divergent 5' Coding Region of Genomes of Arteriviruses, Including the Novel Possum Nidovirus. *J Virol* 2017, 91(6).

Lehmann KC, **Gulyaeva AA**, Zevenhoven-Dobbe JC, Janssen GM, Ruben M, Overkleeft HS, van Veelen PA, Samborskiy DV, Kravchenko AA, Leontovich AM *et al*: Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses. *Nucleic Acids Res* 2015, 43(17):8416-8434.

Lehmann KC, Hooghiemstra L, **Gulyaeva AA**, Samborskiy DV, Zevenhoven-Dobbe JC, Snijder EJ, Gorbalenya AE, Posthuma CC: Arterivirus nsp12 versus the coronavirus nsp16

2'-O-methyltransferase: comparison of the C-terminal cleavage products of two nidovirus pp1ab polyproteins. *J Gen Virol* 2015, 96(9):2643-2655.

[#]equal contribution

ICTV proposals

Gorbalenya AE, Brinton MA, de Groot RJ, **Gulyaeva AA**, Lauber C, Neuman BW, Ziebuhr J: Pending ICTV taxonomic proposal 2019.023S Create five new families and a new suborder of vertebrate viruses in the order *Nidovirales*. 2019.

Brinton MA, **Gulyaeva AA**, Balasuriya UBR, Dunowska M, Faaberg KS, Goldberg T, Leung F-C, Nauwynck HJ, Snijder EJ, Stadejek T *et al*: Pending ICTV taxonomic proposal 2019.020S Create one new genus (*Nuarterivirus*); move the existing subgenus *Pedartevirus* to the genus *Iotaarterivirus*; rename one species from the subgenus *Pedartevirus*; create one new species in the new genus *Nuarterivirus*; create one new subgenus and two new species in the existing genus *Betaarterivirus*. 2019.

Ziebuhr J, Baker S, Baric RS, de Groot RJ, Drosten C, **Gulyaeva AA**, Haagmans BL, Neuman BW, Perlman S, Poon LLM *et al*: Pending ICTV taxonomic proposal 2019.021S Create ten new species and a new genus in the subfamily *Orthocoronavirinae* of the family *Coronaviridae* and five new species and a new genus in the subfamily *Serpentovirinae* of the family *Tobaniviridae*. 2019.

Gorbalenya AE, **Gulyaeva AA**, Hobson-Peters J, Junglen S, Morita K, Sawabe K, Vasilakis N, Ziebuhr J: Pending ICTV taxonomic proposal 2019.022S Create one new species in the genus *Alphamesonivirus* of the family *Mesoniviridae* and one new species in the genus *Okavirus* of the family *Roniviridae*. 2019.

Gorbalenya AE, Brinton MA, Cowley J, de Groot R, **Gulyaeva AA**, Lauber C, Neuman B, Ziebuhr J: ICTV taxonomic proposal 2017.015S Reorganization and expansion of the order *Nidovirales* at the family and sub-order ranks. 2017.

Brinton MA, **Gulyaeva AA**, Balasuriya UBR, Dunowska M, Faaberg KS, Goldberg T, Leung FC-C, Nauwynck HJ, Snijder EJ, Stadejek T *et al*: ICTV taxonomic proposal 2017.012S Expansion of the rank structure of the family *Arteriviridae* and renaming its taxa. 2017.

Ziebuhr J, Baric RS, Baker S, de Groot RJ, Drosten C, **Gulyaeva AA**, Haagmans BL, Neuman BW, Perlman S, Poon LLM *et al*: ICTV taxonomic proposal 2017.013S Reorganization of the family *Coronaviridae* into two families, *Coronaviridae* (including the current subfamily *Coronavirinae* and the new subfamily *Letovirinae*) and the new family *Tobaniviridae*

(accommodating the current subfamily *Torovirinae* and three other subfamilies), revision of the genus rank structure and introduction of a new subgenus rank. 2017.

Gorbalenya AE, Brinton MA, Cowley J, de Groot R, **Gulyaeva AA**, Lauber C, Neuman B, Ziebuhr J: ICTV taxonomic proposal 2017.014S Establishing taxa at the ranks of subfamily, genus, sub-genus and species in six families of invertebrate nidoviruses. 2017.

Brinton MA, **Gulyaeva AA**, Balasuriya UBR, Dunowska M, Faaberg KS, Leung FC, Nauwynck HJ, Snijder EJ, Stadejek T, Gorbalenya AE: ICTV taxonomic proposal 2015.014a-cS In the family *Arteriviridae* create 10 species (1 unassigned, 9 in the genus *Arterivirus*) and rename one species. 2015.

Ziebuhr J, Baric RS, Baker S, de Groot RJ, Drosten C, **Gulyaeva AA**, Haagmans BL, Lauber C, Neuman BW, Perlman S *et al*: ICTV taxonomic proposal 2015.003a-eS Create 12 species in the family *Coronaviridae*. 2015.

Gorbalenya AE, **Gulyaeva AA**, Hobson-Peters J, Junglen S, Morita K, Sawabe K, Vasilakis N, Ziebuhr J: ICTV taxonomic proposal 2015.004a,bS In the family *Mesoniviridae*, create four species in genus *Alphamesonivirus* and two unassigned in the family. 2015.

ACKNOWLEDGEMENTS

I would like to thank all the people who supported me on my PhD journey. First of all, I want to express my gratitude to my promotor Sasha Gorbalenya and co-promotor Igor Sidorov. Sasha, thank you for your guidance, advice, sharing your knowledge and ideas with me, and for all the opportunities that you gave me. Igor, thank you for always being there for me when I had difficulties. I am very grateful to my Moscow colleague Dmitry Samborskiy for finding amazing solutions to the most difficult bioinformatics problems. I would like to thank all my co-authors for pleasant and productive collaborations. Over the past five years, I have been fortunate to supervise several students of the MoBiLe Bioinformatics Summer School: Vanya Kuznetsov, Sveta Iarovenko, Andrey Sigorskih, Lena Ocheredko and Dima Penzar, thank you for your excellent work. I loved this experience and learned a lot from it! I would like to thank my former teachers and mentors from the Lomonosov Moscow State University, my supervisors from the MoBiLe Bioinformatics Summer School, and an MSc. thesis advisor Andrey Mikhailovich Leontovich, all of whom inspired me to pursue career in bioinformatics. I am also very grateful to Louis Kroes and Eric Snijder for the support of my work at the Department of Medical Microbiology, and to all my colleagues for wonderful and stimulating research environment. I would like to thank Ien Dobbelaar, Ineke van Ballegooijen-Molijn, Manon Stijnman, Sophie Greve, Esther Quakkelaar, Annemieke Hofman-Jansen and Marianne Parlevliet-de Gelder for helping me with administrative issues, and Hans van der Geest for maintaining computational environment for my work over these years. I am very grateful to Jeroen Corver and Tim Dalebout for translating thesis summary into Dutch. Finally, from the bottom of my heart, I would like to thank my family, and most importantly, my mum. Without your love, encouragement and support, my PhD journey would be impossible.