Nextclade Result Parameter Diagnostic Plots

Full Genome

Diagnostic plot to estimate min−length filter for phylogenetic analysis (all data = 14548)
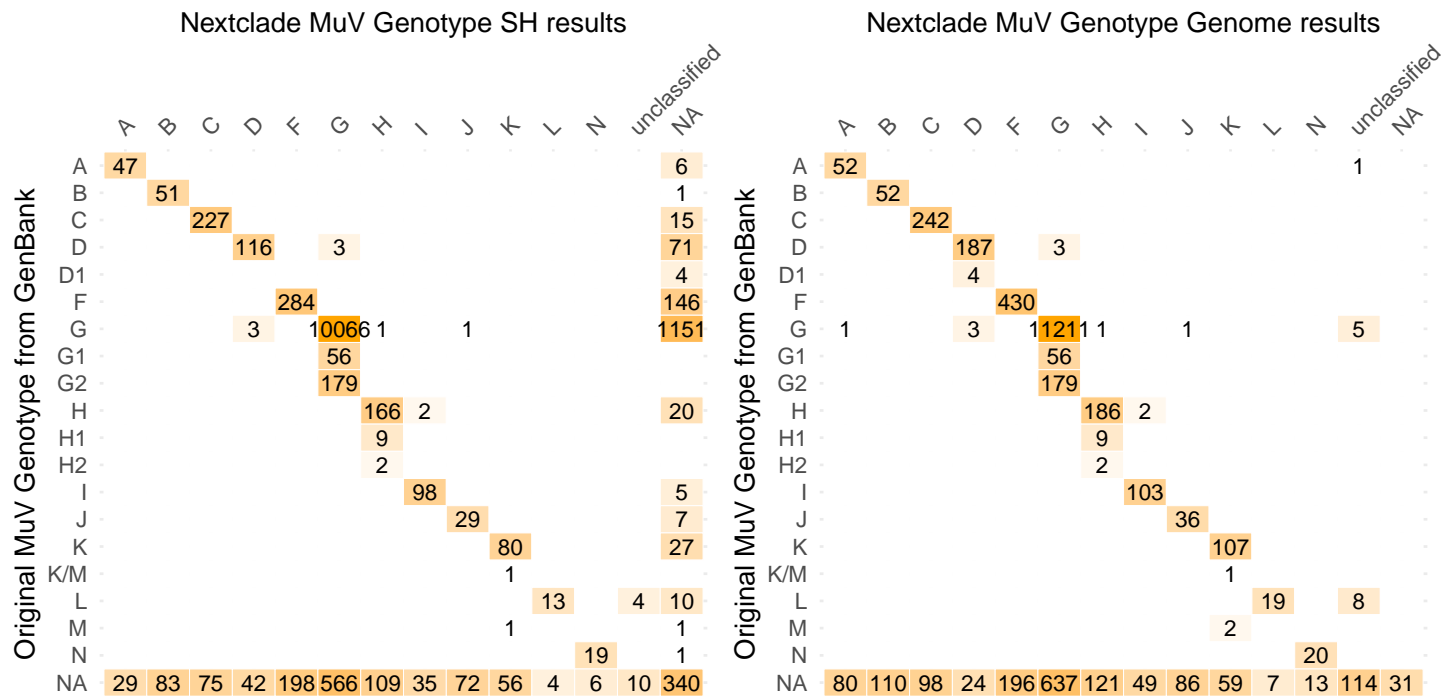
Count — Nucleotide lengths (binwidth=200)

Red dashed line markers: 10769 (70%), 11538 (75%), 12307 (80%), 13846 (90%), 15384 (100%)

SH Region

Diagnostic plot to estimate min−length filter for phylogenetic analysis (all data = 11657)

Count — Nucleotide lengths (binwidth=20)

Red dashed line markers: 220 (70%), 236 (75%), 252 (80%), 284 (90%), 315 (100%)

Nextclade compared with GenBank annotations:

Nextclade MuV Genotype SH results

Original MuV Genotype from GenBank (rows) vs Nextclade result (columns)

	A	B	C	D	F	G	H	I	J	K	L	N	unclassified	NA
A	47													6
B		51												1
C			227											15
D				116	3									71
D1														4
F					284									146
G			3			10066	1		1					1151
G1						56								
G2						179								
H							166	2						20
H1							9							
H2							2							
I								98						5
J									29					7
K										80				27
K/M								1						
L									13		4		10	
M							1							1
N												19		1
NA	29	83	75	42	198	566	109	35	72	56	4	6	10	340

Nextclade MuV Genotype Genome results

Original MuV Genotype from GenBank (rows) vs Nextclade result (columns)

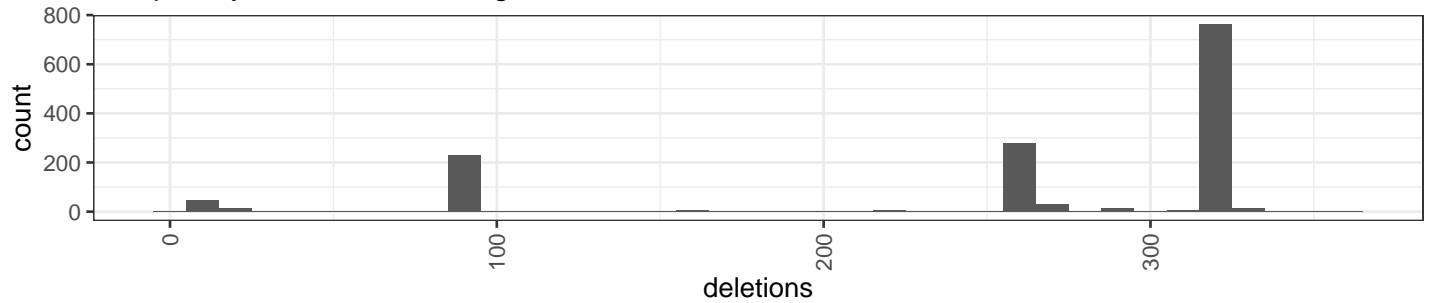|	A	B	C	D	F	G	H	I	J	K	L	N	unclassified	NA
A	52												1	
B		52												
C			242											
D				187	3									
D1				4										
F					430									
G	1		3			1121	1	1	1				5	
G1						56								
G2						179								
H							186	2						
H1							9							
H2							2							
I								103						
J									36					
K										107				
K/M								1						
L										19			8	
M									2					
N												20		
NA	80	110	98	24	196	637	121	49	86	59	7	13	114	31

SH gene Nextclade Dataset

Alignment check

Mumps/SH/example

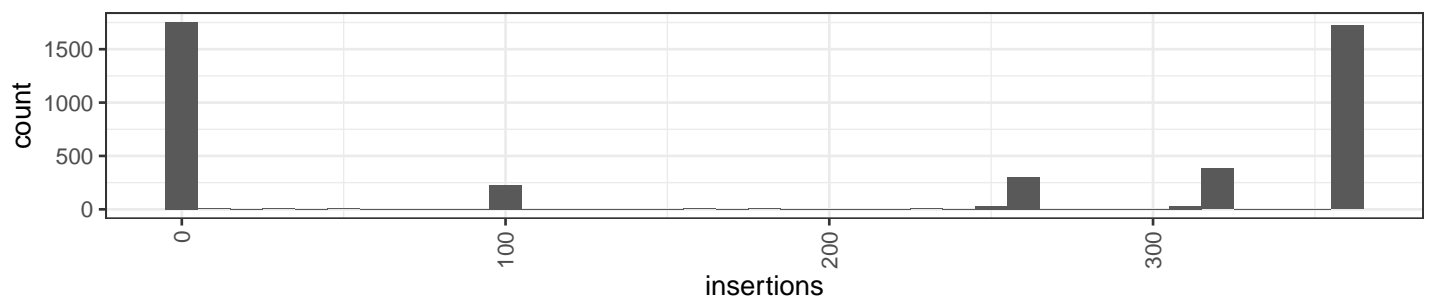### Frequency of deletions along SH
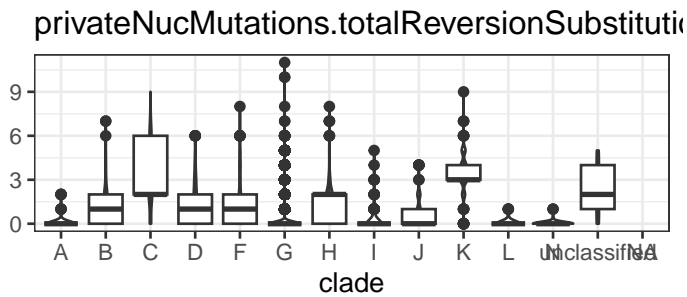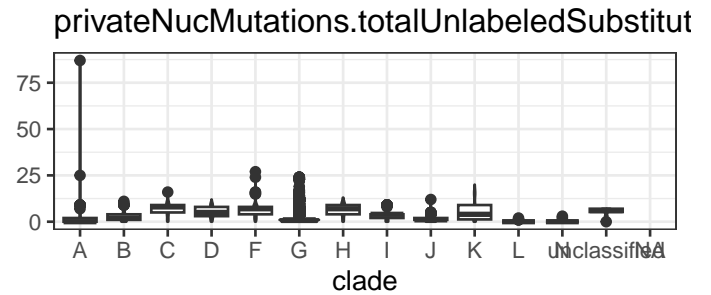

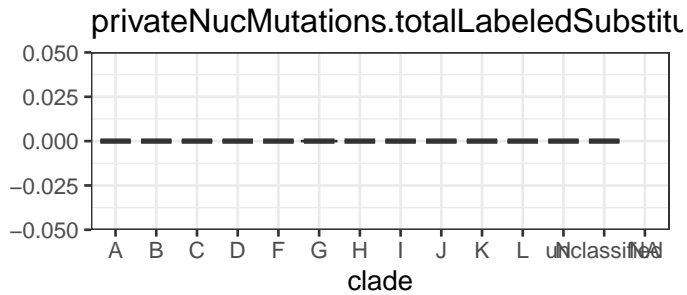### Frequency of insertions along SH

Table 1: Top Frameshifts

frameShifts	n
SH:7	227
SH:empty range	32
SH:49	4
SH:27-28	3
SH:27-58	2
SH:28	2
SH:31-33	2
SH:15-58	1
SH:18-26	1
SH:2-3	1

Missing Data

```
--scoreBias 300               # First 300 not penalized
--missingDataThreshold 3000   # Top threshold, scale the next set of missing data from 0 to 100
```

Since this is the SH gene (315nt), we do not reach past the 300 minimum.
Mixed Sites (M)

```
--mixedSitesThreshold 10 # More than this number of mixed sites results in a "bad" score
```


Well below the 10 threshold (but once again this is for a reference that is 315nt long).

Private Mutations

```json
"privateMutations": {
  "enabled": true,
  "typical": 8,
  "cutoff": 24,
  "weightLabeledSubstitutions": 4,
  "weightReversionSubstitutions": 6,
  "weightUnlabeledSubstitutions": 1
},
```

privateNucMutations.totalLabeledSubstitut

privateNucMutations.totalUnlabeledSubstitut

privateNucMutations.totalReversionSubstitutic

privateNucMutations.totalPrivateSubstitution

Mutation clusters

```json
"snpClusters": {
  "enabled": true,
  "windowSize": 100,
  "clusterCutOff": 6,
  "scoreWeight": 50
},
```

totalSubstitutions

totalAminoacidSubstitutions

totalAminoacidInsertions

totalAminoacidDeletions

4

Stop Codons

```
"stopCodons": {
     "enabled": true,
     "ignoredStopCodons": [
       {
         "geneName": "ORF8",
         "codon": 26
       },
       {
         "geneName": "ORF8",
         "codon": 67
       }
     ]
}
```

qc.stopCodons.totalStopCodons

Frameshifts

```
"frameShifts": {
   "enabled": true,
   "ignoredFrameShifts": [
     {
       "geneName": "ORF3a",
       "codonRange": {
         "begin": 256,
         "end": 276
       }
     },
     {
       "geneName": "ORF3a",
       "codonRange": {
         "begin": 258,
         "end": 276
       }
     }
   ]
},
```

totalFrameShifts

Genome Nextclade Dataset

Alignment check

Mumps/Genome/example

Frequency of deletions along Genome
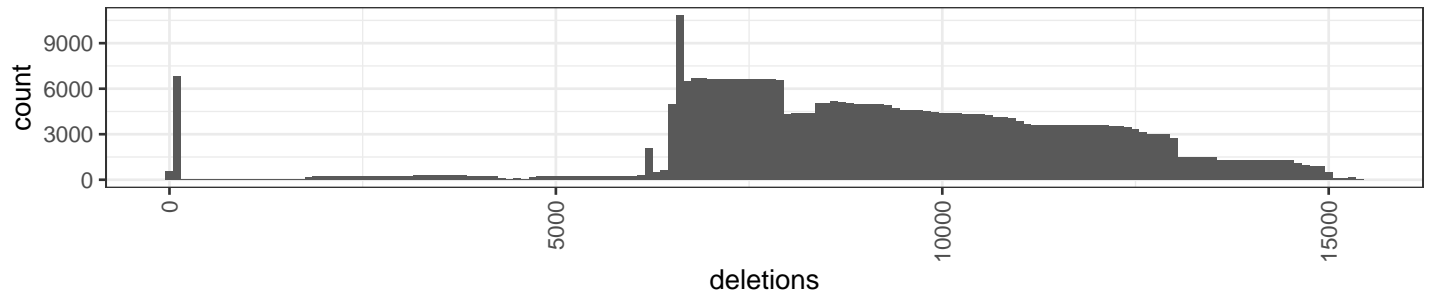
Frequency of insertions along Genome
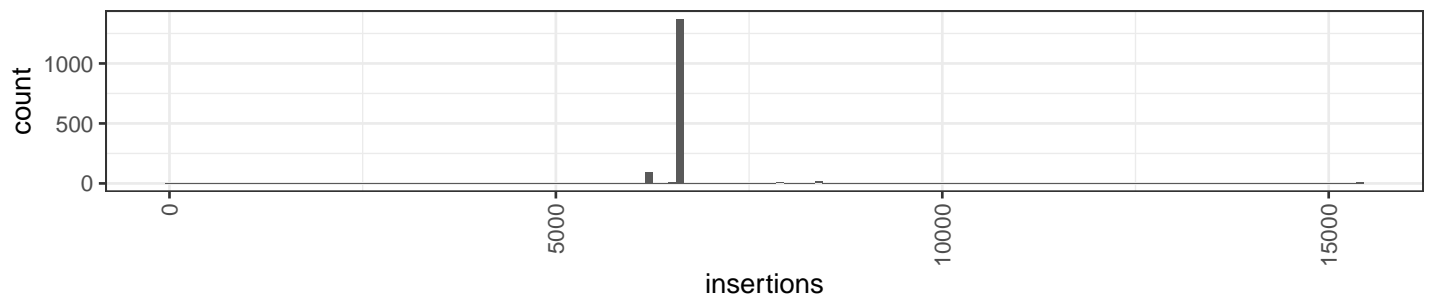
Table 2: Top Frameshifts

frameShifts	n
HN:445-446	7
P:156-391	6
SH:55-58	4
L:1385-2262	3
NP:114-550	2
SH:27-58	2
F:483-488	1
F:495-503	1
F:503-510	1
F:517-520	1

Missing Data

```
--scoreBias 300               # First 300 not penalized
--missingDataThreshold 3000   # Top threshold, scale the next set of missing data from 0 to 100
```
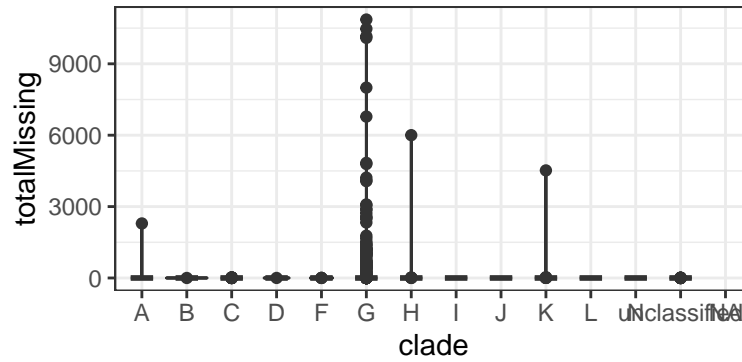
The full genome does reach the 300 minimum, especially for samples that are only reporting one gene (SH region or 315 out of 14k genome length)
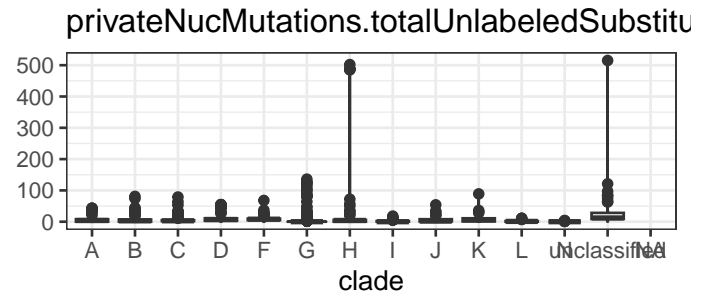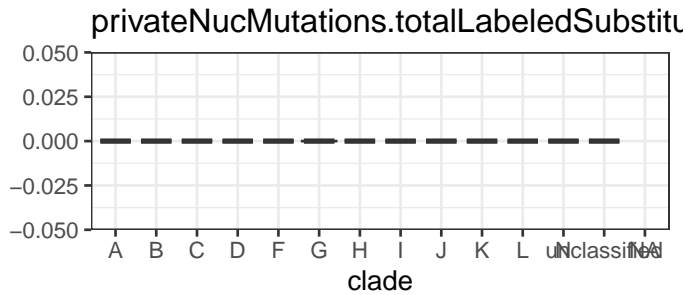
Mixed Sites (M)

```
--mixedSitesThreshold 10 # More than this number of mixed sites results in a "bad" score
```
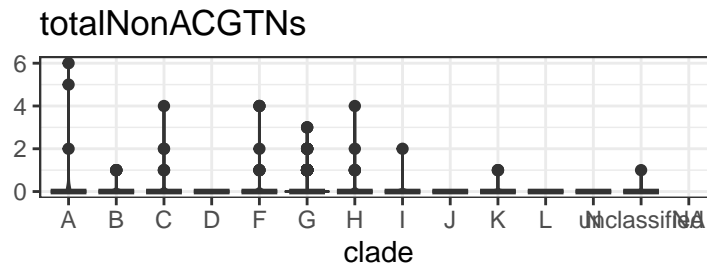

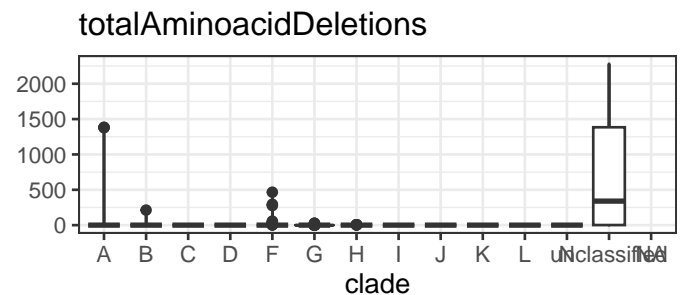This is still well below the 10 threshold.
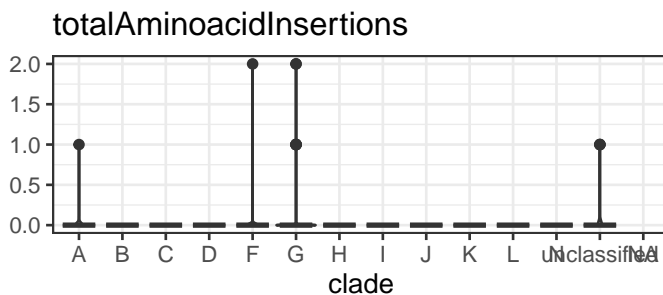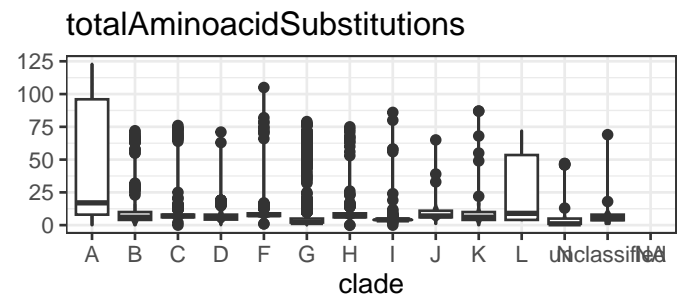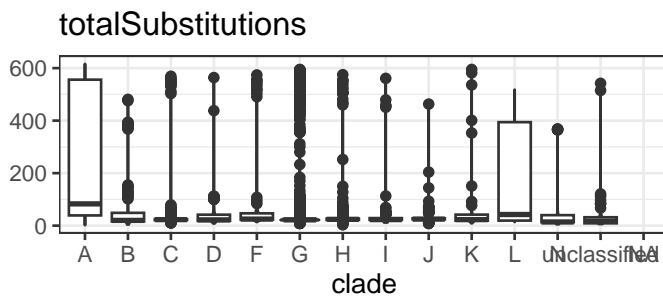
Private Mutations

```
"privateMutations": {
  "enabled": true,
  "typical": 8,
  "cutoff": 24,
  "weightLabeledSubstitutions": 4,
  "weightReversionSubstitutions": 6,
  "weightUnlabeledSubstitutions": 1
},
```
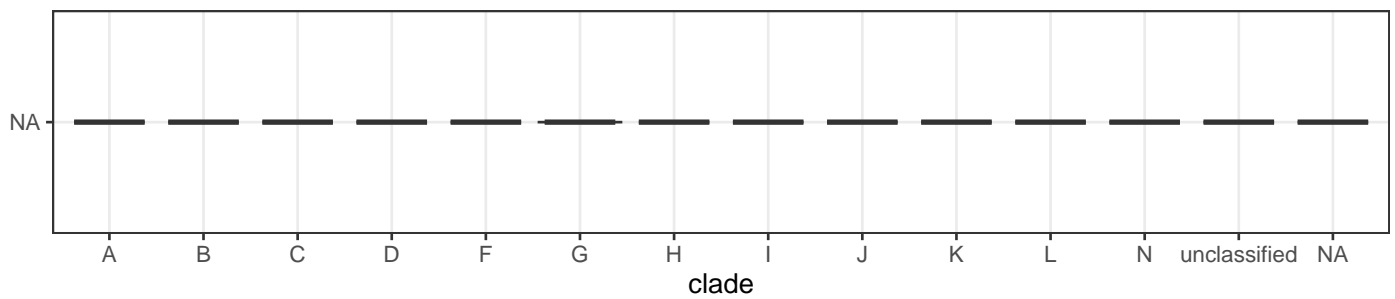

privateNucMutations.totalLabeledSubstitu

privateNucMutations.totalUnlabeledSubstitu

privateNucMutations.totalReversionSubstitut

privateNucMutations.totalPrivateSubstition
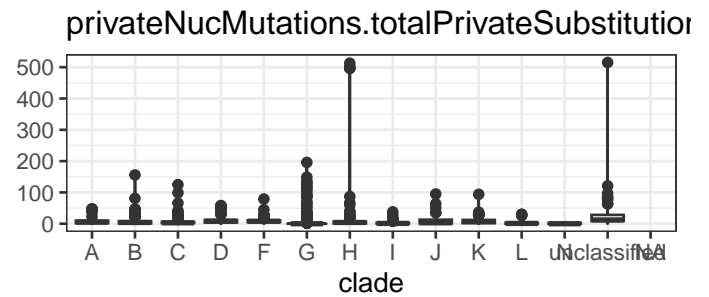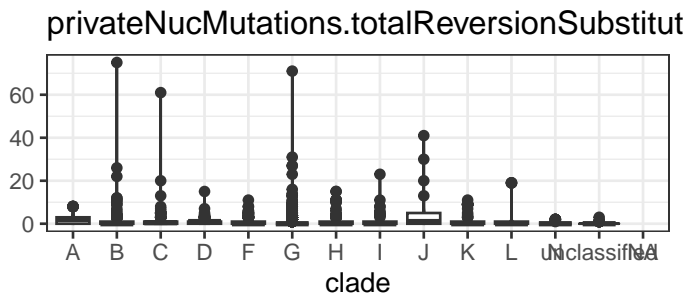
Mutation clusters

```
"snpClusters": {
  "enabled": true,
  "windowSize": 100,
  "clusterCutOff": 6,
  "scoreWeight": 50
},
```


totalSubstitutions

totalAminoacidSubstitutions

totalAminoacidInsertions

totalAminoacidDeletions

8

Stop Codons

```
"stopCodons": {
    "enabled": true,
    "ignoredStopCodons": [
      {
        "geneName": "ORF8",
        "codon": 26
      },
      {
        "geneName": "ORF8",
        "codon": 67
      }
    ]
}
```

qc.stopCodons.totalStopCodons

Frameshifts

```
"frameShifts": {
  "enabled": true,
  "ignoredFrameShifts": [
    {
      "geneName": "ORF3a",
      "codonRange": {
        "begin": 256,
        "end": 276
      }
    },
    {
      "geneName": "ORF3a",
      "codonRange": {
        "begin": 258,
        "end": 276
      }
    }
  ]
},
```

totalFrameShifts