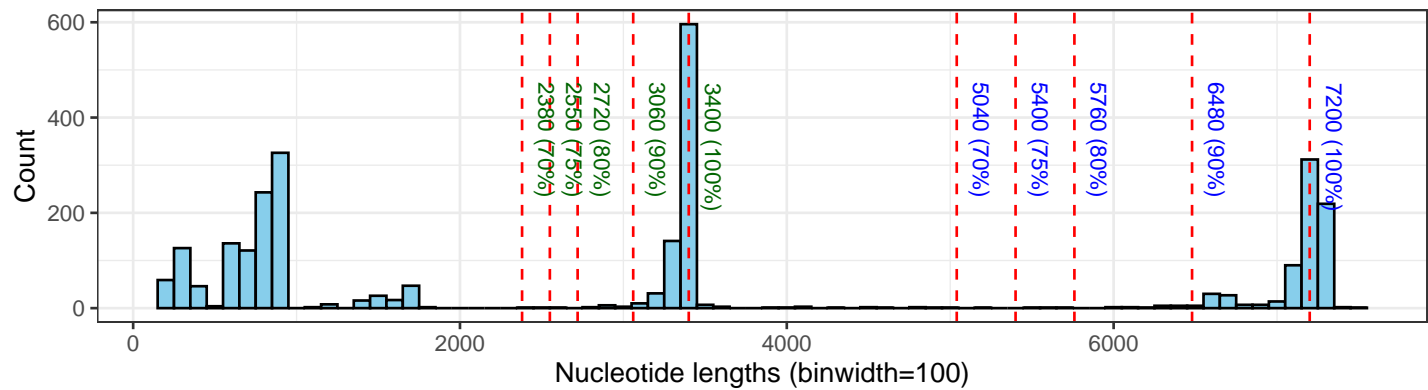


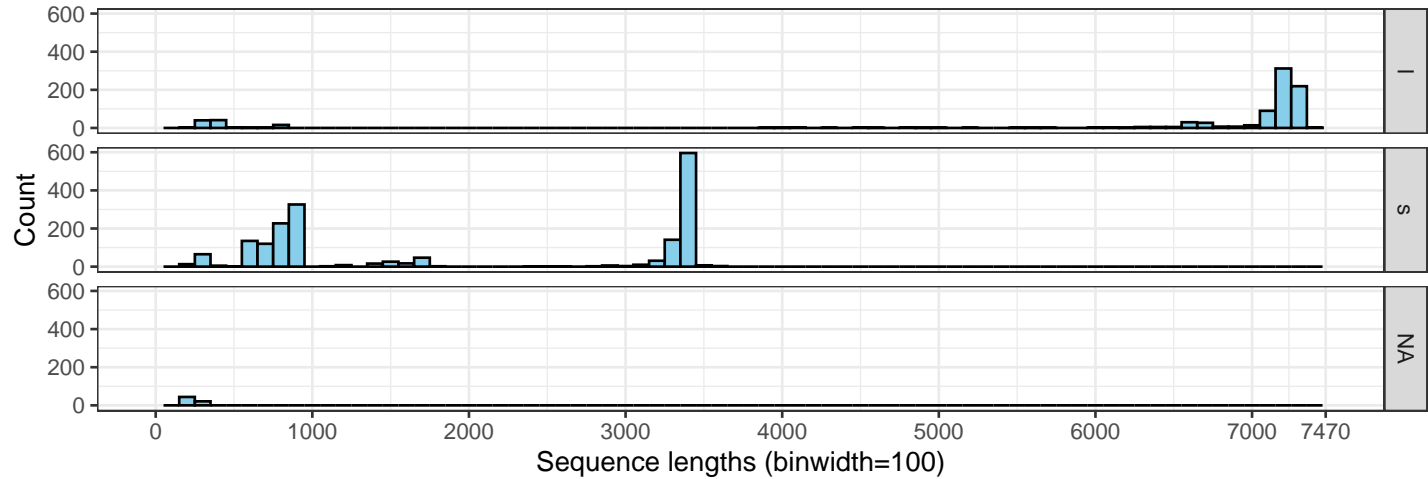
Background

Exploratory Graphics

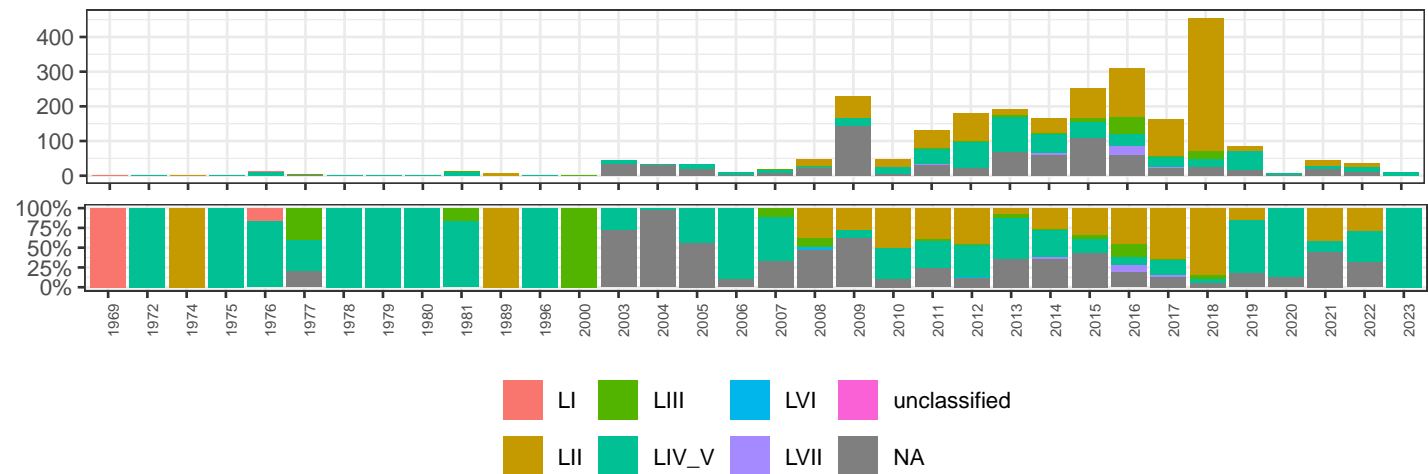
Diagnostic plot to estimate min-length filter for phylogenetic analysis (all data = 2727)



Sequence lengths (all data = 2727)



Frequency and proportion (n=2552)



L segment Nextclade Dataset

Alignment check

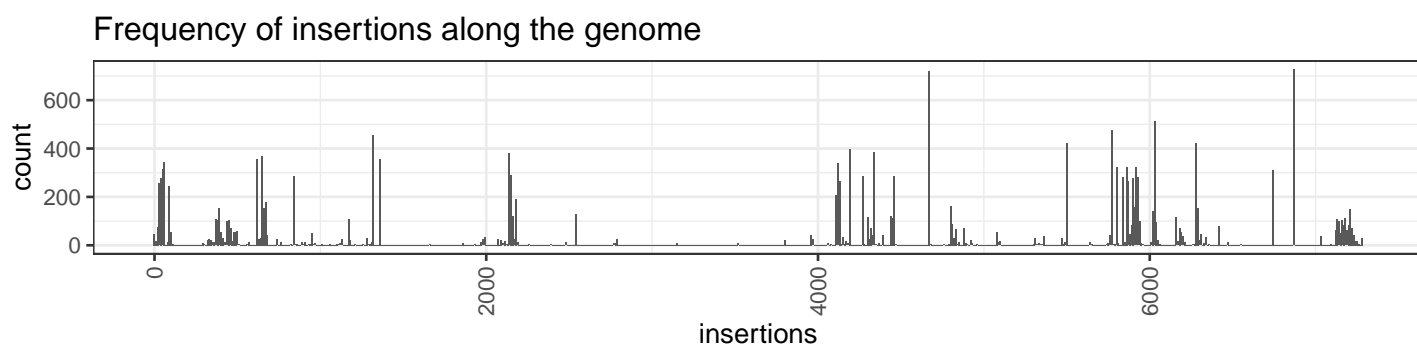
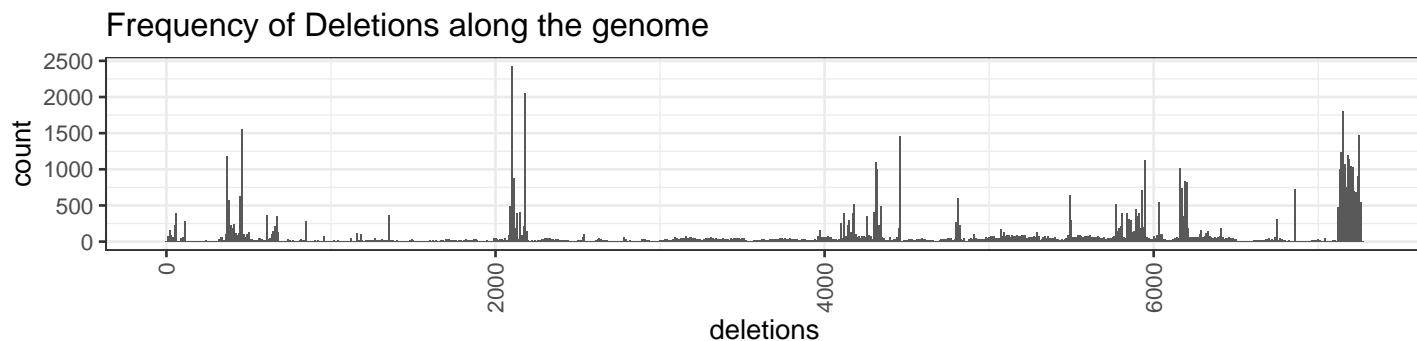
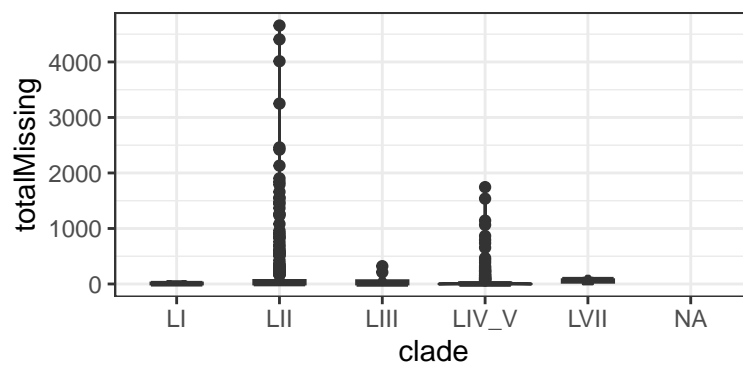


Table 1: Top Frameshifts

frameShifts	n
L:85-87	650
L:451-452	402
L:1663	364
L:1923-1924	341
L:439-442	321
L:1001	319
L:2168-2170	287
L:126-128	276
L:2094-2095	255
L:952-955	245

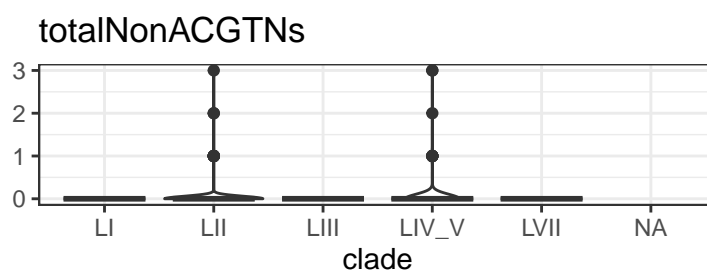
Missing Data

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



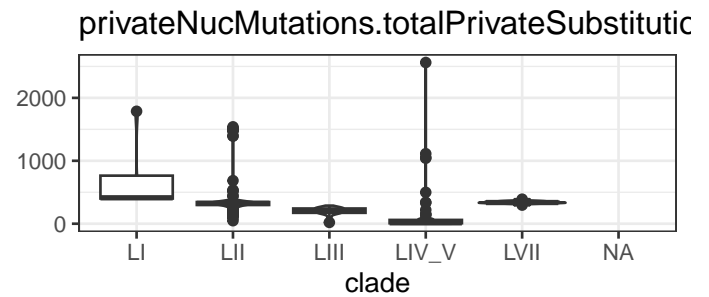
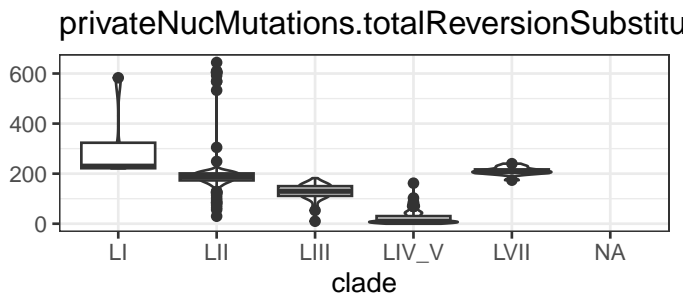
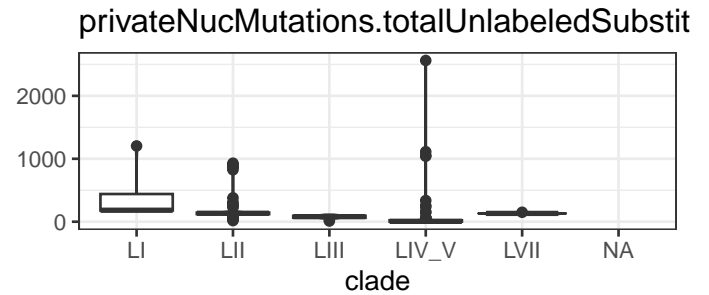
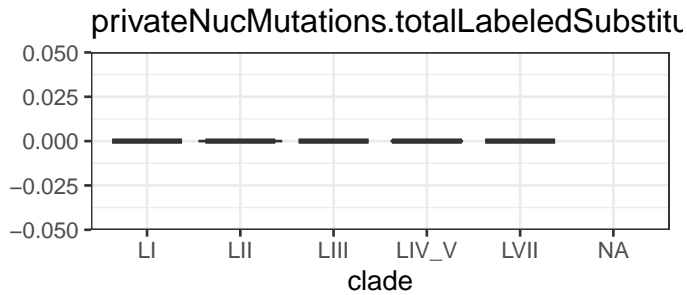
Mixed Sites (M)

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



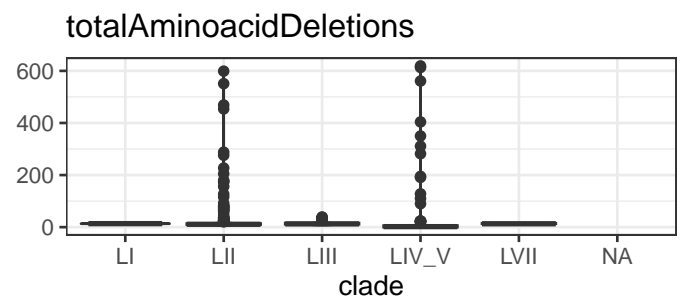
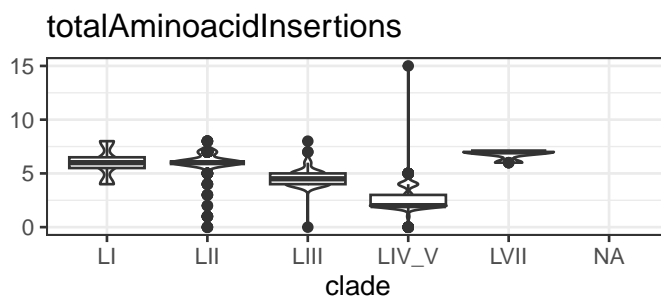
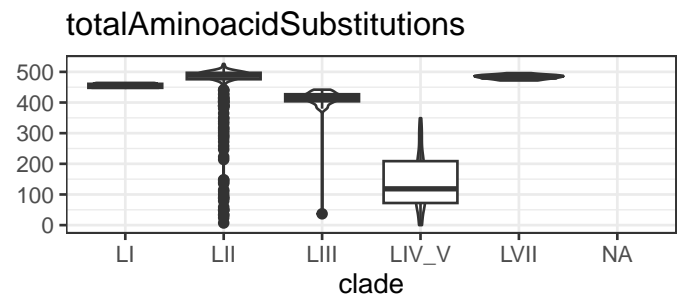
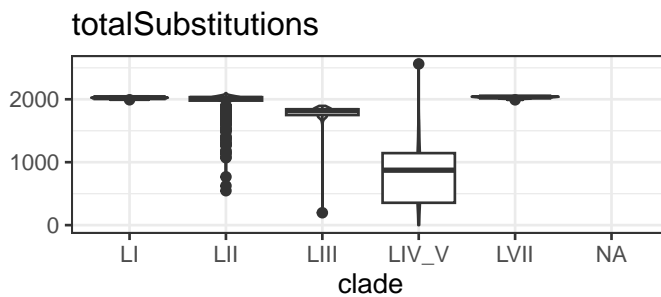
Private Mutations

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



Mutation clusters

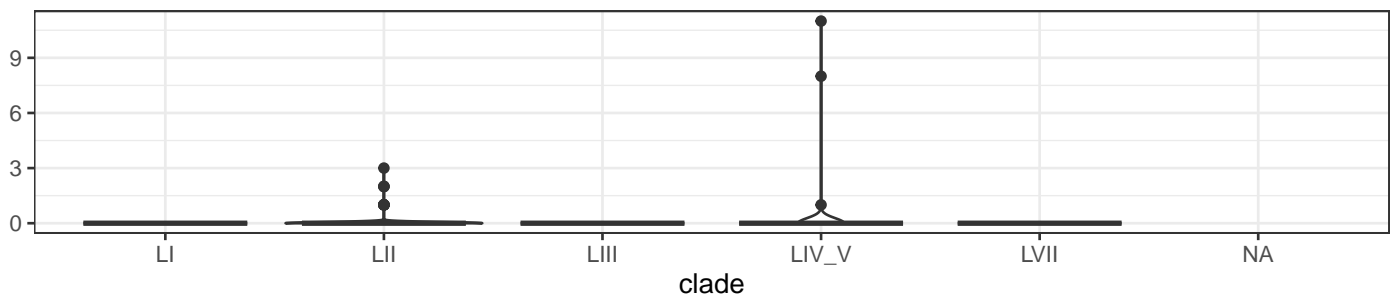
```
"snpClusters": {
  "enabled": true,
  "windowSize": 100,
  "clusterCutOff": 6,
  "scoreWeight": 50
},
```



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

