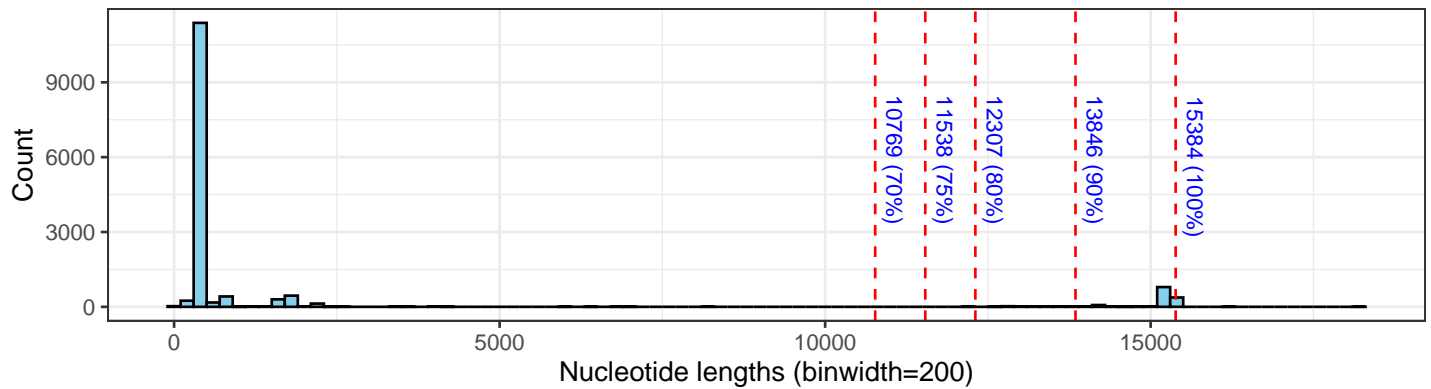


Background

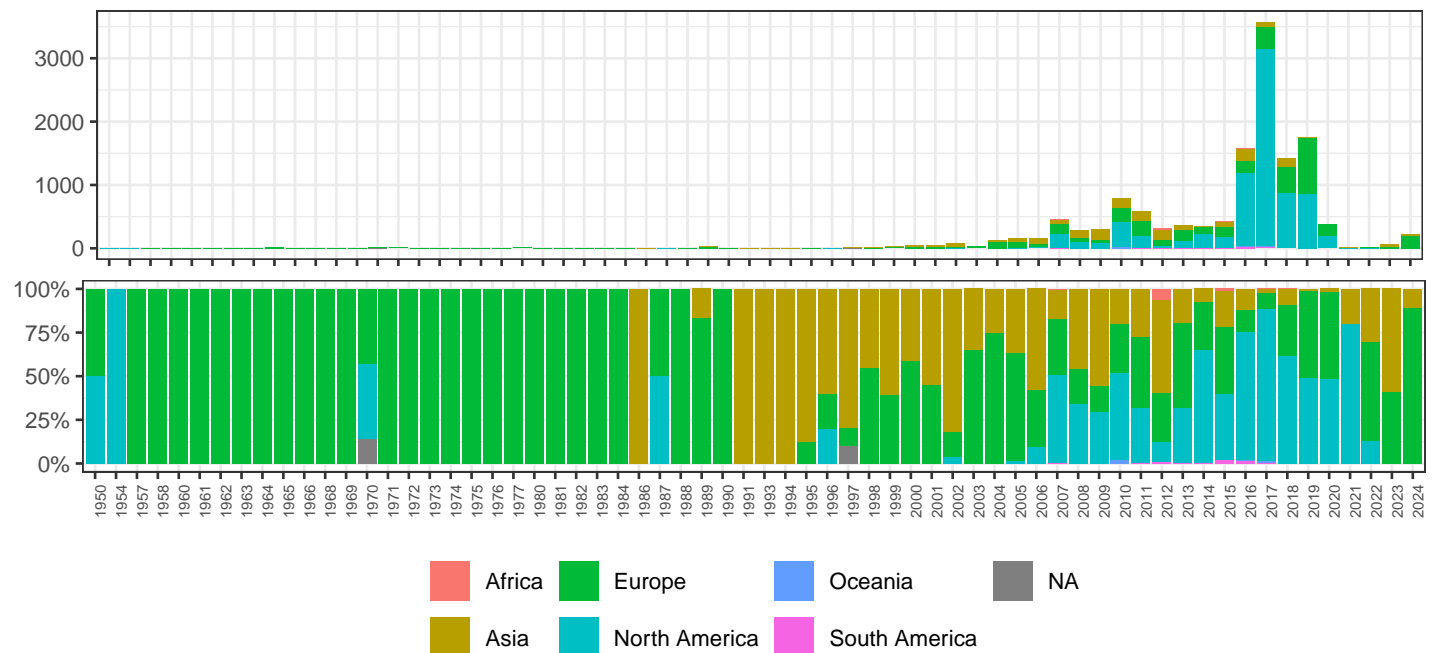
Exploratory Graphics

NCBI GenBank records

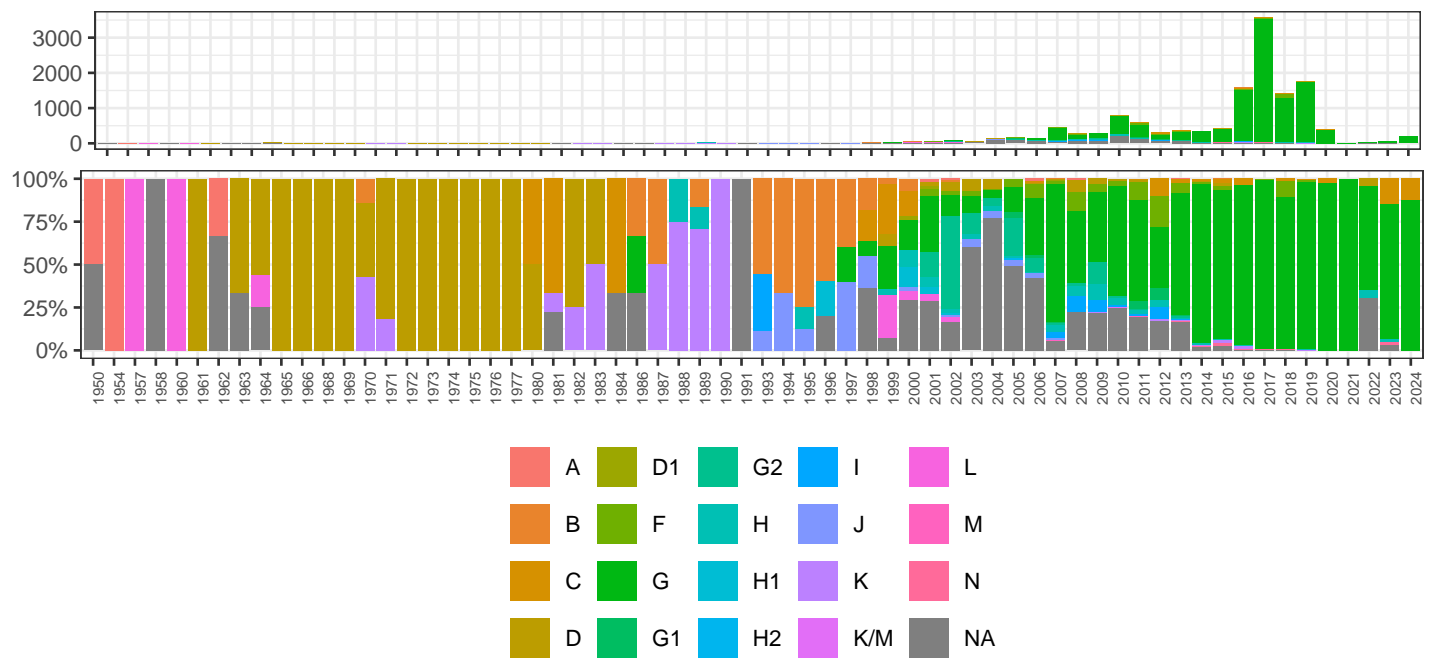
Diagnostic plot to estimate min-length filter for phylogenetic analysis (all data = 14548)



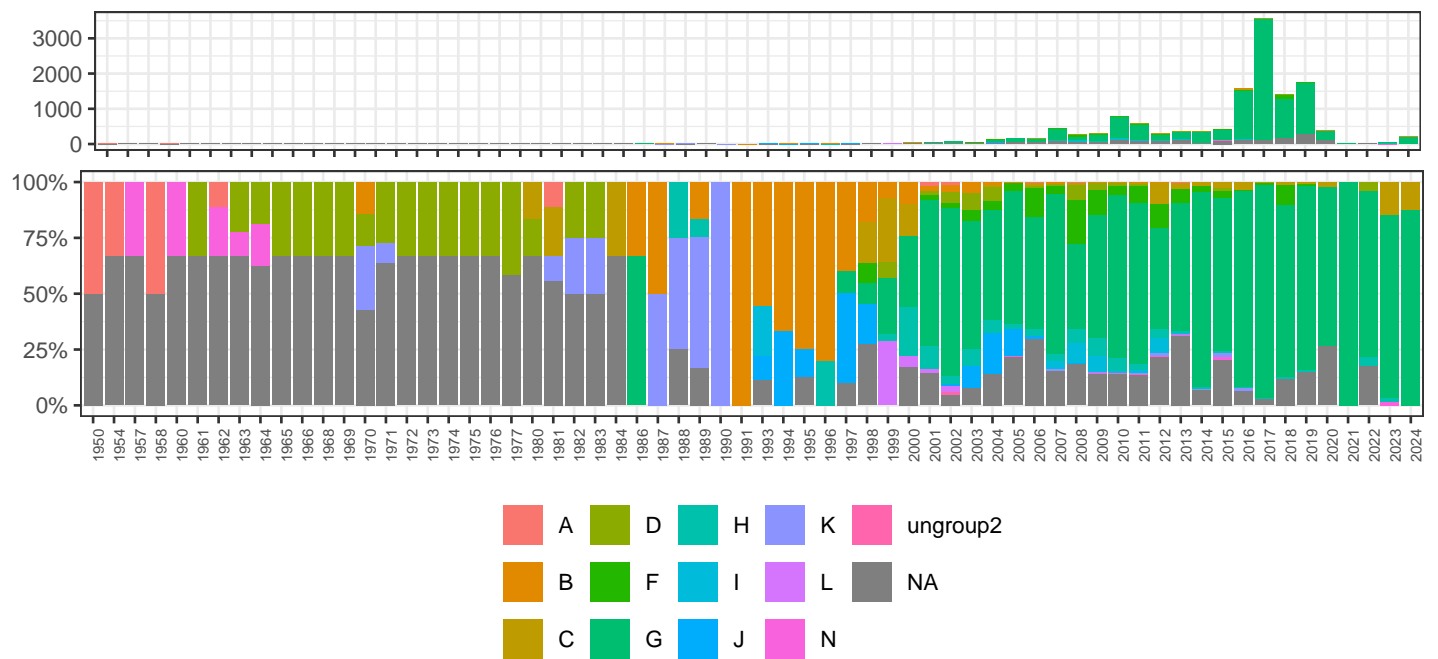
Frequency and proportion (n=13790)



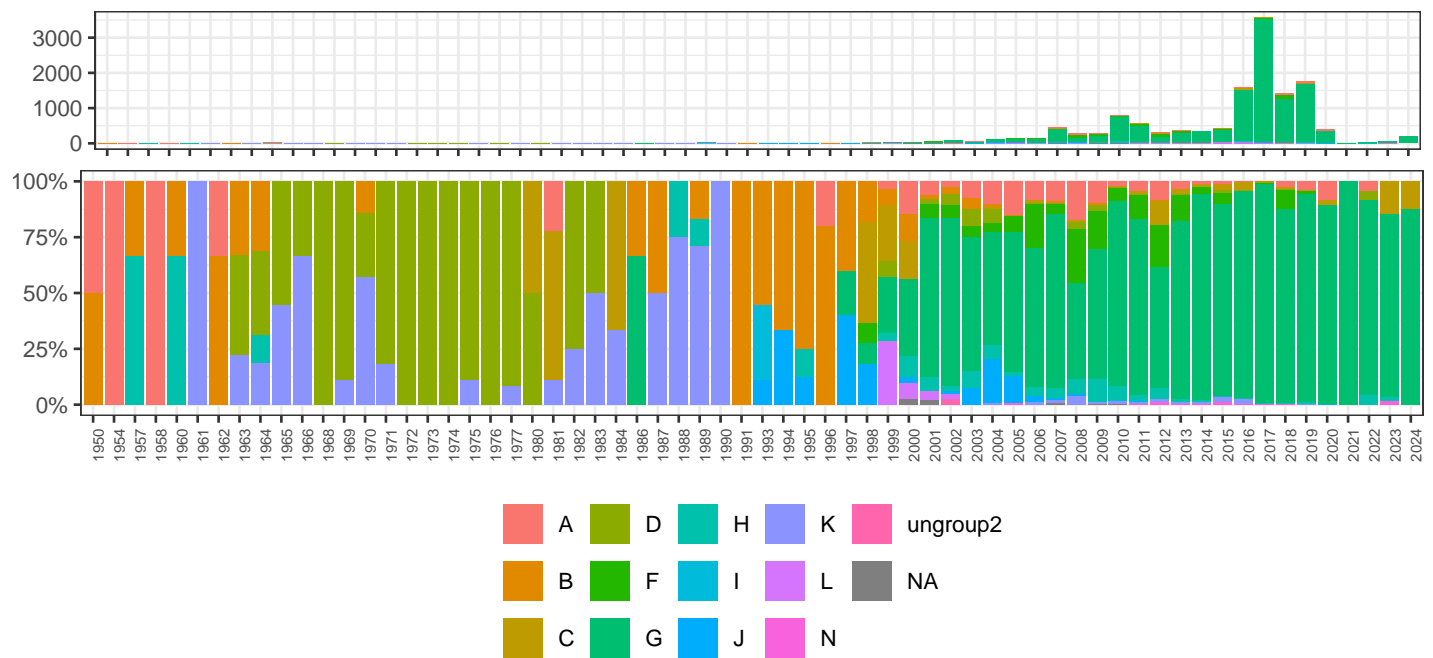
MuV Genotype (n=13790)



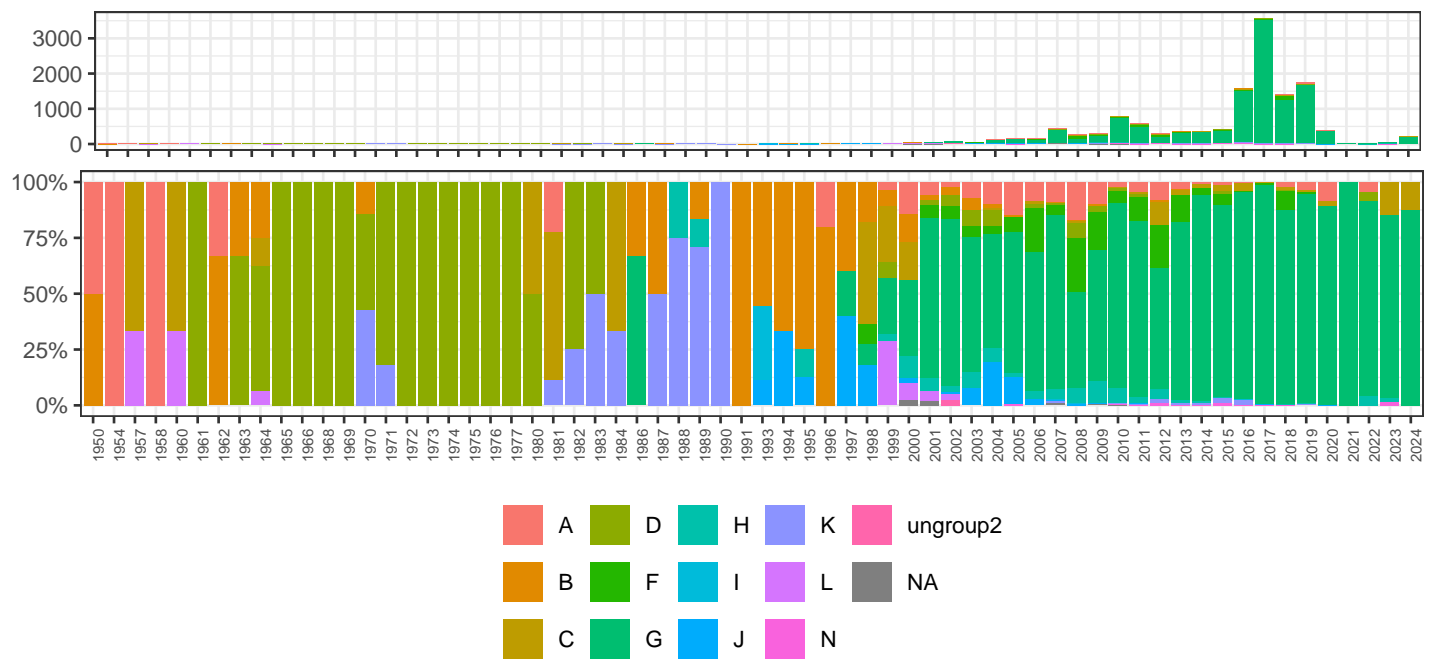
MuV Genotype_nextclade_sh (n=13790)



MuV Genotype_nextclade_genome (n=13790)



MuV Genotype_nextclade_genomesample (n=13790)



Nextclade dataset

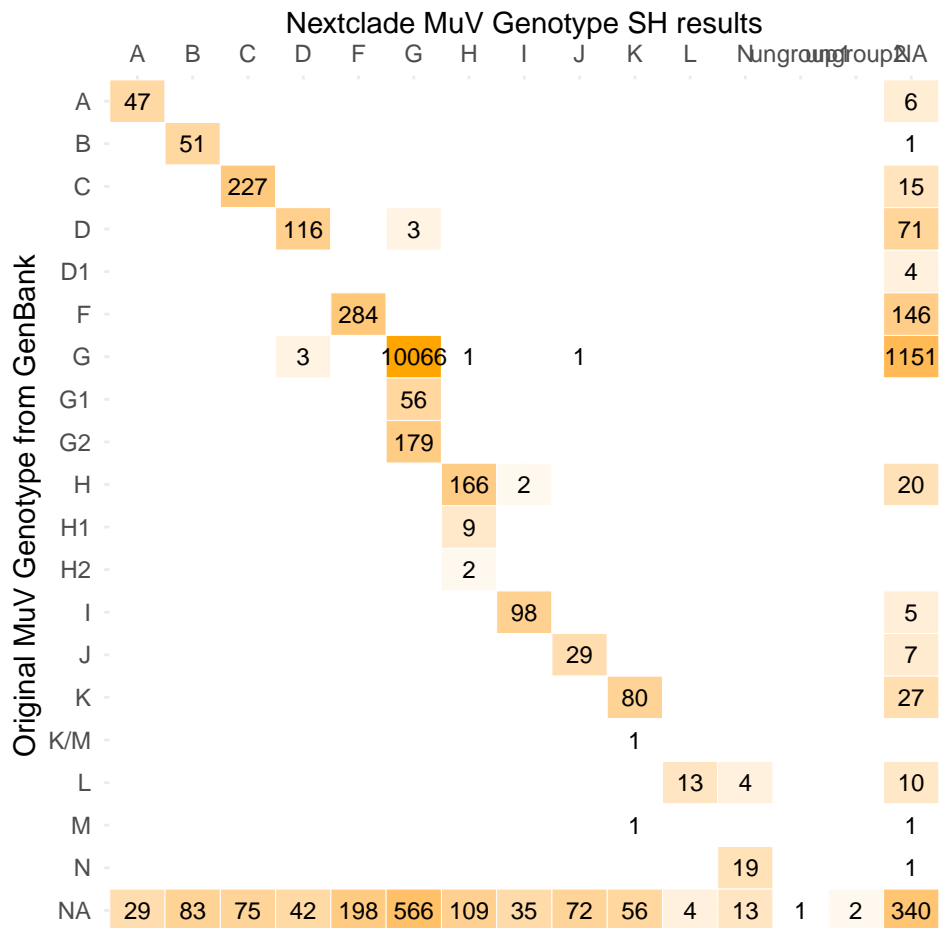


Table 1: Mismatched Genotype between GenBank and Nextclade datasets

accession	strain	length	genotype	MuV_genotype	SH
MN436859	MuVs/Segovia.ESP/15.09/[D]	316	Mumps virus genotype D	D	G
MN436862	MuVs/Madrid.ESP/21.09/[D]	316	Mumps virus genotype D	D	G
MN436869	MuVs/Madrid.ESP/36.09/[D]	316	Mumps virus genotype D	D	G
MZ913526	MuVs/Washington.USA/15.20[G]	316	Mumps virus genotype G	G	J
MN436857	MuVs/Ceuta.ESP/14.09/[G]	316	Mumps virus genotype G	G	D
MN436860	MuVs/Ceuta.ESP/16.09/[G]	316	Mumps virus genotype G	G	D
MN436868	MuVs/Ceuta.ESP/30.09/[G]	316	Mumps virus genotype G	G	D
MN511294	MuVs/Stockholm.SWE/19.19[G]	316	Mumps virus genotype G	G	H
AB600942	ODATE-3	15384	Mumps orthorubulavirus	H	I
KP213049	MuVi/Incheon.KOR/50.11/12[H]	316	Mumps virus genotype H	H	I
KX136901	MuVi/NLD/0.57[L]	316	Mumps virus genotype L	L	N
KX136902	MuVi/NLD/0.60/1[L]	316	Mumps virus genotype L	L	N
KX136903	MuVi/NLD/0.60/2[L]	316	Mumps virus genotype L	L	N
KX136908	MuVi/NLD/0.64/1[L]	316	Mumps virus genotype L	L	N

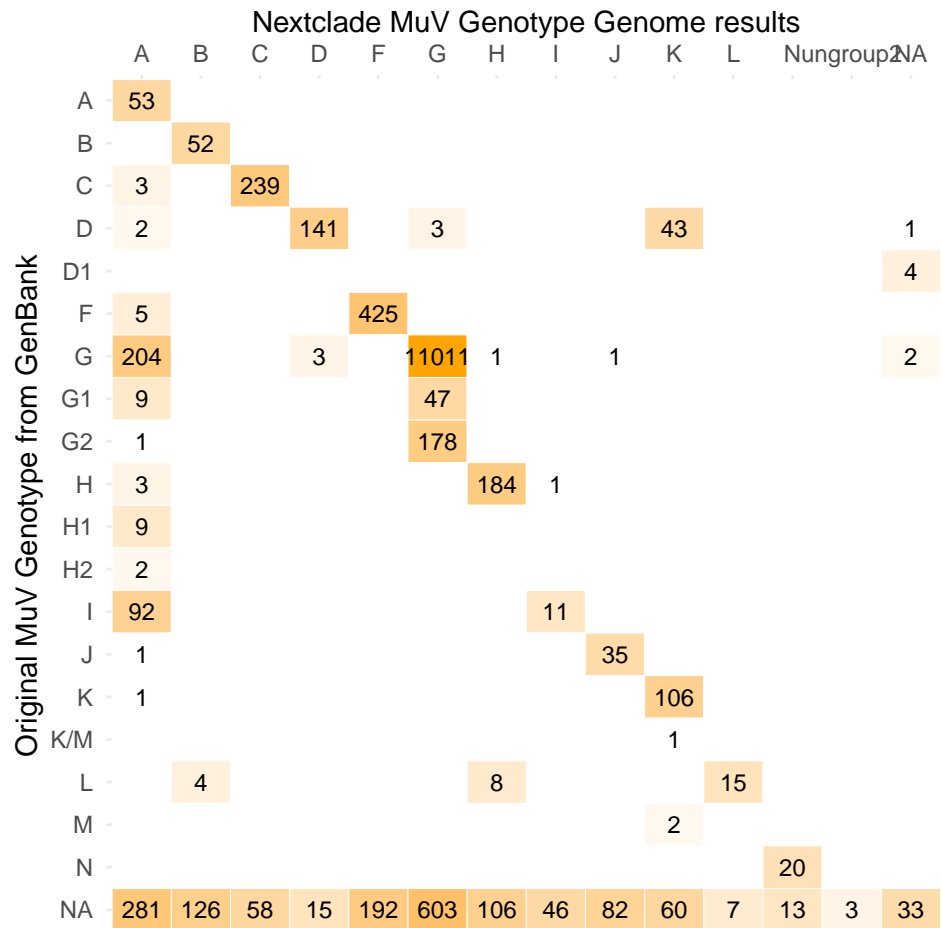


Table 2: Mismatched Genotype between GenBank and Nextclade datasets

accession	strain	length	MuV_genotype	SH	Genome
MH509394	MuV/ITA/52.00/INMI-2[C]	425	C	C	A
FJ919369	1610NE	156	C	C	A
KC429766	MuVi/Chennai.IND/45.12.4[C]	483	C	C	A
MN920214	MuVs/Quebec.CAN/26.19[D]	316	D	D	A
MW819866	MuVi/Sint Philipsland.NLD/02.08[D]	15375	D	D	K
MF522126	RIE26	606	D	D	K
MN436859	MuVs/Segovia.ESP/15.09/[D]	316	D	G	G
MN436862	MuVs/Madrid.ESP/21.09/[D]	316	D	G	G
MN436869	MuVs/Madrid.ESP/36.09/[D]	316	D	G	G
JQ034464	MuVi/Nottingham.GBR/19.04[D]	1749	D	NA	K

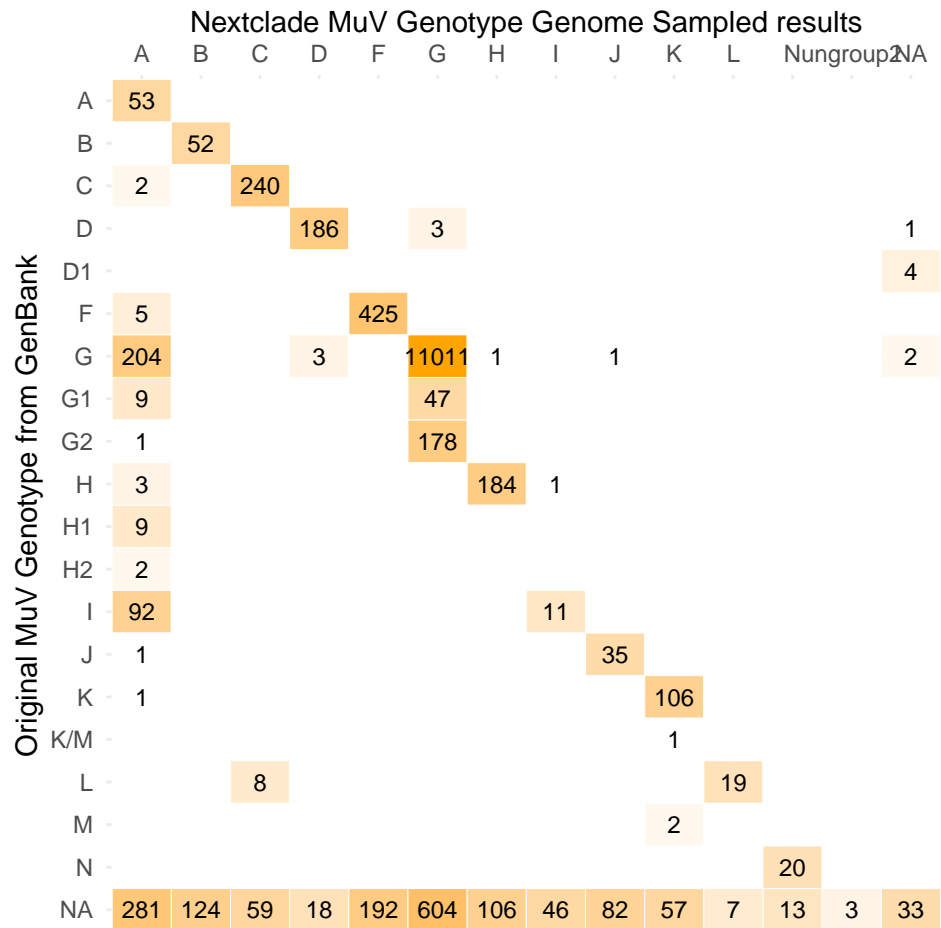


Table 3: Mismatched Genotype between GenBank and Nextclade datasets

accession	strain	length	MuV_genotype	SH	Genome	GenomeSample
MH509394	MuV/ITA/52.00/INMI-2[C]	425	C	C	A	A
FJ919369	1610NE	156	C	C	A	A
MN436859	MuVs/Segovia.ESP/15.09/[D]	316	D	G	G	G
MN436862	MuVs/Madrid.ESP/21.09/[D]	316	D	G	G	G
MN436869	MuVs/Madrid.ESP/36.09/[D]	316	D	G	G	G
MH521278	IQ-1	221	F	F	A	A
MH521279	IQ-2	221	F	F	A	A
MH521280	IQ-3	221	F	F	A	A
MH521281	IQ-4	221	F	F	A	A
MH521282	IQ-5	221	F	F	A	A

Debugging Nextclade Assignments

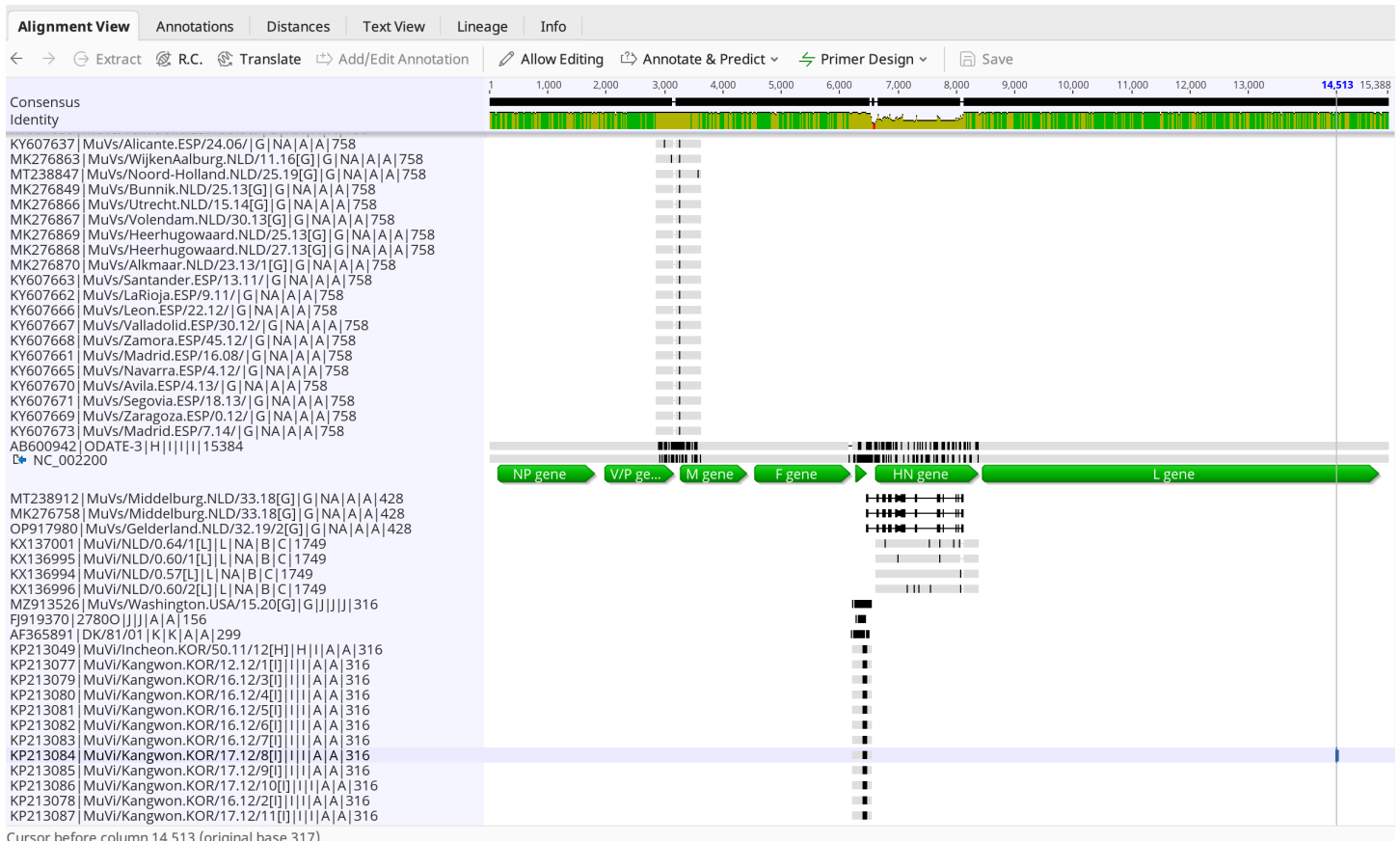
```
nextstrain shell .

#export BUILD=sh
#export METADATA_COLUMNS="strain,MuV_genotype,SH,length"
#export BUILD=genome
#export METADATA_COLUMNS="strain,MuV_genotype,SH,Genome,length"
export BUILD=genomesample
export METADATA_COLUMNS="strain,MuV_genotype,SH,Genome,GenomeSample,length"

augur filter \
  --metadata mumps/mismatch_${BUILD}.tsv \
  --metadata-id-columns accession \
  --sequences ~/Desktop/MUMPS/mumps/ingest/results/sequences.fasta \
  --output-sequences mumps/mismatch_${BUILD}_raw.fasta

python ~/github/j23414/basic-phylogenetic-pipeline/bin/annotate_header.py \
  --metadata mumps/mismatch_${BUILD}.tsv \
  --metadata-columns ${METADATA_COLUMNS} \
  --sequences mumps/mismatch_${BUILD}_raw.fasta \
  > mumps/mismatch_${BUILD}.fasta

nextclade3 run \
  mumps/mismatch_${BUILD}.fasta \
  --input-dataset ~/github/nextstrain/mumps-branches/add-nextclade-workflow-wgs/nextclade_data/${BUILD} \
  --output-tsv mumps/nextclade_${BUILD}.tsv \
  --output-fasta mumps/nextclade_${BUILD}.fasta \
  --silent
```



- Mis-assignments seem to be in the gap between V/P and M gene (weird indel), SH region for Genotype I, and HN gene.

Table 4: Top 25 most frequent sequence submitters with their region and countries

authors	n	regions	countries
hiebert et al.	3174	North America	Canada
mcnall et al.	1502	North America	USA
castellanos et al.	1061	Europe	Spain
wharton et al.	945	North America, Oceania	USA, Micronesia, Marshall Islands
cui et al.	514	Europe, Asia, North America	United Kingdom, China, Dominican Republic
bodewes et al.	488	Europe	Netherlands
gouma et al.	485	Europe, North America	Netherlands, USA
rota et al.	425	North America, NA	USA, NA
kidokoro et al.	392	Asia, NA	Japan, Mongolia, NA
frost et al.	319	Europe, North America	United Kingdom, Canada
bryant et al.	315	North America	USA
kim et al.	285	North America, Asia, NA	USA, South Korea, NA, Korea
hickman et al.	254	North America	USA
peran-ramos et al.	253	Europe	Spain
aoki et al.	249	Asia	Japan
gavilan et al.	227	Europe	Spain
rubalskaia et al.	218	Europe	Russia
ma et al.	178	Asia	China
catellanos et al.	172	Europe	Spain
byrne et al.	169	North America	USA
moncla et al.	166	North America	USA
anton et al.	155	Europe	Spain
shah et al.	154	Europe, North America	Netherlands, Canada
rivailler et al.	144	Oceania, North America	Guam, USA
jin et al.	142	NA, Europe, Asia, North America	NA, Sweden, Japan, United Kingdom, USA, Germany, Malaysia,

Table 5: Top 20 earliest records

date_adjusted	accession	strain	country	authors
1950-01-01	JQ946042	MuVi/Taylor.GBR/0.50s	United Kingdom	Jin et al.
1950-01-01	KF876715	MuVi/Kilham.USA/0.50[A]	USA	Jin et al.
1954-01-01	KX136900	MuVi/Albany.USA/0.54[A]	USA	Gouma et al.
1954-01-01	KX136946	MuVi/Albany.USA/0.54[A]	USA	Gouma et al.
1954-01-01	KX136993	MuVi/Albany.USA/0.54[A]	USA	Gouma et al.
1957-01-01	KX136901	MuVi/NLD/0.57[L]	Netherlands	Gouma et al.
1957-01-01	KX136947	MuVi/NLD/0.57[L]	Netherlands	Gouma et al.
1957-01-01	KX136994	MuVi/NLD/0.57[L]	Netherlands	Gouma et al.
1958-01-01	JQ034458	Enders-58-2	United Kingdom	Cui et al.
1958-01-01	JQ034507	ENDERS-58-2	United Kingdom	Cui et al.
1960-01-01	KX136902	MuVi/NLD/0.60/1[L]	Netherlands	Gouma et al.
1960-01-01	KX136903	MuVi/NLD/0.60/2[L]	Netherlands	Gouma et al.
1960-01-01	KX136948	MuVi/NLD/0.60/1[L]	Netherlands	Gouma et al.
1960-01-01	KX136949	MuVi/NLD/0.60/2[L]	Netherlands	Gouma et al.
1960-01-01	KX136995	MuVi/NLD/0.60/1[L]	Netherlands	Gouma et al.
1960-01-01	KX136996	MuVi/NLD/0.60/2[L]	Netherlands	Gouma et al.
1961-01-01	KX136904	MuVi/NLD/0.61[D]	Netherlands	Gouma et al.
1961-01-01	KX136950	MuVi/NLD/0.61[D]	Netherlands	Gouma et al.
1961-01-01	KX136997	MuVi/NLD/0.61[D]	Netherlands	Gouma et al.
1962-01-01	KX136905	MuVi/NLD/0.62/1	Netherlands	Gouma et al.

Table 6: Top 20 latest records

date_adjusted	accession	strain	country	authors
2024-03-22	PQ311690	MuVs/KIPMR/Chennai.IND/4.24/3[C]	India	Kaveri et al.
2024-03-23	PQ311691	MuVs/KIPMR/Chennai.IND/4.24/4[C]	India	Kaveri et al.
2024-03-23	PQ311692	MuVs/KIPMR/Chennai.IND/4.24/5[C]	India	Kaveri et al.
2024-03-25	PQ311693	MuVs/KIPMR/Chennai.IND/5.24[C]	India	Kaveri et al.
2024-05-14	PQ311694	MuVs/KIPMR/Chennai.IND/3.24[C]	India	Kaveri et al.
2024-05-14	PQ311695	MuVs/KIPMR/Chennai.IND/3.24/2[C]	India	Kaveri et al.
2024-05-14	PQ311696	MuVs/KIPMR/Chennai.IND/3.24/3[C]	India	Kaveri et al.
2024-05-14	PQ311697	MuVs/KIPMR/Chennai.IND/3.24/4[C]	India	Kaveri et al.
2024-05-17	PQ451425	MuVs/Odisha.INDIA/20.24[C]	India	Mamidi et al.
2024-05-17	PQ451426	MuVs/Odisha.INDIA/20.24/1[C]	India	Mamidi et al.
2024-05-17	PQ451427	MuVs/Odisha.INDIA/20.24/2[C]	India	Mamidi et al.
2024-05-17	PV072739	MuVs/Odisha.INDIA/20.24/8[C]	India	Mishra et al.
2024-05-27	PV072740	MuVs/Odisha.INDIA/20.24/9[C]	India	Mishra et al.
2024-06-07	PQ001008	MuVs/Makhchkala.RUS/23.24[C]	Russia	Zamotaeva et al.
2024-06-07	PQ001009	MuVs/Makhchkala.RUS/23.24[G]	Russia	Zamotaeva et al.
2024-06-19	PV072734	MuVs/Odisha.INDIA/20.24/3[C]	India	Mishra et al.
2024-07-05	PV072735	MuVs/Odisha.INDIA/20.24/4[C]	India	Mishra et al.
2024-07-31	PV072738	MuVs/Odisha.INDIA/20.24/7[C]	India	Mishra et al.
2024-08-09	PV072737	MuVs/Odisha.INDIA/20.24/6[C]	India	Mishra et al.
2024-08-29	PV072736	MuVs/Odisha.INDIA/20.24/5[C]	India	Mishra et al.