



# Rotavirus diversity, evolution, and lineage classification

**Dr Celeste Donato, PhD, MASM, BBiomedSchons**

- 1. Centre for Pathogen Genomics, University of Melbourne, Melbourne, Australia**
- 2. Department of Microbiology and Immunology, Peter Doherty Institute for Infection & Immunity, University of Melbourne, Melbourne, Australia**
- 3. Enteric Diseases Group, Murdoch Children's Research Institute, Melbourne, Australia**
- 4. Department of Paediatrics, University of Melbourne, Melbourne, Australia**
- 5. Department of Microbiology, Monash University, Melbourne, Australia**

# *Rotavirus species*

- Nine species recognised by ICTV (as of 2022)

1. *Rotavirus A*

2. *Rotavirus B*

3. *Rotavirus C*

4. *Rotavirus D*

○ ~~*Rotavirus E*~~

5. *Rotavirus F*

6. *Rotavirus G*

7. *Rotavirus H*

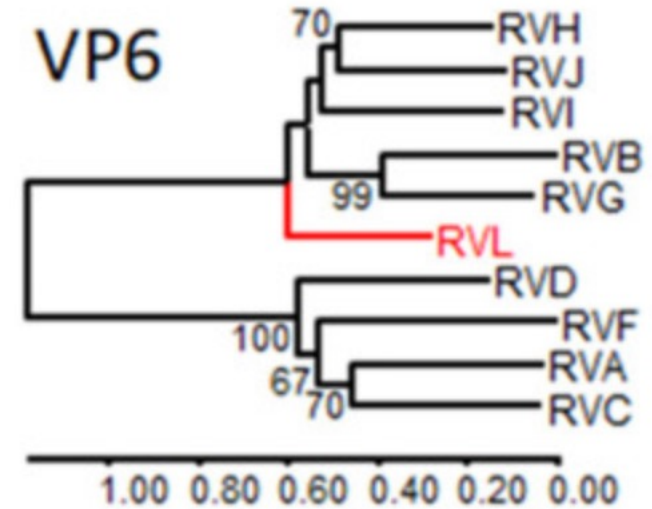
8. *Rotavirus I*

9. *Rotavirus J*

○ *Rotavirus K\*\**

○ *Rotavirus L\*\**

○ *Unclassified*



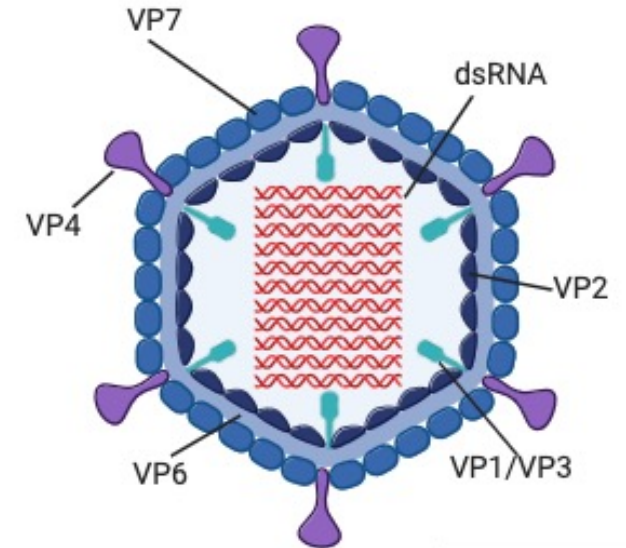
Neighbor-joining phylogenetic tree of the VP6 gene of rotavirus species. Scaled amino acid substitutions per site.

Johne R, et al. *Viruses*. 2022 24;14(3):462.

# Rotavirus genetic diversity

- **11 segments of dsRNA encoding 11-12 proteins**
- **Genome is approximately 18,555 base pairs**
- **Recombination is rare**
- **Reassortment is common – not between different rotavirus species**
- **Nomenclature for strain names**

**RV group/species of origin/country/common name/DOC/G- and P-type**



# Group B classification

- **1787 gene sequences in GenBank**
- **Whole genome classification proposed in 2018 by Douglas Marthaler & colleagues**
- **>80% of open reading frame of each gene**
- **No recent updates**
- **250 sequences added to GenBank since this scheme was proposed**
- **No online classification tool**

Gene segment	Gene	Genotypes	Nucleotide Cutoff
VP7	G	26	80%
VP4	P	5	80%
VP6	I	13	81%
VP1	R	5	78%
VP2	C	5	79%
VP3	M	5	77%
NSP1	A	8	76%
NSP2	N	10	83%
NSP3	T	6	78%
NSP4	E	4	76%
NSP5	H	7	79%

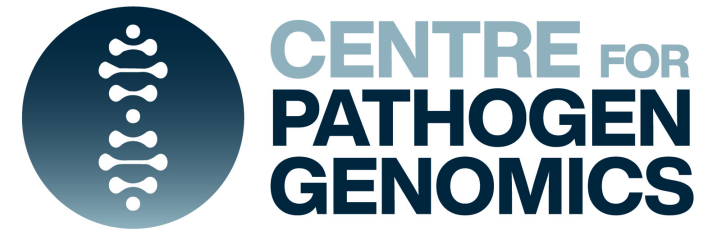


# Group C

- **5139 gene sequences in GenBank**
- **Whole genome classification proposed in 2017 by Tohru Suzuki & Ayako Hasebe**
- **>80% of open reading frame of each gene**
- **No recent updates**
- **2260 sequences added to GenBank since this scheme was proposed**
- **No online classification tool**

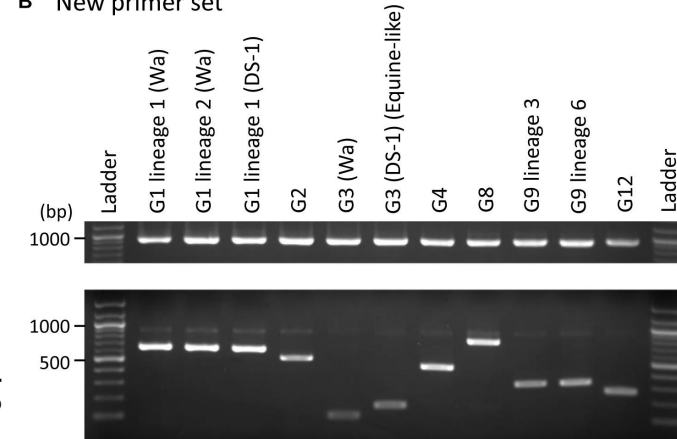
Gene segment	Gene	Genotypes	Nucleotide Cutoff
VP7	G	18	85%
VP4	P	21	85%
VP6	I	13	87%
VP1	R	4	74%
VP2	C	6	85%
VP3	M	6	85%
NSP1	A	9	84%
NSP2	N	8	87%
NSP3	T	6	85%
NSP4	E	5	81%
NSP5	H	4	80%

# Group A – historical classification



- **Classified into subgroups (SG) based on the antigenic specificity of the VP6 protein (SG I, II, I+II, non-I or non-II) using MAbs**
- **In 1989 G (glycoprotein, VP7) and P (protease sensitive, VP4) serotyping developed**
  - **14 G and 14 P serotypes**
- **In 2008 hemi-nested multiplex RT-PCR became routine for G and P genotyping**
  - **G types (1, 2, 3, 4, 8, 9, and 12)**
  - **P types ([4], [6], [8], [9], [10] and [11])**
- **Used by the WHO Global Rotavirus Surveillance Network and many national surveillance programs – sequencing not routine**
- **RT-qPCR assays exist but not routinely used**

B New primer set



# Group A – current classification

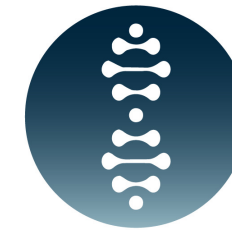
- **Rotavirus Classification Working Group – established in 2008**
- **26 members of the RCWG**
- **New sequences are submitted for classification – manually analyzed and the RCWG votes on agreement of new genotypes**
- **500bp minimum or 50% of ORF**
- **Issues**
  - **RotaC offline for 3+ years and unlikely to be made available again**
  - **Some new genotype sequences not in GenBank**



Jelle Matthijssens,  
KU Leuven

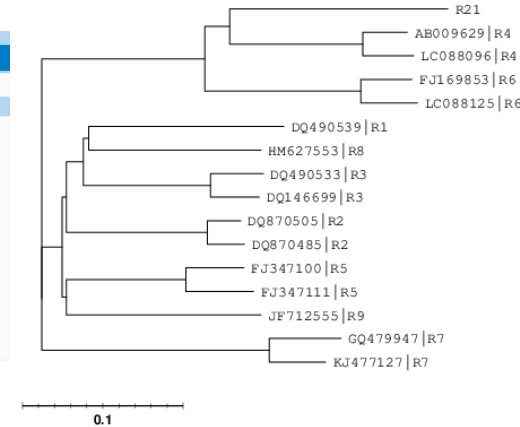
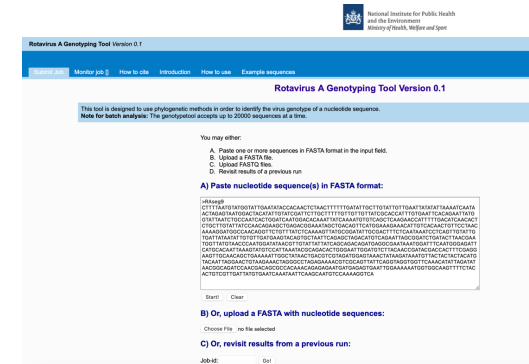
Gene segment	Gene	Genotypes	Nucleotide Cutoff
VP7	G	42	80%
VP4	P	58	80%
VP6	I	32	85%
VP1	R	28	83%
VP2	C	24	84%
VP3	M	24	81%
NSP1	A	39	79%
NSP2	N	28	85%
NSP3	T	28	85%
NSP4	E	32	85%
NSP5	H	28	91%

# Group A – current classification



**CENTRE FOR  
PATHOGEN  
GENOMICS**

- **Rotavirus A classification tool - National Institute for Public Health and the Environment, Netherlands**
  - Many new genotypes are not included – 2015 onwards not reliable
  - Incorrectly assigns some genotypes – based on phylogeny and bootstrap support




- **RV-BRC tool**
  - Appears to be limited to what was published in the 2011 classification paper
- **Neither tool is contacting RCWG for updates / info**

Services

**Subspecies Classification**  

The Subspecies Classification tool assigns the genotype/subtype of a virus, based on the genotype/subtype assignments maintained by the International Committee on Taxonomy of Viruses (ICTV). This tool infers the genotype/subtype for a query sequence from its position within a reference tree. The service uses the pplacer tool with a reference tree and reference alignment and includes the query sequence as input. Interpretation of the pplacer result is handled by Cladinator. Link to pplacer and Cladinator.


For further explanation, please see the Subspecies Classification Service Quick Reference Guide and Tutorial.


**Query source** 


Enter sequence  Select FASTA file


Enter one or more query nucleotide or protein sequences to search. Requires FASTA format.

**Species** 

Reoviridae - Rotavirus A [complete genome, genomic RNA] 

**Output Folder** 

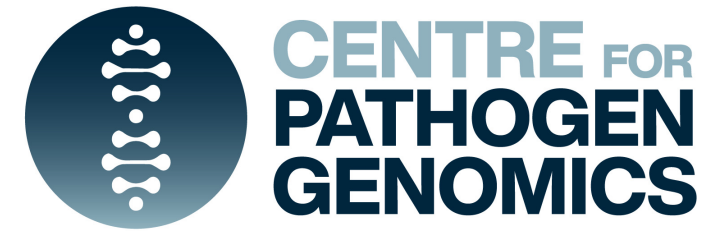
 Output Folder 

**Output Name** 

 Output Name



# Semi-curated databases



## Virus variation

- 74,381/114,164 sequences in GenBank
- Selecting by genotype is problematic
  - e.g. 350 G4 VP7 not 1300

## NCBI Virus

- Is up-to-date
- Can't filter by genotype

**Select sequence type**

Protein  Nucleotide  Full-length sequences only

**Define search set**

Species	Host	Region/Country	Segment	Isolation source
any Rotavirus A	any Bird Human Mammal Primate	any regions Africa Asia Europe	any VP1 VP2 VP3 VP4	any abdominal cavity animal tissues blood bursa

Collection date: [ ] [ ] [ ] to [ ] [ ] [ ]  
Release date: [ ] [ ] [ ] to [ ] [ ] [ ]  
Year Month Day Year Month Day

**Additional filters** [ ] [ ]

Keyword [ ] Search in [ ] sequence pattern [ ]

**Segments required in genome set:**  Select all  Full-length only

VP1	VP2	VP3	VP4	NSP1	NSP2	NSP3	NSP4	NSP5/6	VP6	VP7
any R1	any C1	any M1	any P[1]	any A1	any N1	any T1	any E1	any H1	any I1	any G1
R2	C2	M2	P[2]	A2	N2	T2	E2	H2	I2	G2
R3	C3	M3	P[3]	A3	N3	T3	E3	H3	I3	G3
R4	C4	M4	P[4]	A4	N4	T4	E4	H4	I4	G4
R5	C5	M5	P[5]	A5	N5	T5	E5	H5	I5	G5

Select the genotypes of required segments above.

NIH National Library of Medicine  
National Center for Biotechnology Information

NCBI Virus  
Sequences for discovery

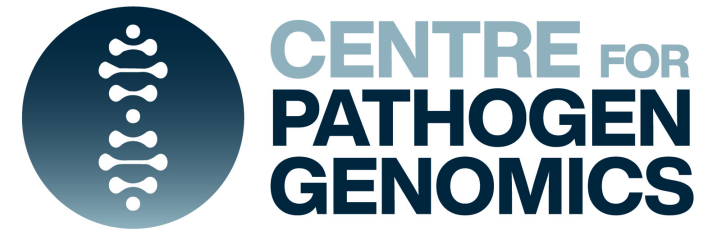
Explore Virus Data [Download] Popular Searches Influenza virus Dengue virus Zika virus Ebolavirus  
Rotavirus West Nile virus MERS coronavirus SARS-CoV-2 coronavirus

Advanced Filters for GenBank Sequences Visual Filters for GenBank Sequences Selected Results: 0 [Align] [Build Phylogenetic Tree]

**New! Randomized subsets in Downloads**  
You now have the option of downloading a smaller, randomized subset of the data shown in the Results table. Begin by using filters to refine your dataset, select the Nucleotide, Protein, or RefSeq Genome tab above the table for the datatype you would like to download, then follow the prompts in the Download menu. Our [Help documentation](#) has more information.

Refine Results		Nucleotide (114,041)	Protein (114,724)	RefSeq Genome (1)		
Accession	Organism Name	Submitters	Organization	Release Date	Isolate	
<input type="checkbox"/>	LC779915	Rotavirus A	Kumazaki, M., et al.	Microbiological Testing a...	2024-04-03	
<input type="checkbox"/>	LC779916	Rotavirus A	Kumazaki, M., et al.	Microbiological Testing a...	2024-04-03	

# Other genotype classification



## Random forest models - Tran et al 2023

- **VP7 and VP4 only**
- **Genotype where the total count was <10 were excluded to prevent classification of genotypes with insufficient amount of data to train the algorithm**
- **\*\*new classification tools need to be aware of existing classification used such a G and P typing or the RCWG recognized genotypes**

## Classification of group A rotavirus VP7 and VP4 genotypes using random forest

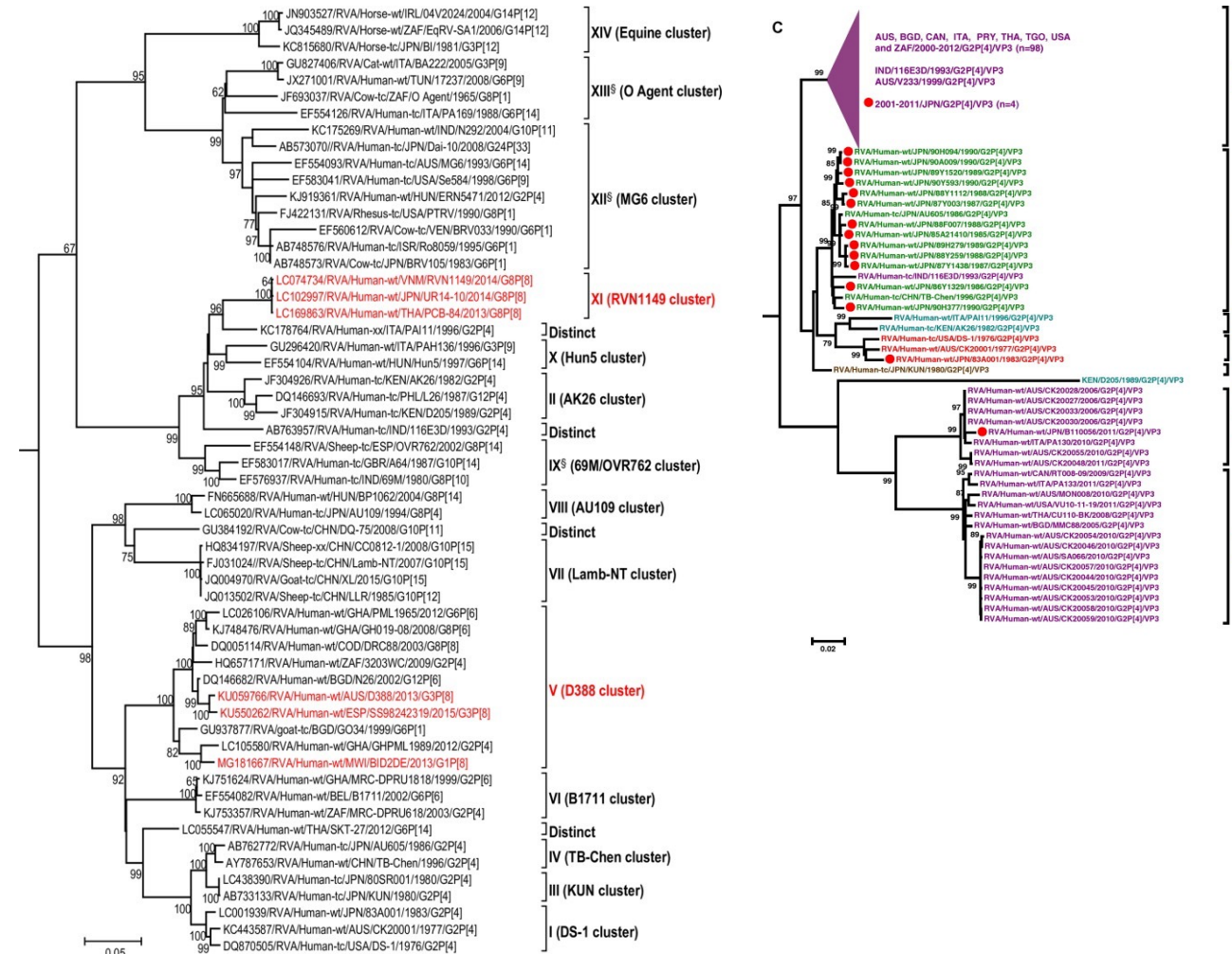
Hoc Tran\*, Robert Friendship and Zvonimir Poljak

Department of Population Medicine, Ontario Veterinary College, University of Guelph, Guelph, ON, Canada

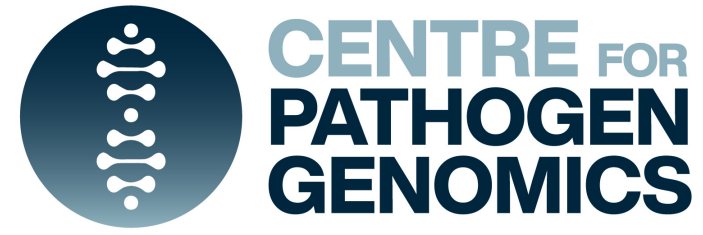
# Genotype lineage classification



- **Sub-lineage classification exists for many VP7 G and VP4 P genotypes**
- **Group 2 non-capsid genes**
- **Issues**
  - **Static - no updates since 2019**
  - **Trees were made in MEGA - topology changes if using RAxML or IQTree**
  - **Sequencing of older strains impacts tree topology**



# Conclusions



- **Issues**
  - **Diversity within genotypes is substantial**
  - **Sequencing is not routine and not timely – often 3-5 years behind**
  - **Divergent strains emerge via zoonotic transmission which can alter tree topology**
- **Automated, up-to-date, genotyping tool is required**
- **Existing tools need to be updated**