

# COVID-19 subject 2749

*2021-04-30*

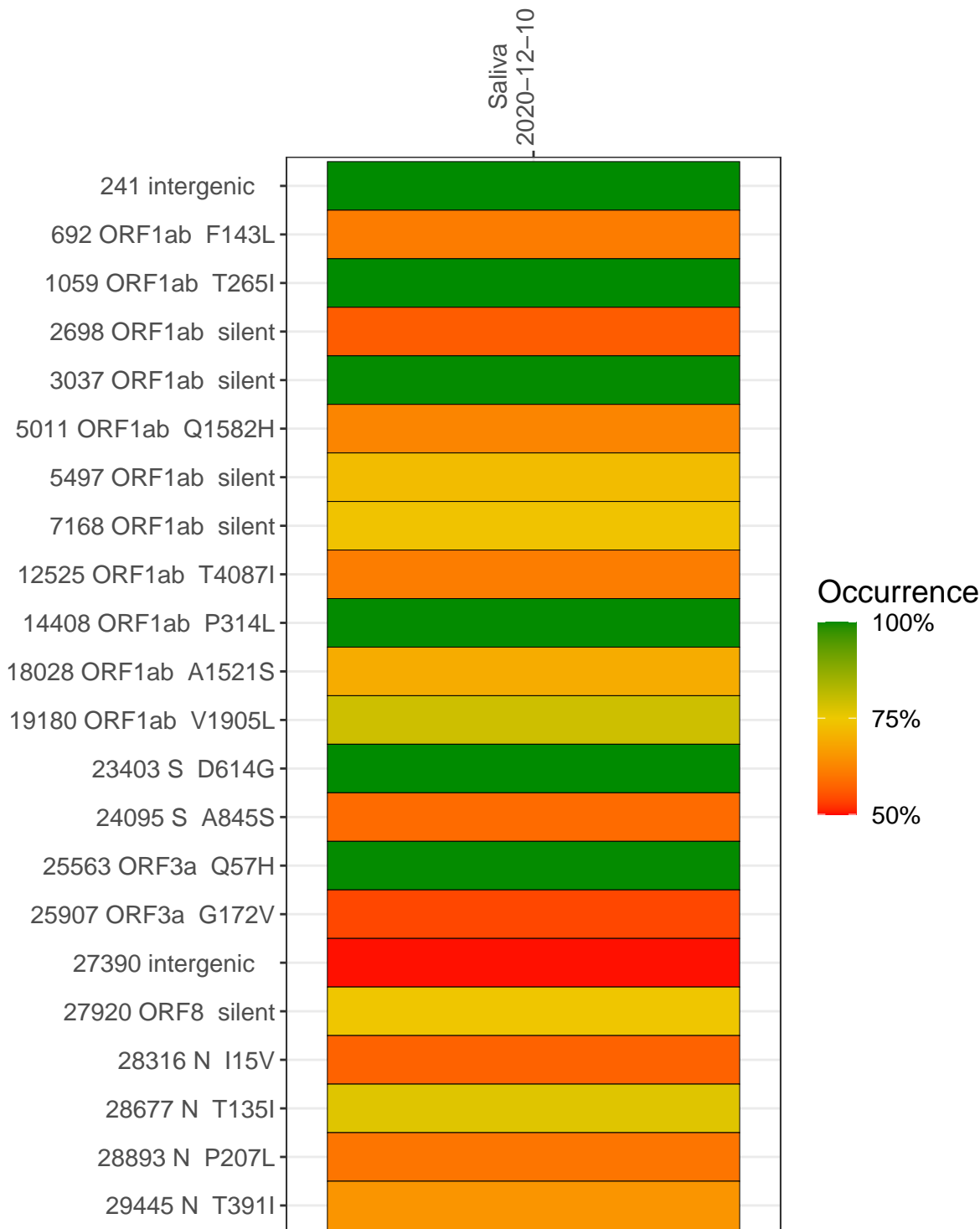
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

Table 1. Sample summary.

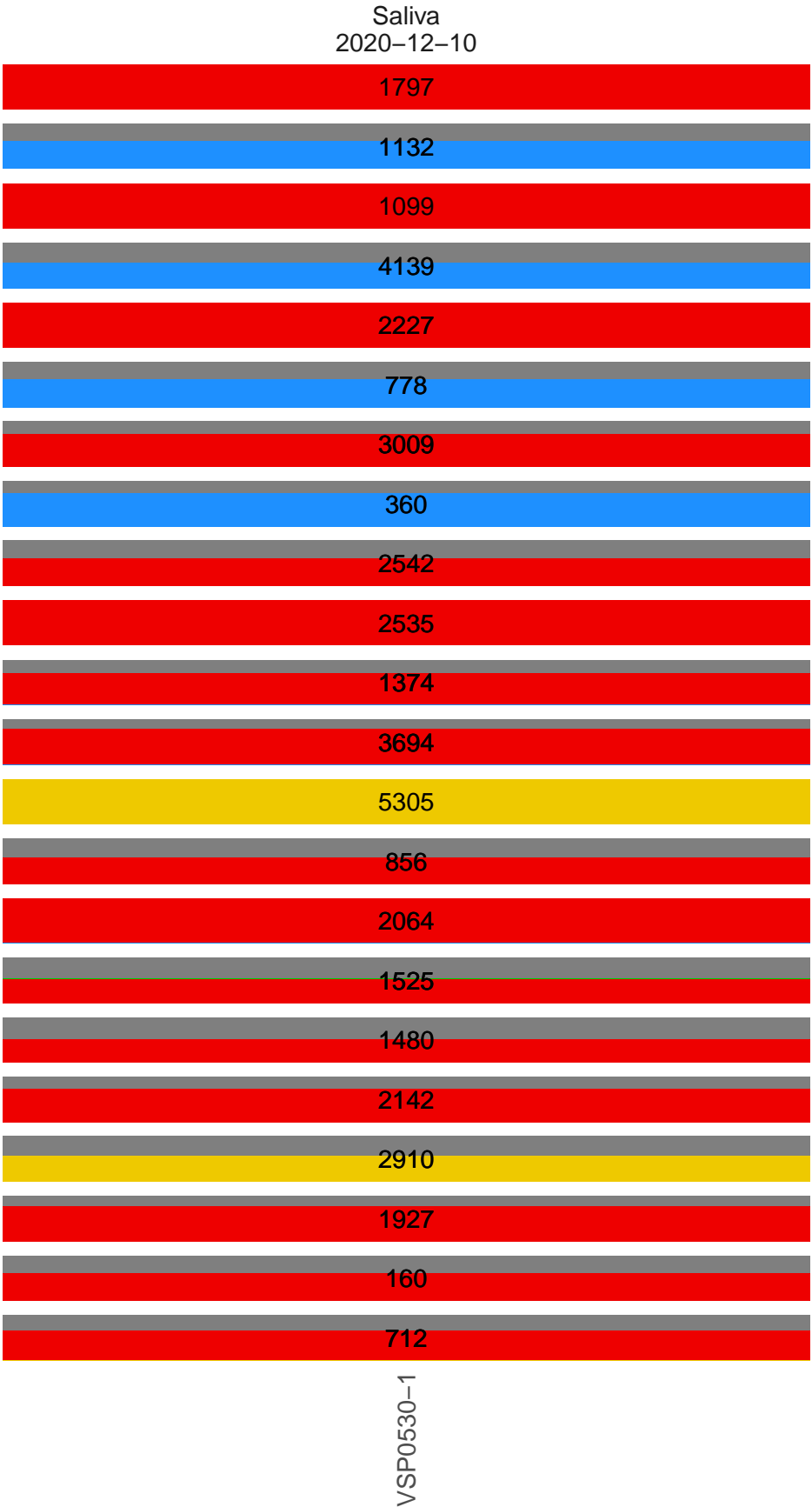
Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage ( $\geq 5$ reads)
VSP0530-1	single experiment	NA	Saliva	2020-12-10	21.64	B.1.311	99.3%	99.3%

## Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/Wuhan-Hu-1) 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.



5 N 3 N 17 N 6 N 2 RF1 inter RF3 RF3a RF3b RF3c RF3d RF3e RF3f RF3g RF3h RF3i RF3j RF3k RF3l RF3m RF3n RF3o RF3p RF3q RF3r RF3s RF3t RF3u RF3v RF3w RF3x RF3y RF3z RF3aa RF3ab RF3ac RF3ad RF3ae RF3af RF3ag RF3ah RF3ai RF3aj RF3ak RF3al RF3am RF3an RF3ao RF3ap RF3aq RF3ar RF3as RF3at RF3au RF3av RF3aw RF3ax RF3ay RF3az RF3ba RF3bb RF3bc RF3bd RF3be RF3bf RF3bg RF3bh RF3bi RF3bj RF3bk RF3bl RF3bm RF3bn RF3bo RF3bp RF3bq RF3br RF3bs RF3bt RF3bu RF3bv RF3bw RF3bx RF3by RF3bz RF3ca RF3cb RF3cc RF3cd RF3ce RF3cf RF3cg RF3ch RF3ci RF3cj RF3ck RF3cl RF3cm RF3cn RF3co RF3cp RF3cq RF3cr RF3cs RF3ct RF3cu RF3cv RF3cw RF3cx RF3cy RF3cz RF3da RF3db RF3dc RF3dd RF3de RF3df RF3dg RF3dh RF3di RF3dj RF3dk RF3dl RF3dm RF3dn RF3do RF3dp RF3dq RF3dr RF3ds RF3dt RF3du RF3dv RF3dw RF3dx RF3dy RF3dz RF3ea RF3eb RF3ec RF3ed RF3ee RF3ef RF3eg RF3eh RF3ei RF3ej RF3ek RF3el RF3em RF3en RF3eo RF3ep RF3eq RF3er RF3es RF3et RF3eu RF3ev RF3ew RF3ex RF3ey RF3ez RF3fa RF3fb RF3fc RF3fd RF3fe RF3ff RF3fg RF3fh RF3fi RF3fj RF3fk RF3fl RF3fm RF3fn RF3fo RF3fp RF3fq RF3fr RF3fs RF3ft RF3fu RF3fv RF3fw RF3fx RF3fy RF3fz RF3ga RF3gb RF3gc RF3gd RF3ge RF3gf RF3gg RF3gh RF3gi RF3gj RF3gk RF3gl RF3gm RF3gn RF3go RF3gp RF3gq RF3gr RF3gs RF3gt RF3gu RF3gv RF3gw RF3gx RF3gy RF3gz RF3ha RF3hb RF3hc RF3hd RF3he RF3hf RF3hg RF3hi RF3hj RF3hk RF3hl RF3hm RF3hn RF3ho RF3hp RF3hq RF3hr RF3hs RF3ht RF3hu RF3hv RF3hw RF3hx RF3hy RF3hz RF3ia RF3ib RF3ic RF3id RF3ie RF3if RF3ig RF3ih RF3ii RF3ij RF3ik RF3il RF3im RF3in RF3io RF3ip RF3iq RF3ir RF3is RF3it RF3iu RF3iv RF3iw RF3ix RF3iy RF3iz RF3ja RF3jb RF3jc RF3jd RF3je RF3jf RF3jg RF3jh RF3ji RF3jj RF3jk RF3jl RF3jm RF3jn RF3jo RF3jp RF3jq RF3jr RF3js RF3jt RF3ju RF3jv RF3jw RF3jx RF3jy RF3jz RF3ka RF3kb RF3kc RF3kd RF3ke RF3kf RF3kg RF3kh RF3ki RF3kj RF3kk RF3kl RF3km RF3kn RF3ko RF3kp RF3kq RF3kr RF3ks RF3kt RF3ku RF3kv RF3kw RF3kx RF3ky RF3kz RF3la RF3lb RF3lc RF3ld RF3le RF3lf RF3lg RF3lh RF3li RF3lj RF3lk RF3ll RF3lm RF3ln RF3lo RF3lp RF3lq RF3lr RF3ls RF3lt RF3lu RF3lv RF3lw RF3lx RF3ly RF3lz RF3ma RF3mb RF3mc RF3md RF3me RF3mf RF3mg RF3mh RF3mi RF3mj RF3mk RF3ml RF3mm RF3mn RF3mo RF3mp RF3mq RF3mr RF3ms RF3mt RF3mu RF3mv RF3mw RF3mx RF3my RF3mz RF3na RF3nb RF3nc RF3nd RF3ne RF3nf RF3ng RF3nh RF3ni RF3nj RF3nk RF3nl RF3nm RF3nn RF3no RF3np RF3nq RF3nr RF3ns RF3nt RF3nu RF3nv RF3nw RF3nx RF3ny RF3nz RF3oa RF3ob RF3oc RF3od RF3oe RF3of RF3og RF3oh RF3oi RF3oj RF3ok RF3ol RF3om RF3on RF3oo RF3op RF3oq RF3or RF3os RF3ot RF3ou RF3ov RF3ow RF3ox RF3oy RF3oz RF3pa RF3pb RF3pc RF3pd RF3pe RF3pf RF3pg RF3ph RF3pi RF3pj RF3pk RF3pl RF3pm RF3pn RF3po RF3pp RF3pq RF3pr RF3ps RF3pt RF3pu RF3pv RF3pw RF3px RF3py RF3pz RF3qa RF3qb RF3qc RF3qd RF3qe RF3qf RF3qg RF3qh RF3qi RF3qj RF3qk RF3ql RF3qm RF3qn RF3qo RF3qp RF3qq RF3qr RF3qs RF3qt RF3qu RF3qv RF3qw RF3qx RF3qy RF3qz RF3ra RF3rb RF3rc RF3rd RF3re RF3rf RF3rg RF3rh RF3ri RF3rj RF3rk RF3rl RF3rm RF3rn RF3ro RF3rp RF3rq RF3rr RF3rs RF3rt RF3ru RF3rv RF3rw RF3rx RF3ry RF3rz RF3sa RF3sb RF3sc RF3sd RF3se RF3sf RF3sg RF3sh RF3si RF3sj RF3sk RF3sl RF3sm RF3sn RF3so RF3sp RF3sq RF3sr RF3ss RF3st RF3su RF3sv RF3sw RF3sx RF3sy RF3sz RF3ta RF3tb RF3tc RF3td RF3te RF3tf RF3tg RF3th RF3ti RF3tj RF3tk RF3tl RF3tm RF3tn RF3to RF3tp RF3tq RF3tr RF3ts RF3tu RF3tv RF3tw RF3tx RF3ty RF3tz RF3ua RF3ub RF3uc RF3ud RF3ue RF3uf RF3ug RF3uh RF3ui RF3uj RF3uk RF3ul RF3um RF3un RF3uo RF3up RF3uq RF3ur RF3us RF3ut RF3uu RF3uv RF3uw RF3ux RF3uy RF3uz RF3va RF3vb RF3vc RF3vd RF3ve RF3vf RF3vg RF3vh RF3vi RF3vj RF3vk RF3vl RF3vm RF3vn RF3vo RF3vp RF3vq RF3vr RF3vs RF3vt RF3vu RF3vv RF3vw RF3vx RF3vy RF3vz RF3wa RF3wb RF3wc RF3wd RF3we RF3wf RF3wg RF3wh RF3wi RF3wj RF3wk RF3wl RF3wm RF3wn RF3wo RF3wp RF3wq RF3wr RF3ws RF3wt RF3wu RF3wv RF3ww RF3wx RF3wy RF3wz RF3xa RF3xb RF3xc RF3xd RF3xe RF3xf RF3xg RF3xh RF3xi RF3xj RF3xk RF3xl RF3xm RF3xn RF3xo RF3xp RF3xq RF3xr RF3xs RF3xt RF3xu RF3xv RF3xw RF3xx RF3xy RF3xz RF3ya RF3yb RF3yc RF3yd RF3ye RF3yf RF3yg RF3yh RF3yi RF3yj RF3yk RF3yl RF3ym RF3yn RF3yo RF3yp RF3yq RF3yr RF3ys RF3yt RF3yu RF3yv RF3yw RF3yx RF3yy RF3yz RF3za RF3zb RF3zc RF3zd RF3ze RF3zf RF3zg RF3zh RF3zi RF3zj RF3zk RF3zl RF3zm RF3zn RF3zo RF3zp RF3zq RF3zr RF3zs RF3zt RF3zu RF3zv RF3zw RF3zx RF3zy RF3zz



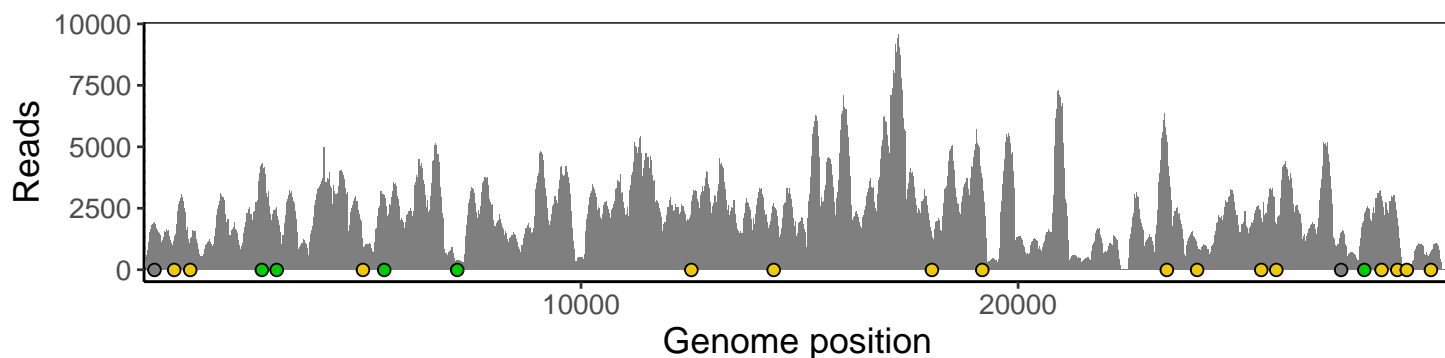
Base change

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

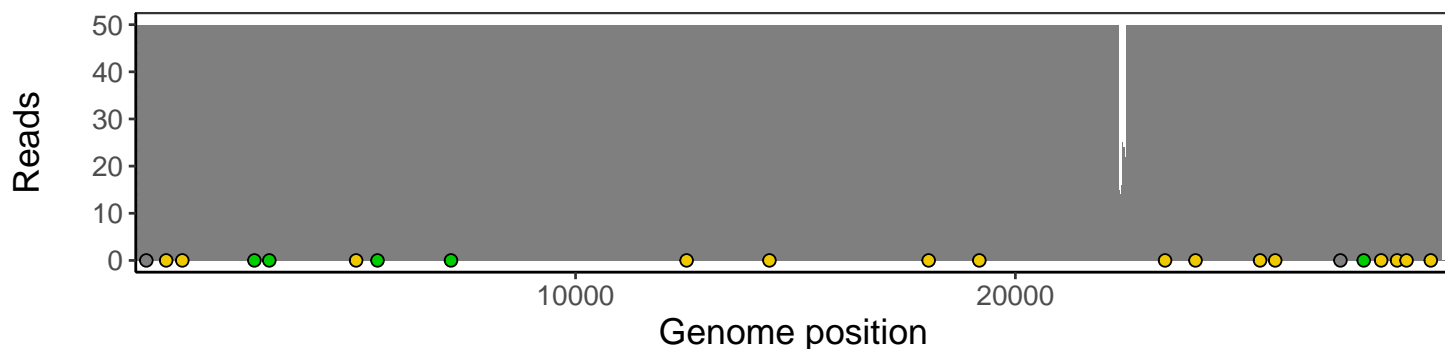
## Analyses of individual experiments and composite results

VSP0530-1 | 2020-12-10 | Saliva | 2749 | genomes | single experiment

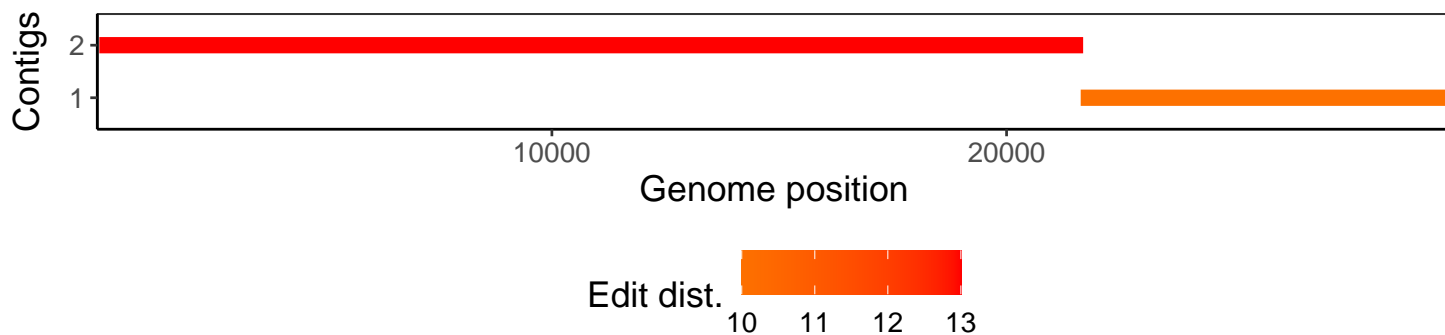
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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