

COVID-19 subject 100667644

2021-04-17

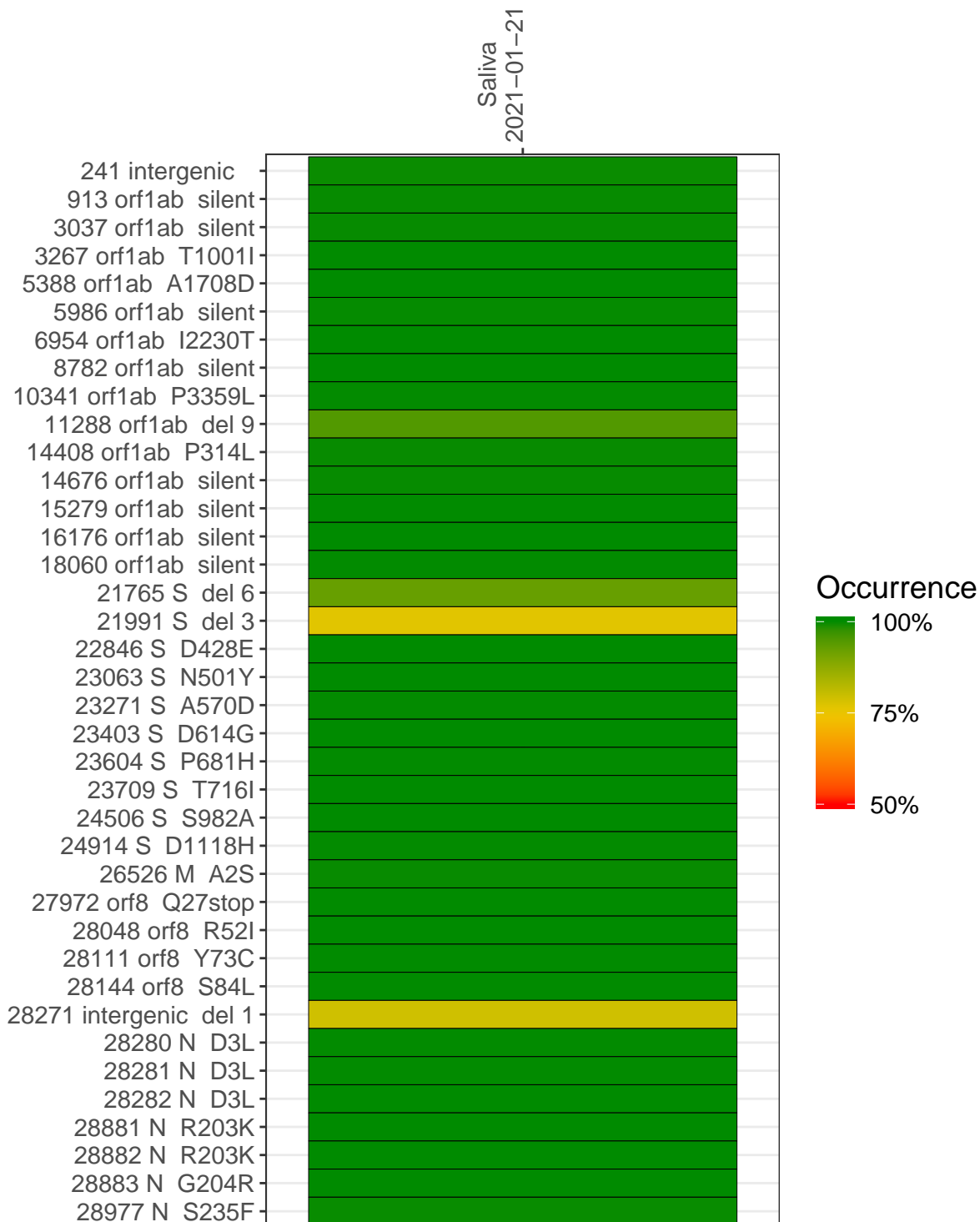
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with $> 90\%$ sequence coverage.

Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (≥ 5 reads)
VSP0623	composite	NA	Saliva	2021-01-21	29.76	B.1.1.7	99.2%	99.2%
VSP0623-1	single experiment	NA	Saliva	2021-01-21	29.65	B.1.1.7	99.2%	99.2%
VSP0623-2	single experiment	NA	Saliva	2021-01-21	15.04	B.1.1.7	99.2%	99.2%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/USA-WA1-2020) are shared across subject samples where the percent variance is colored. Variants are called if a variant position is covered by 5 or more reads, the alternative base is found in > 50% of read pairs and the variant yields a PHRED score > 20. Gray tiles denote positions where the variant was not the major variant or no variants were found. The relative base compositions of each experiment used to calculate tiles are shown in the following plot where the total number of position reads are shown atop of each plot.



Saliva
2021-01-21

241 intergenic	11238	4047
913 orf1ab silent	18390	5963
3037 orf1ab silent	9414	3209
3267 orf1ab T1001I	8937	2821
5388 orf1ab A1708D	7031	2237
5986 orf1ab silent	3621	1049
6954 orf1ab I2230T	3264	1162
8782 orf1ab silent	7894	2320
10341 orf1ab P3359L	11892	3731
11288 orf1ab del 9	11972	3940
14408 orf1ab P314L	14068	4373
14676 orf1ab silent	10148	3340
15279 orf1ab silent	13082	4329
16176 orf1ab silent	9965	3362
18060 orf1ab silent	4329	1363
21765 S del 6	4306	1366
21991 S del 3	1352	398
22846 S D428E	3115	936
23063 S N501Y	2993	1024
23271 S A570D	8322	2657
23403 S D614G	10172	3332
23604 S P681H	8016	2609
23709 S T716I	7174	2932
24506 S S982A	3324	1031
24914 S D1118H	10807	3285
26526 M A2S	3771	1138
27972 orf8 Q27stop	10229	3301
28048 orf8 R52I	7977	2681
28111 orf8 Y73C	9615	3036
28144 orf8 S84L	12054	3978
28271 intergenic del 1	15810	4583
28280 N D3L	12773	3209
28281 N D3L	12773	3209
28282 N D3L	12847	3259
28881 N R203K	932	287
28882 N R203K	930	287
28883 N G204R	930	290
28977 N S235F	343	163

Base change

- Expected
- A
- T
- C
- G
- N
- Ins/Del
- No data

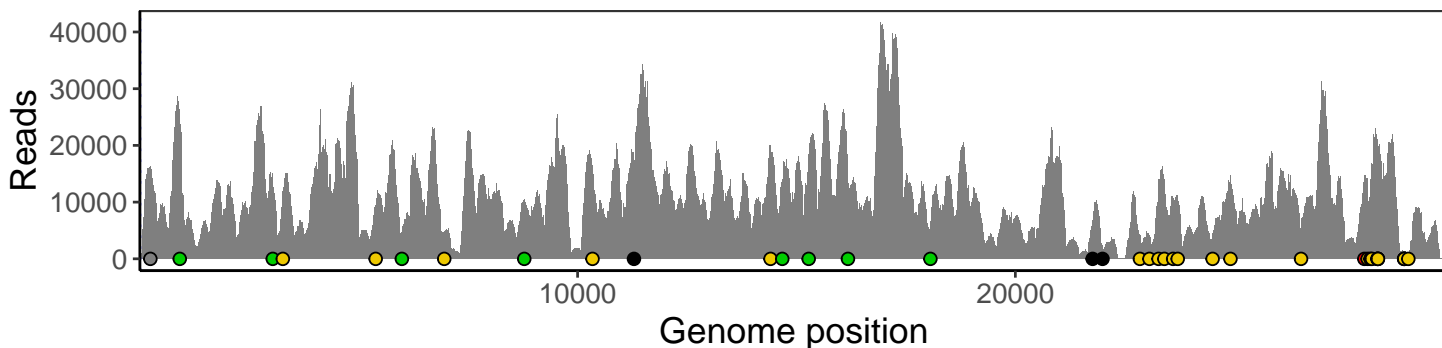
VSP0623-1

VSP0623-2

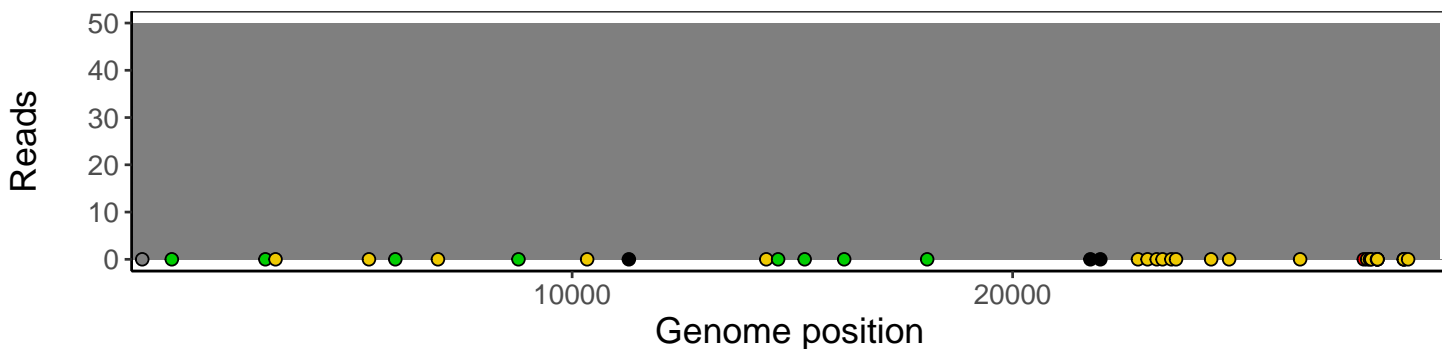
Analyses of individual experiments and composite results

VSP0623 | 2021-01-21 | Saliva | Molpath-Seq2 | composite result

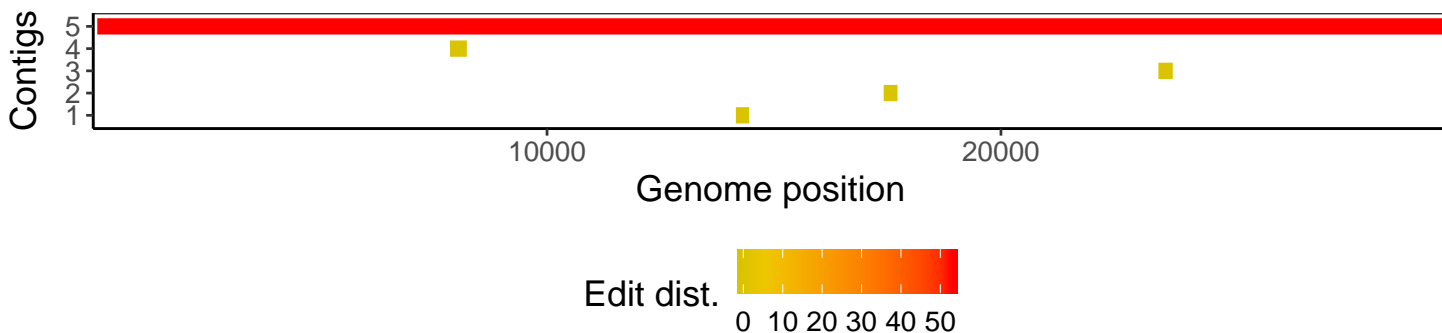
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



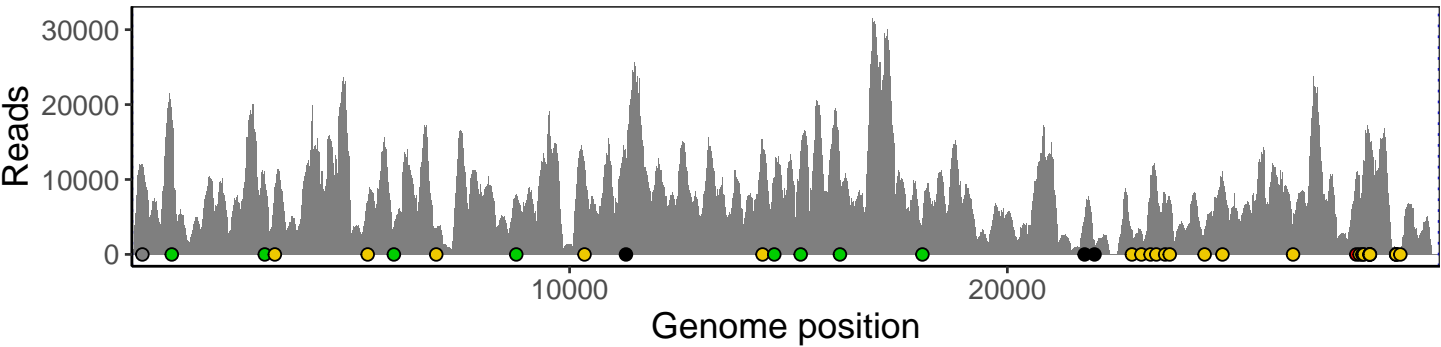
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



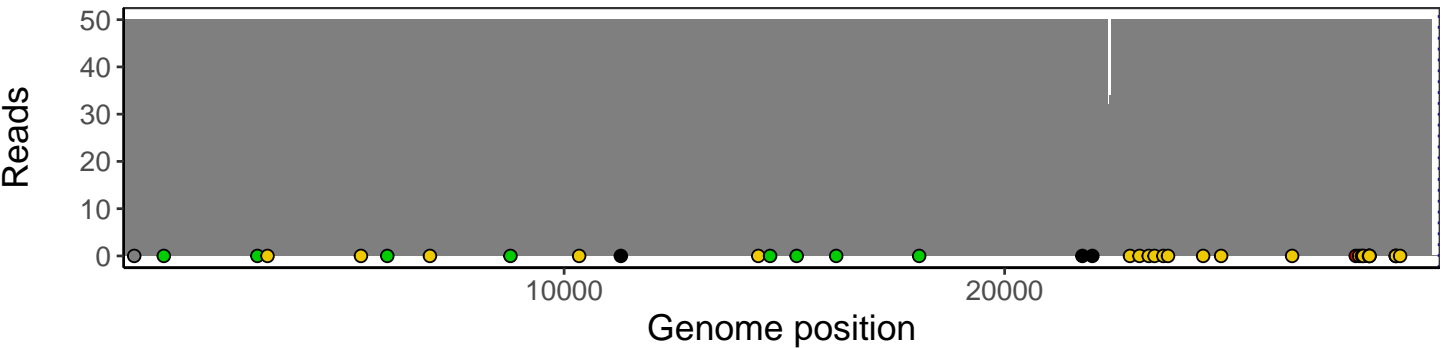
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



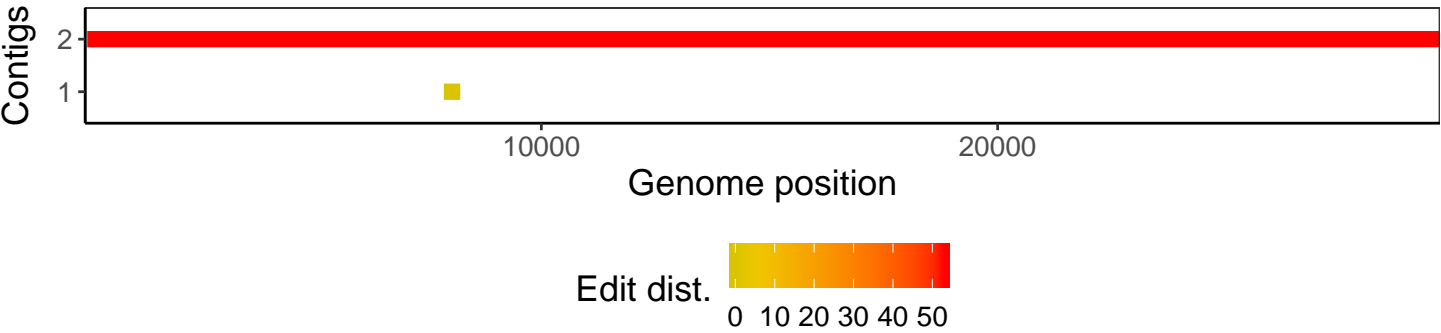
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



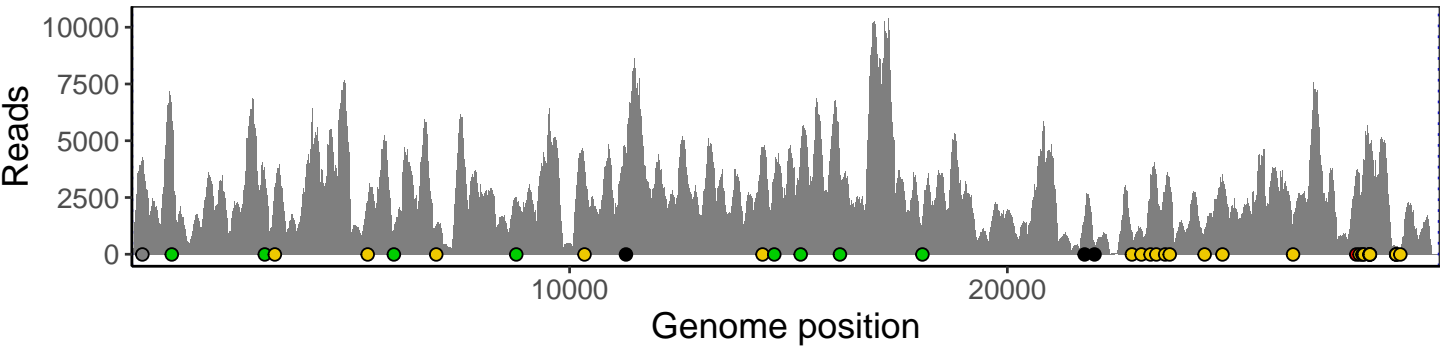
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



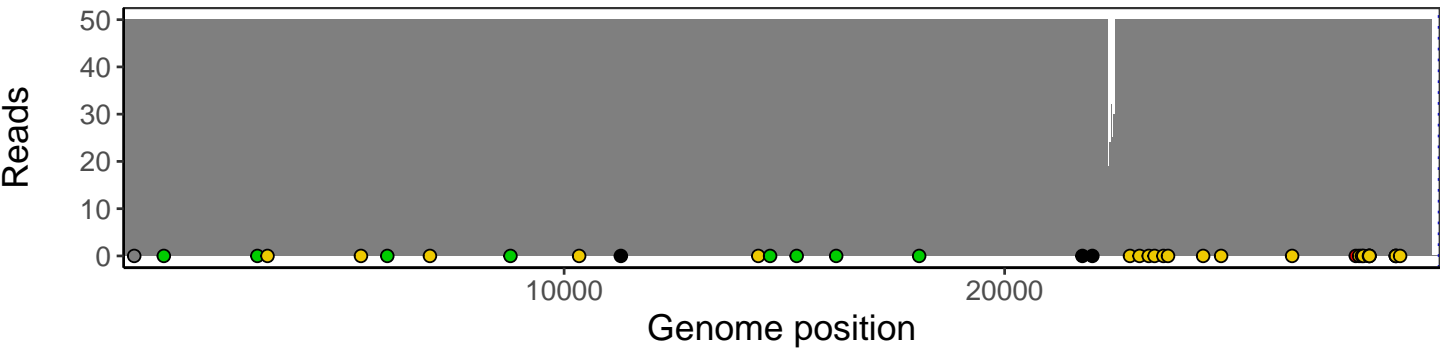
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



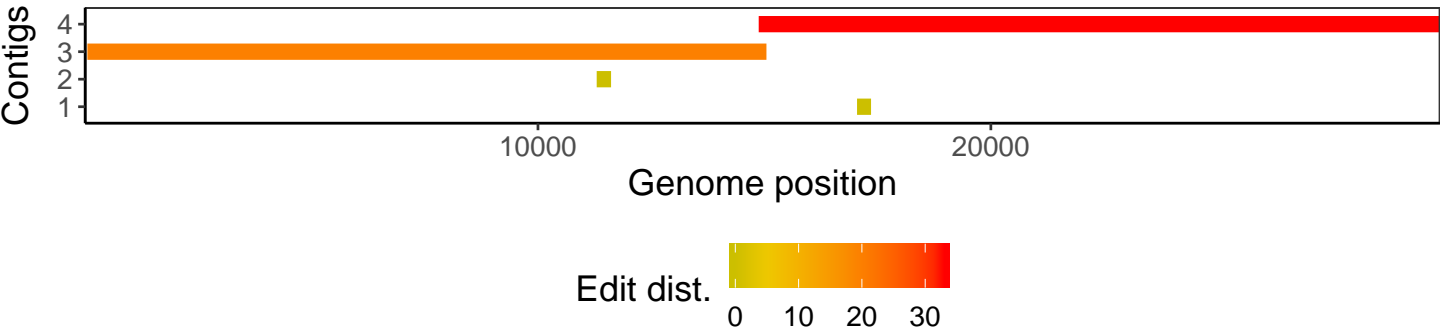
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



Software environment

Software/R package	Version
R	3.4.0
bwa	0.7.17-r1198-dirty
samtools	1.10 Using htlib 1.10
bcftools	1.10.2-34-g1a12af0-dirty Using htlib 1.10.2-57-gf58a6f3
pangolin	2.3.8
genbankr	1.4.0
optparse	1.6.0
forcats	0.3.0
stringr	1.4.0
dplyr	0.8.1
purrr	0.2.5
readr	1.1.1
tidyr	0.8.1
tibble	2.1.2
ggplot2	3.0.0
tidyverse	1.2.1
ShortRead	1.34.2
GenomicAlignments	1.12.2
SummarizedExperiment	1.6.5
DelayedArray	0.2.7
matrixStats	0.54.0
Biobase	2.36.2
Rsamtools	1.28.0
GenomicRanges	1.28.6
GenomeInfoDb	1.12.3
Biostrings	2.44.2
XVector	0.16.0
IRanges	2.10.5
S4Vectors	0.14.7
BiocParallel	1.10.1
BiocGenerics	0.22.1