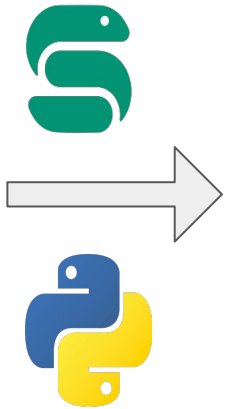
Live Monitoring of MinION Sequencing Runs

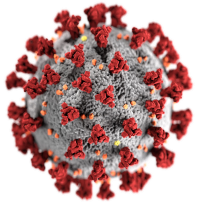
(Monitorovanie MinION sekvenovania)

Jana Černíková
doc. Mgr. Tomáš Vinař, PhD.

https://nanoporetech.com/products/minion

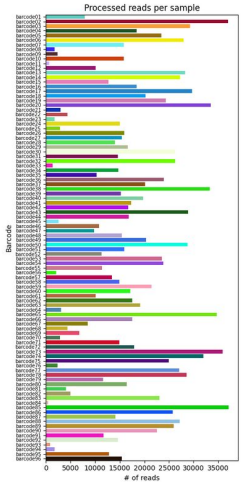


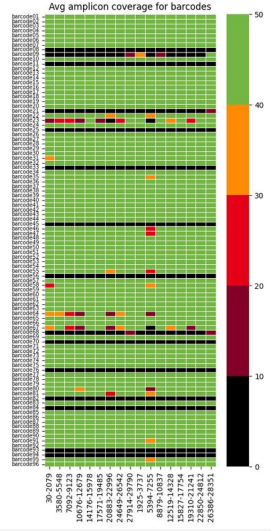
β γ α δ

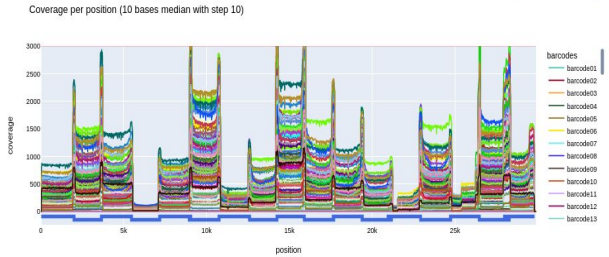


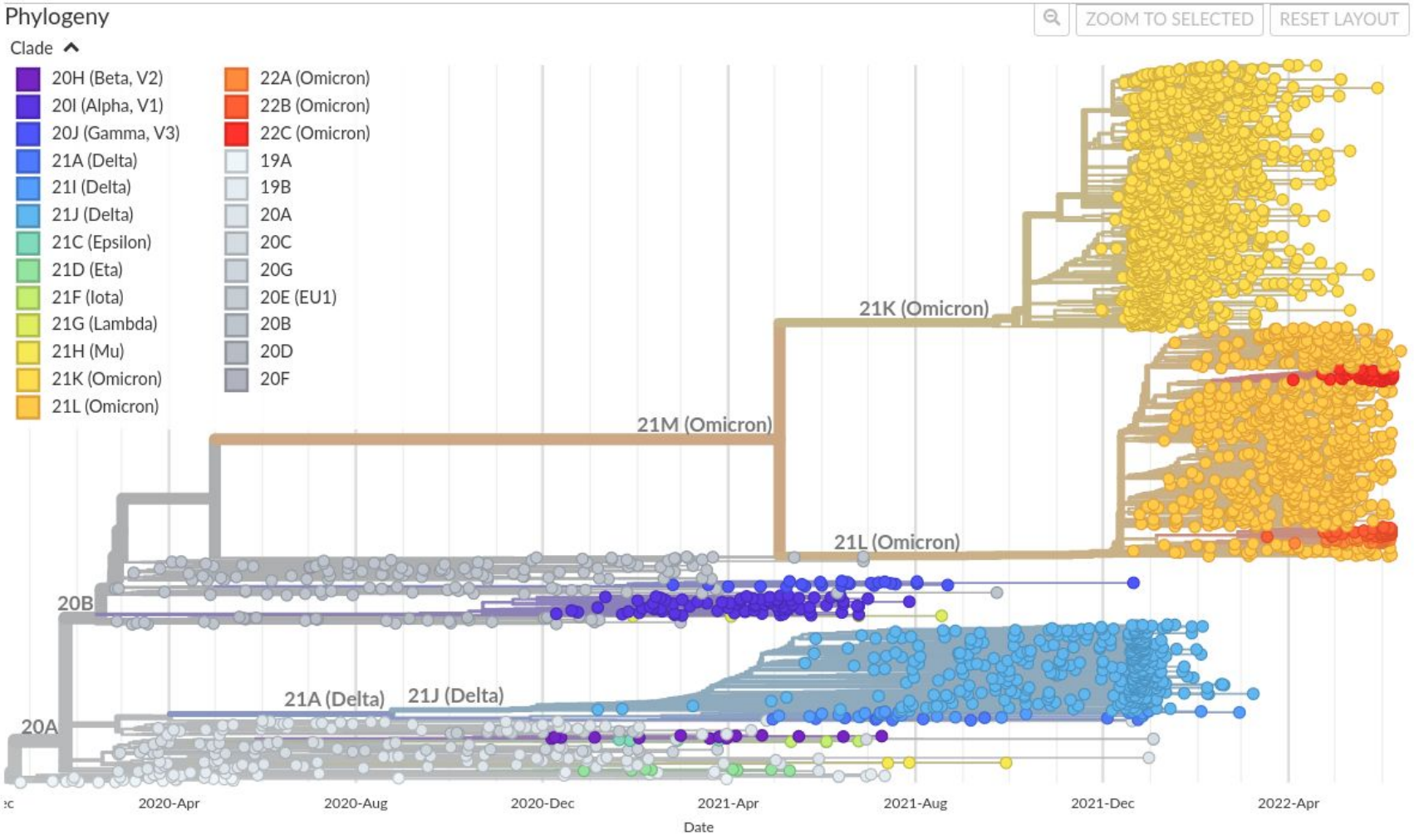
XX.YY.WW.ZZ

o





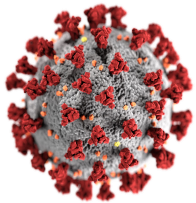




https://nextstrain.org/ncov/gisaid/global

https://phil.cdc.gov/Details.aspx?pid=23312

Skôr než začneme sekvenovať..

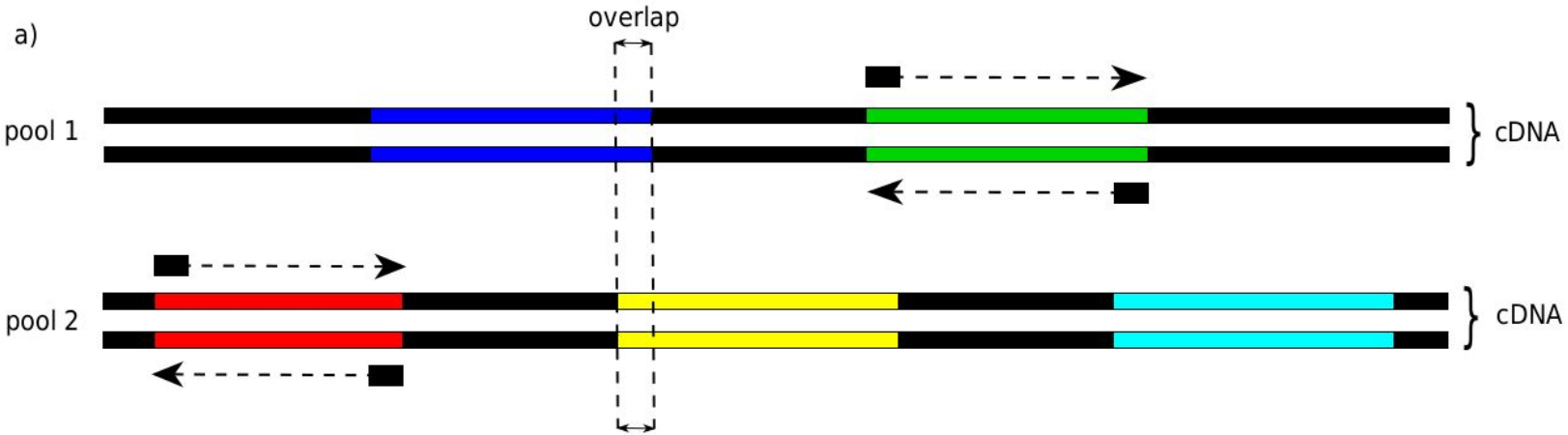


- odobratie vzorky od pacienta

Laboratórium:

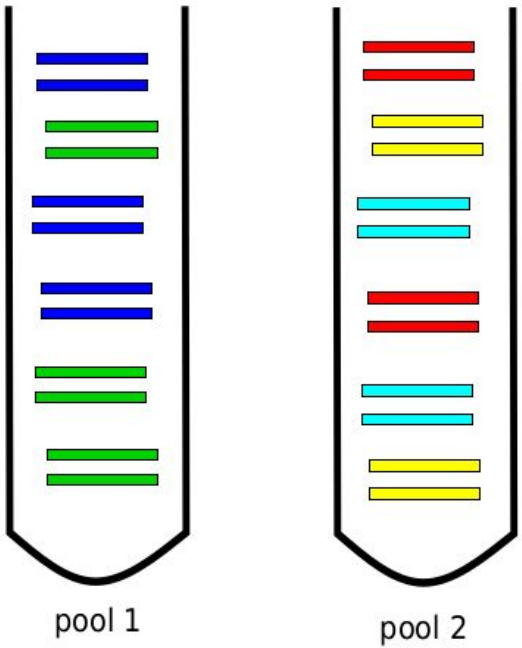
- RNA -> cDNA
- **amplifikácia pomocou tiled PCR**
- **pripojenie barkódových sekvencií**
- pripojenie sekvenovacích adaptérov (sequencing adapters)
- **sekvenovanie**

Tiled PCR



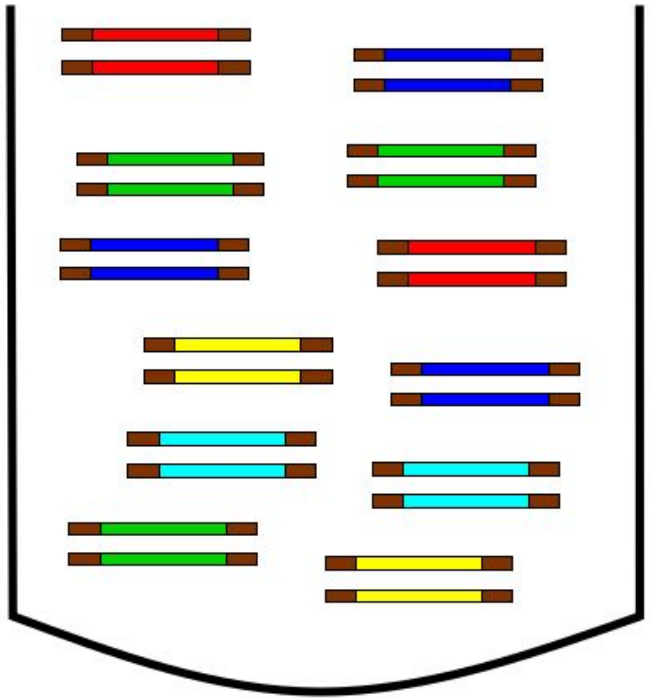
Tiled PCR

b)

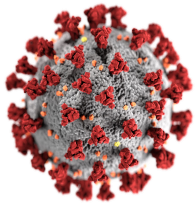


Barkódovanie

c)



Skôr než začneme sekvenovať..

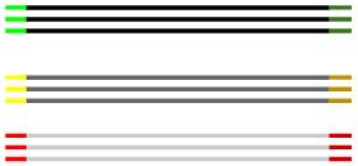


- odobratie vzorky od pacienta

Laboratórium:

- RNA -> cDNA
- **amplifikácia pomocou tiled PCR**
- **pripojenie barkódových sekvencií**
- pripojenie sekvenovacích adaptérov (sequencing adapters)
- **sekvenovanie**

Sekvenovanie

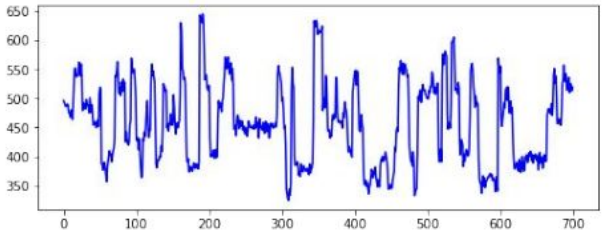


V. Boža et al.: Alternative base callers aid real-time analysis of SARS-CoV-2 sequencing runs

+

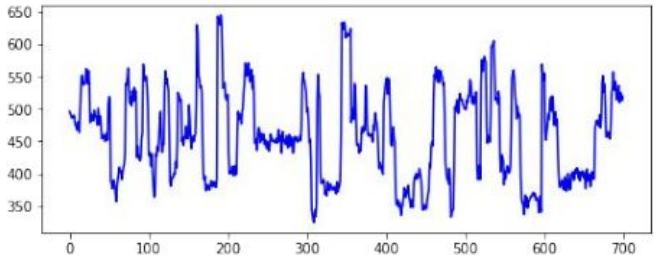
https://nanoporetech.com/products/minion

=



V. Boža et al.: Alternative base callers aid real-time analysis of SARS-CoV-2 sequencing runs

Base calling

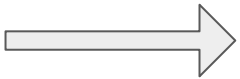


AAAGTAGATCTAAAGCTTACAATTT
GCGATCGTTGACTGATCGTAGCGC
TATGTACTGTTTAGTACGGATCGCG
AAACTGATGTAGCTTGAATGCTTTT
GCGACTTGCTCTTTAGTCAGGCGC
TATATGCTGTGTAGTGCTAGTAGCG

V. Boža et al.: Alternative base callers aid real-time analysis of SARS-CoV-2 sequencing runs

Debarkóding

AAAGTAGATCTAAAGCTTACAATTT
GCGATCGTTGACTGATCGTAGCGC
TATGTACTGTTTAGTACGGATCGCG
AAACTGATGTAGCTTGAATGCTTTT
GCGACTTGCTCTTTAGTCAGGCGC
TATATGCTGTGTAGTGCTAGTAGCG



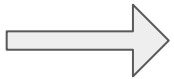
AAA GTAGATCTAAAGCTTACAA TTT
AAA CTGATGTAGCTTGAATGCT TTT

GCG ATCGTTGACTGATCGTAG CGC
GCG ACTTGCTCTTTAGTCAGG CGC

TAT GTACTGTTTAGTACGGATC GCG
TAT ATGCTGTGTAGTGCTAGTA GCG

Zarovnanie

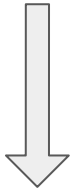
AAA **AGGTGCCACTACATGTGTT** TTT
AAA **TTACCCCAAAATGCTGTTG** TTT

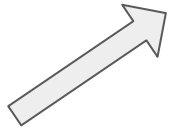


```
AGGTGCCACTACTTGTGGTTACTTACCCCAAAATGCTGTTGTTAAAATTTATTGTCCAGC
AGGTGCCACTACATGTGGTTACTTACCCCAAAA
 GGTGCCACTACATGTGGTTACTTACCCCAAAAT
  GTGCCACTACTTGTGGTTACTTACCCCAA
 GGTGCCACTACATGTGGTTACTTACCCCAAAA
  GTGCCACTACATGTGGTTACTTACCCCAAAA
                          TTACCCCAAAATGCTGTTGTT-AAATTTATTGTCCAGC
                           TACCCCAAAATGCTGTTGTT-AAATTTATTGTCCAG
                         CTTACCCCAAAATGCTGTTGTT-AAATTTATTGTCC
                          TTACCCCAAAATGCTGTTGTT-AAATTTATTGTCCAG
                            ACCCCAAAATGCTGTTGTT-AAATTTATTGTCCAGC
```

V. Boža et al.: Alternative base callers aid real-time analysis of SARS-CoV-2 sequencing runs

Detekcia mutácií





AGGTGCCACTACATGTGGTTACTTACCCCAAAATGCTGTTGTT-AAATTTATTGTCCAGC

zoznam mutácií:

snp, same_as_ref, mutated
C241T,4,81
G376T,2,103
T514C,9,90
C913T,17,79
C3037T,8,34
C3267T,4,63
C5388A,6,158
C5944T,0,7
C5986T,1,6
.
.
.

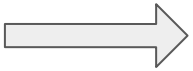
V. Boža et al.: Alternative base callers aid real-time analysis of SARS-CoV-2 sequencing runs

Detekcia variantov

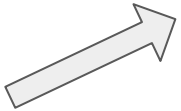
zoznam mutácií:

(pre jeden konkrétny barkód)

snp, same_as_ref,
mutated
C241T,4,81
G376T,2,103
T514C,9,90
C913T,17,79
C3037T,8,34
C3267T,4,63
.
.



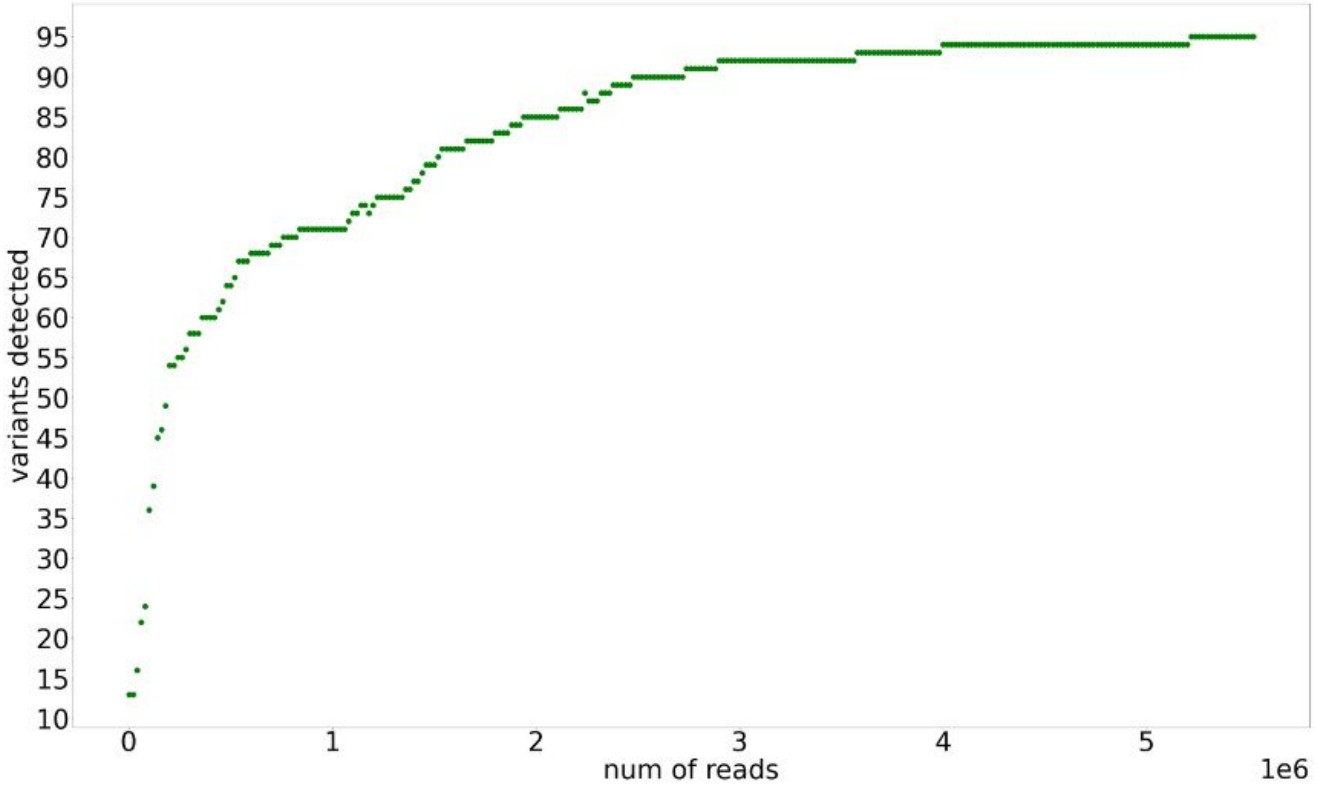
variant **Omicron (B.1.1.529.X)**



variant definovaný v konfiguračnom súbore:

Omicron 7 A2832G T5386G G8393A C10029T C10449A A11537G T13195C C15240T A18163G

Určovanie variantov v závislosti od množstva dát



Zhrnutie

- laboratórium: tiled PCR, vzorky označené barkódmi, sekvenovanie

- bioinformatická analýza
  - base calling
  - debarkóding
  - zarovnanie
  - detekcia mutácií
  - detekcia variantov

Problém

.fast5

Veľa dát v krátkom čase

jeden nový .fast5 súbor každých ~ 5 sekúnd
rádovo 10tky Gb dát na jeden sekvenačný beh

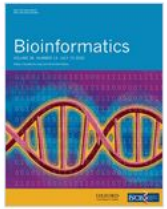
chceme výsledky v reálnom čase

Riešenia?

Časté problémy existujúcich riešení

- potreba podpory GPU, inak pomalý base calling zdržiava ďalšie kroky analýzy

- problém so spracovaním veľkého množstva dát



Volume 36, Issue 14
15 July 2020

Article Contents

DeepNano-blitz: a fast base caller for MinION nanopore sequencers

FREE

Vladimír Boža, Peter Perešíni, Broňa Brejová, Tomáš Vinař ✉ Author Notes

Bioinformatics, Volume 36, Issue 14, 15 July 2020, Pages 4191–4192,
https://doi.org/10.1093/bioinformatics/btaa297
Published: 06 May 2020 **Article history** ▾

Abstract

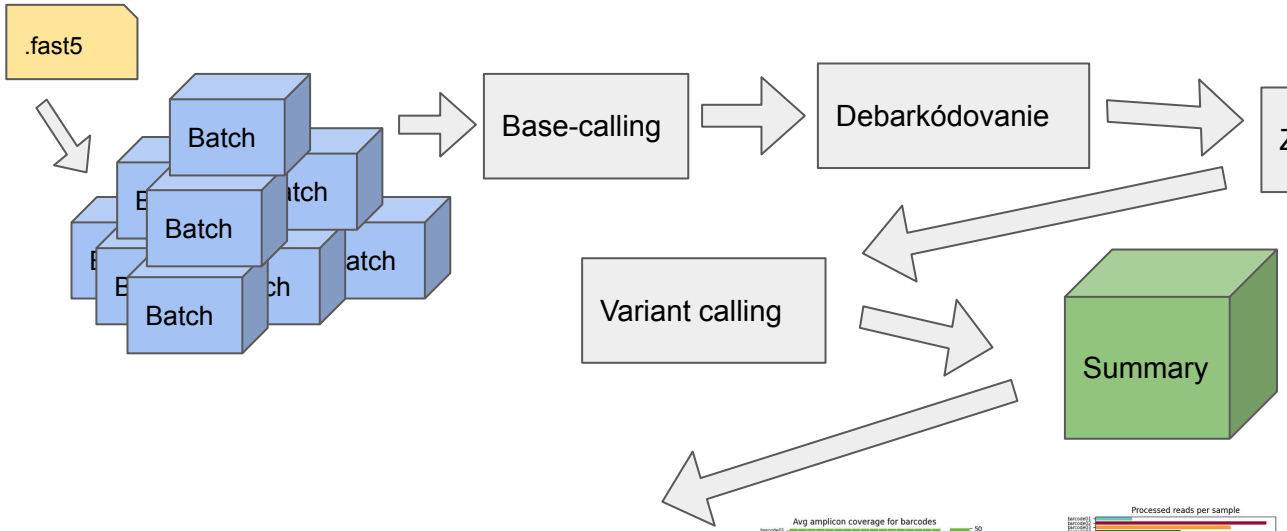
Motivation

Oxford Nanopore MinION is a portable DNA sequencer that is marketed as a device that can be deployed anywhere. Current base callers, however, require a powerful GPU to analyze data produced by MinION in real time, which hampers field applications.

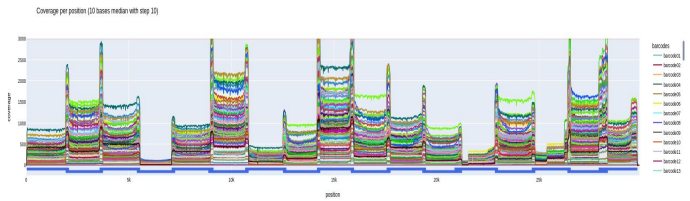
Results

We have developed a fast base caller DeepNano-blitz that can analyze stream from up to two MinION runs in real time using a common laptop CPU (i7-7700HQ), with no GPU requirements. The base caller settings

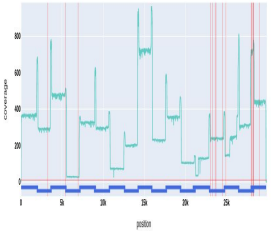
Naše riešenie

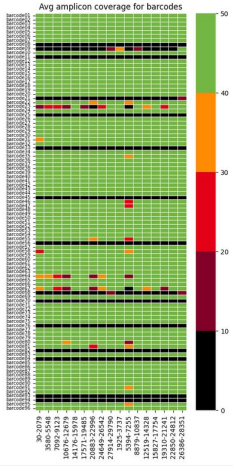


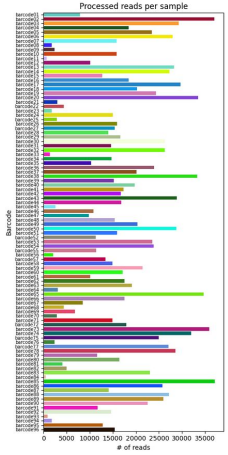
Pipeline



variant	#samples	samples
B.1.1.7	78	barcode01 , barcode03 , barcode04 , barcode05 , barcode26 , barcode27 , barcode28 , barcode29 , barcode51 , barcode52 , barcode53 , barcode54 , barcode74 , barcode75 , barcode77 , barcode78 ,
test	1	barcode01
B.1.258	5	barcode14 , barcode38 , barcode39 , barcode40 ,
B.1.351	1	barcode96





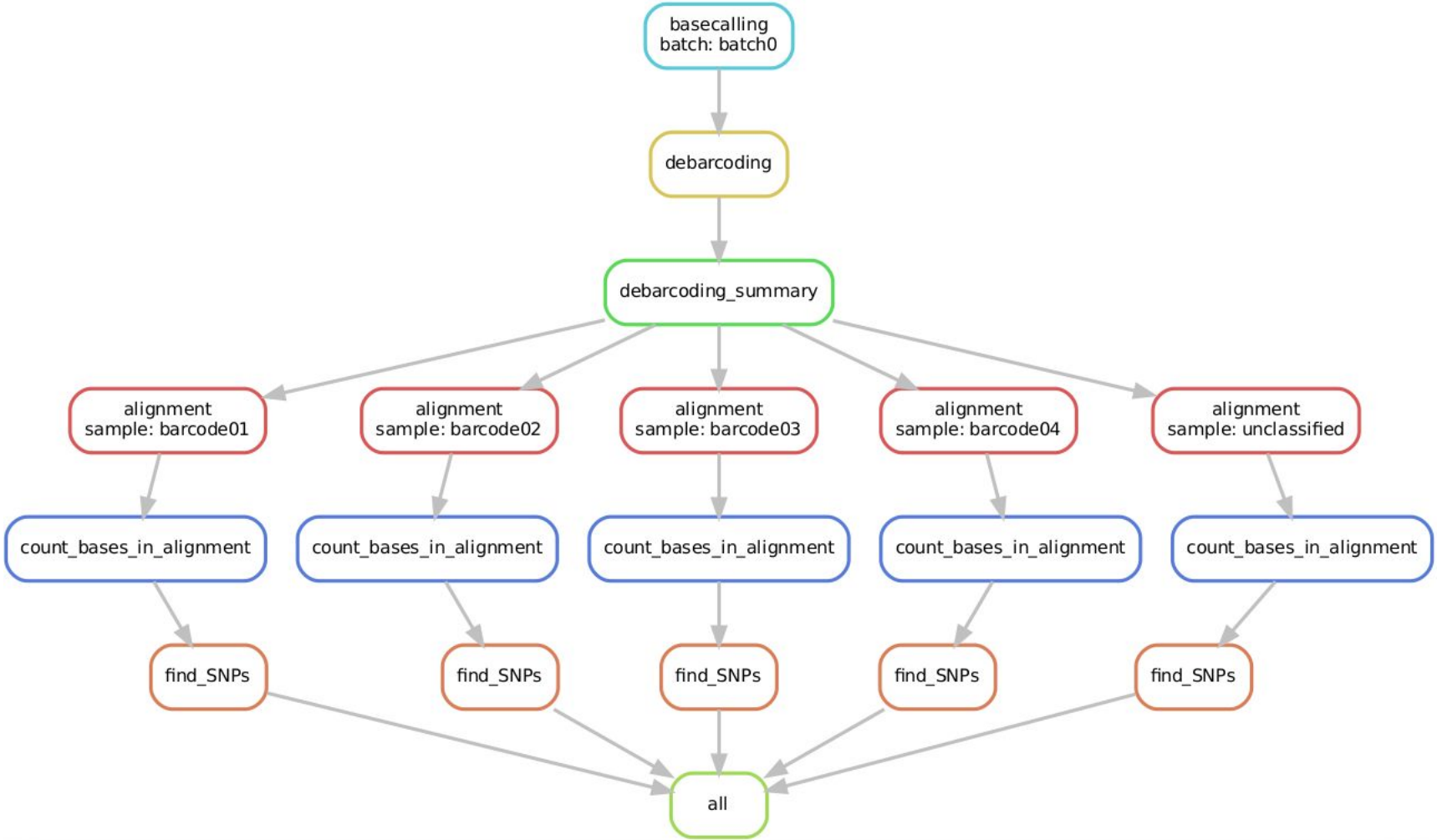


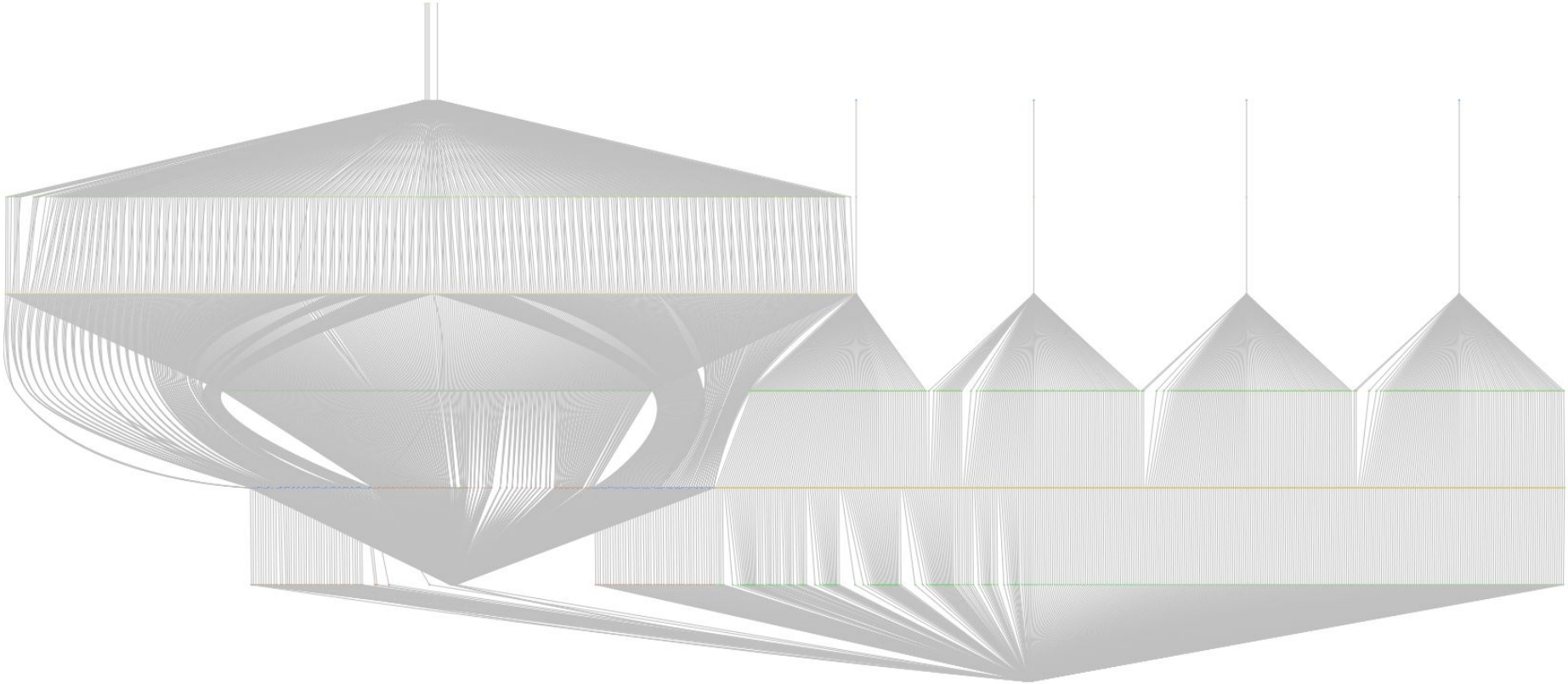
UI

JS, ajax, jinja..

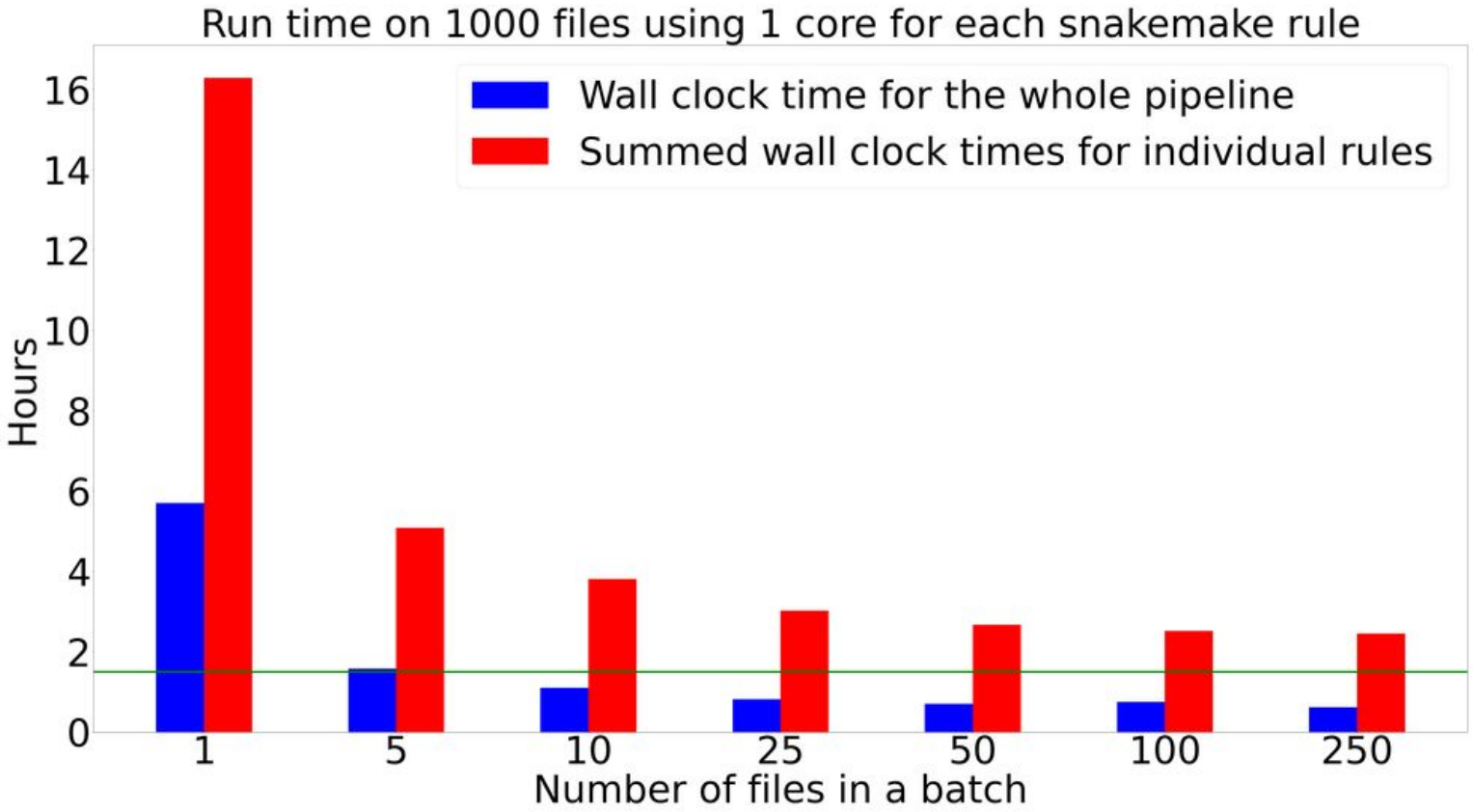
Snakemake

- nástroj na výrobu reprodukovateľných a škálovateľných analýz
- pravidlá -> workflow
- DAG

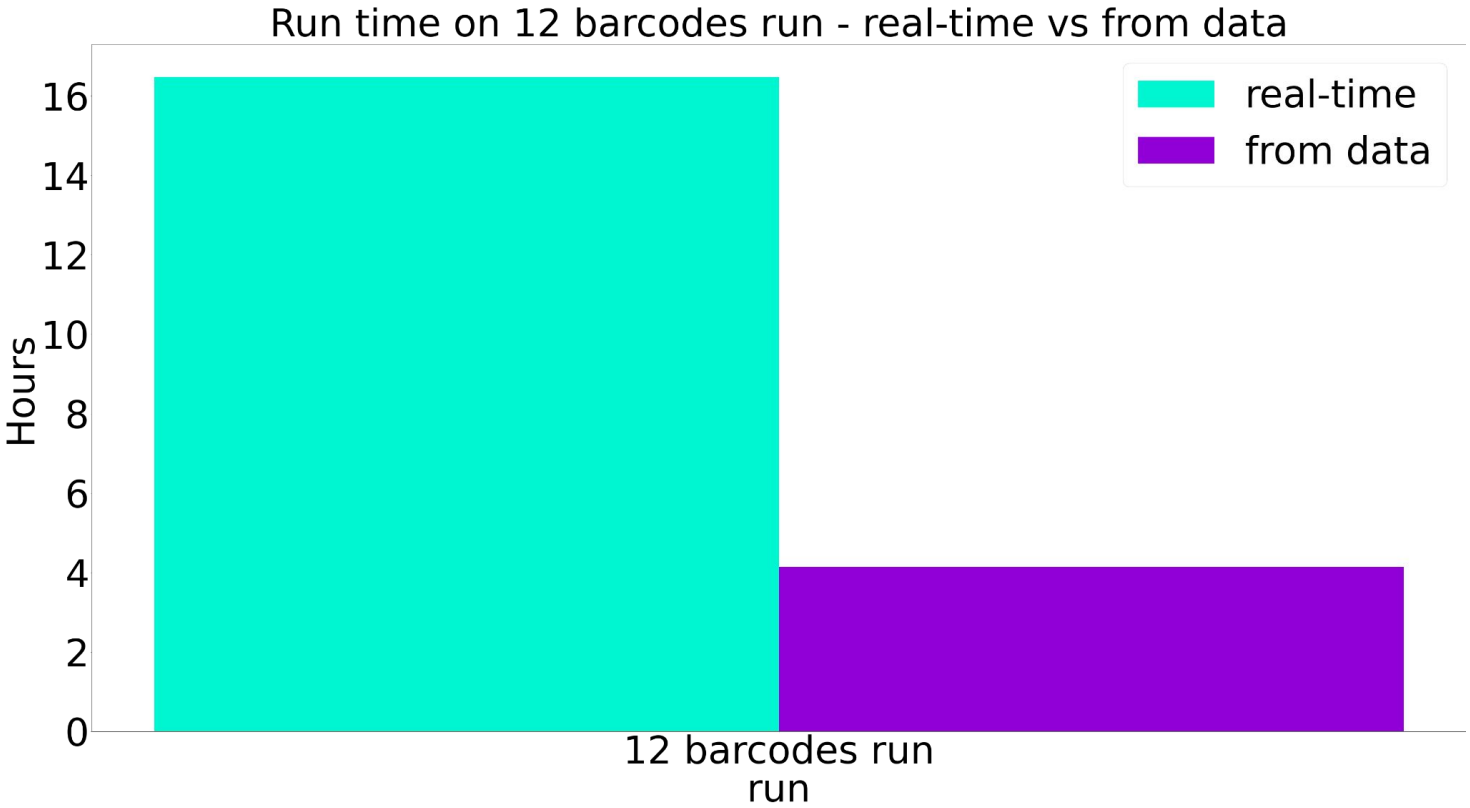

basecalling
batch: batch0
debarcoding
debarcoding_summary
alignment
sample: barcode01
alignment
sample: barcode02
alignment
sample: barcode03
alignment
sample: barcode04
alignment
sample: unclassified
count_bases_in_alignment
count_bases_in_alignment
count_bases_in_alignment
count_bases_in_alignment
count_bases_in_alignment
find_SNPs
find_SNPs
find_SNPs
find_SNPs
find_SNPs
all



Ako veľkosť batch-u ovplyvní čas potrebný na spracovanie dát ?



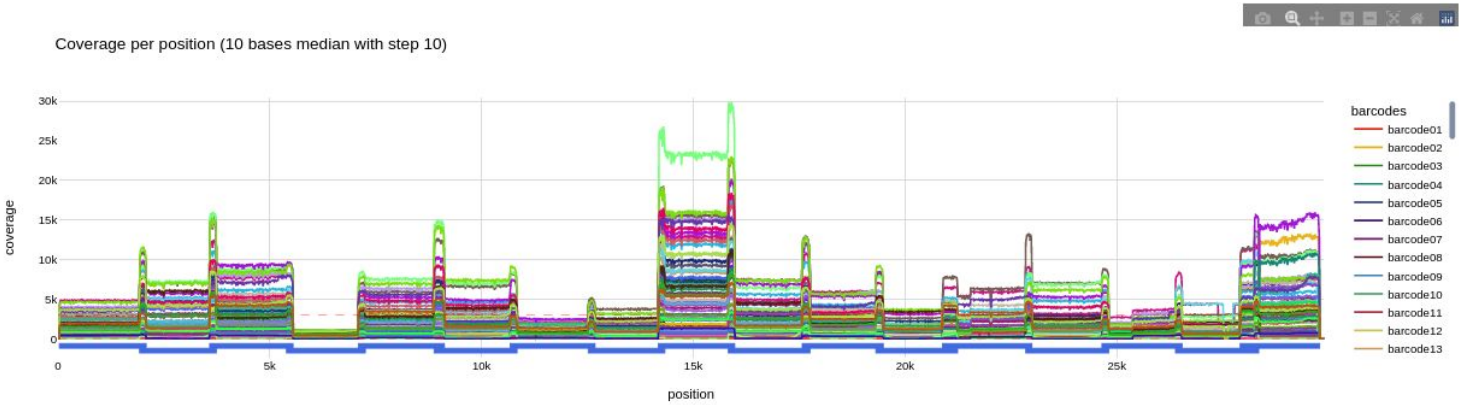
Testovanie na behu v reálnom čase

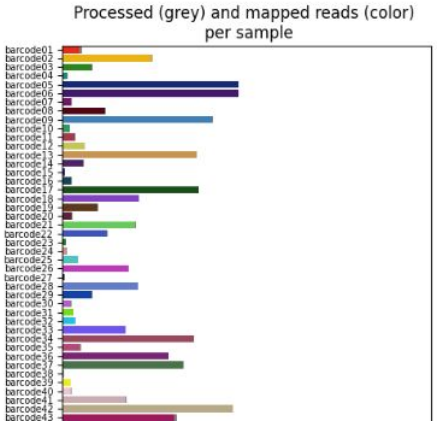


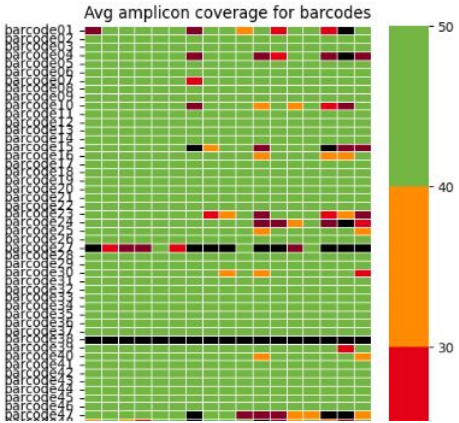
 # reads: 5525055

 # bases: 9595038626

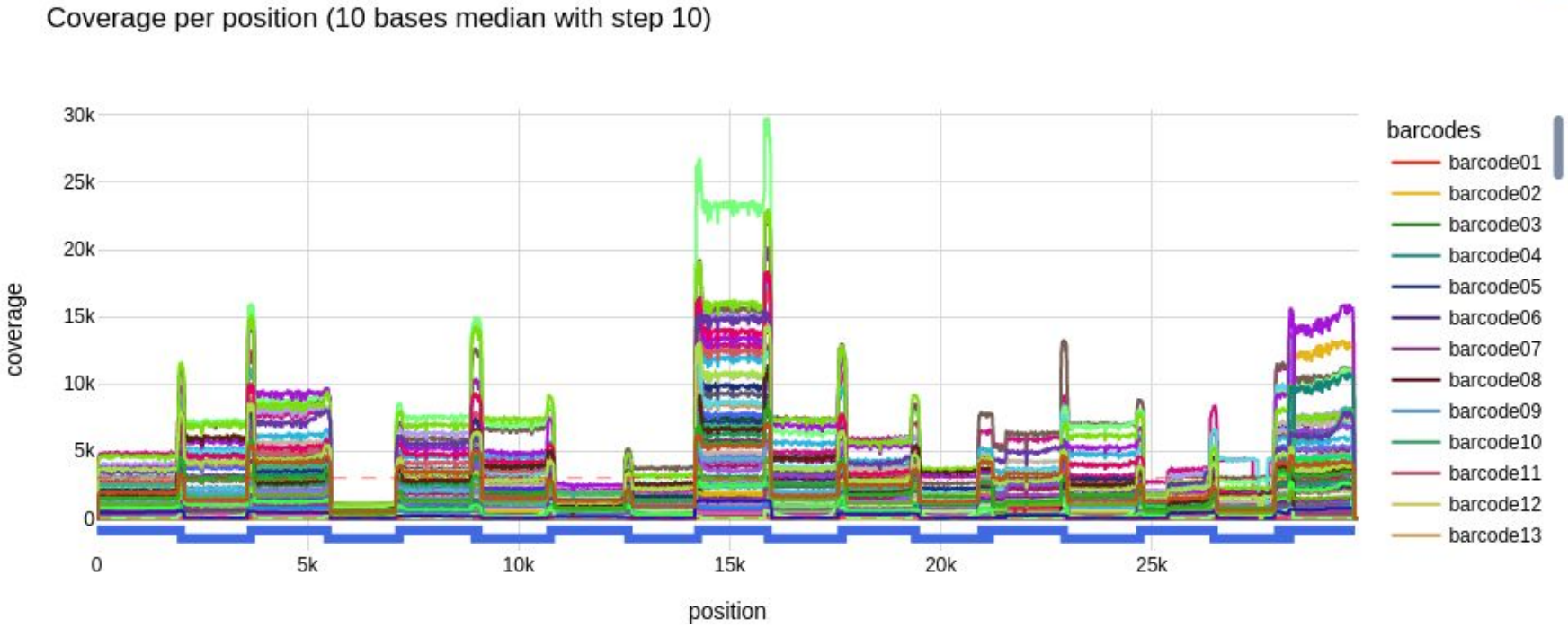
last realod: 2022-04-20 15:29:13

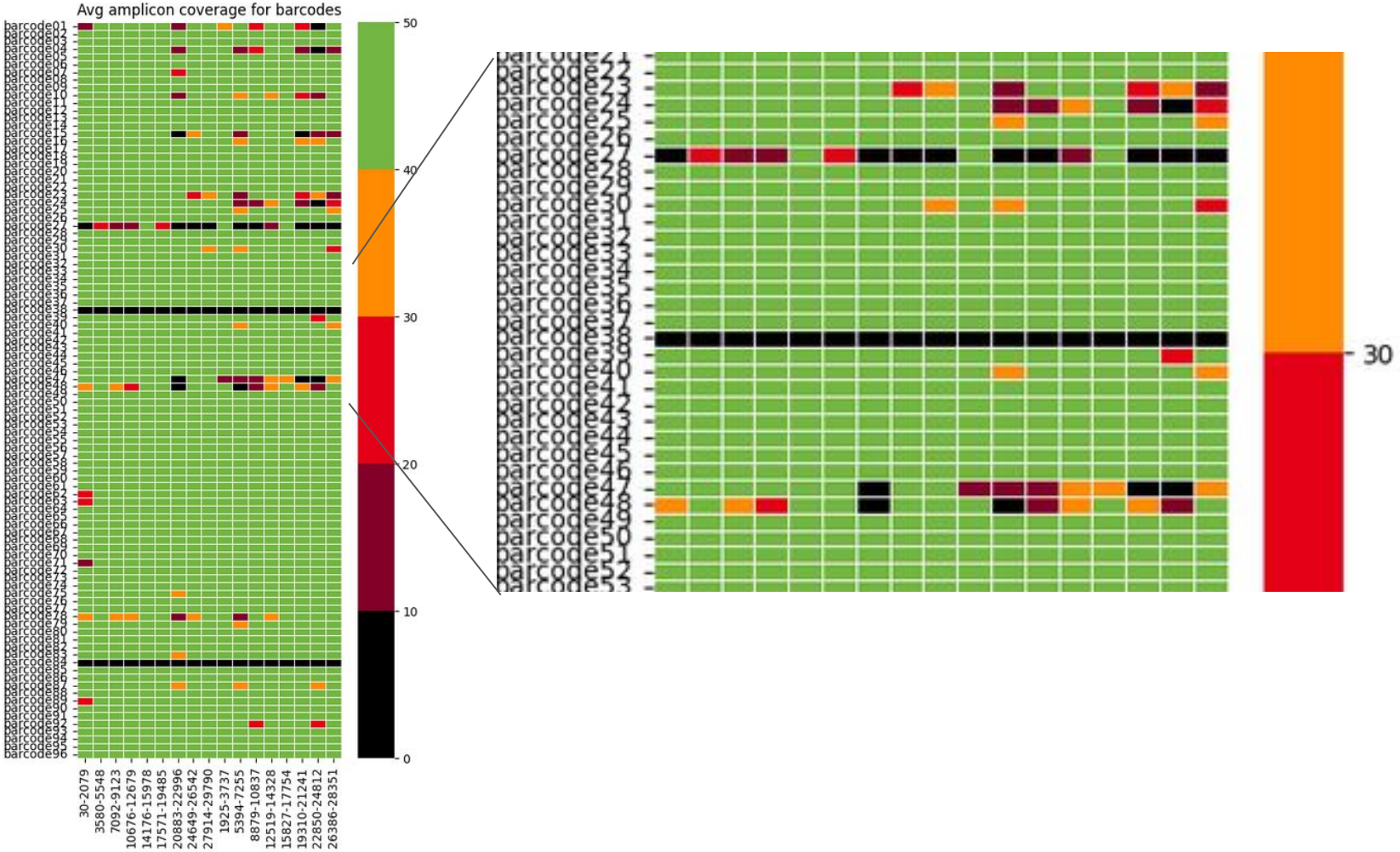






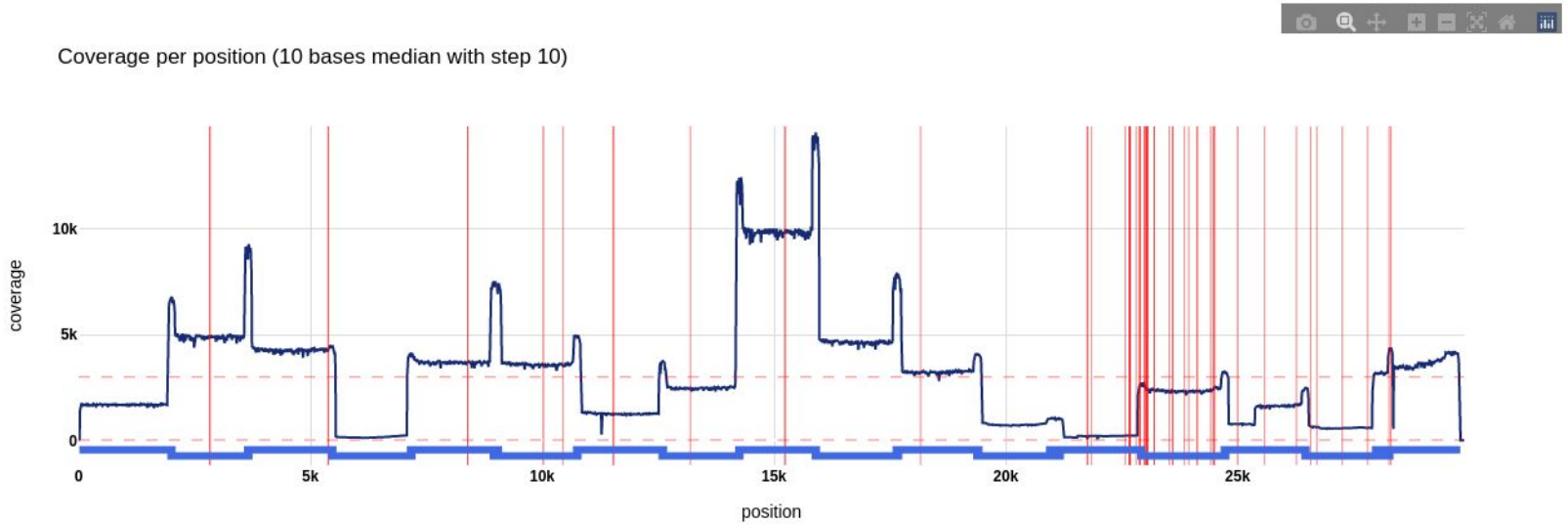
Flask


Coverage per position (10 bases median with step 10)
coverage
30k
25k
20k
15k
10k
5k
0
0
5k
10k
15k
20k
25k
position
barcodes
barcode01
barcode02
barcode03
barcode04
barcode05
barcode06
barcode07
barcode08
barcode09
barcode10
barcode11
barcode12
barcode13


Avg amplicon coverage for barcodes
50
40
30
20
10
0
30-2079
3580-5548
7092-9123
10676-12679
14176-15978
17571-19485
20883-22996
24649-26542
27914-29790
1925-3737
5394-7255
8879-10837
12519-14328
15827-17754
19310-21241
22850-24812
26386-28351
30

Flask

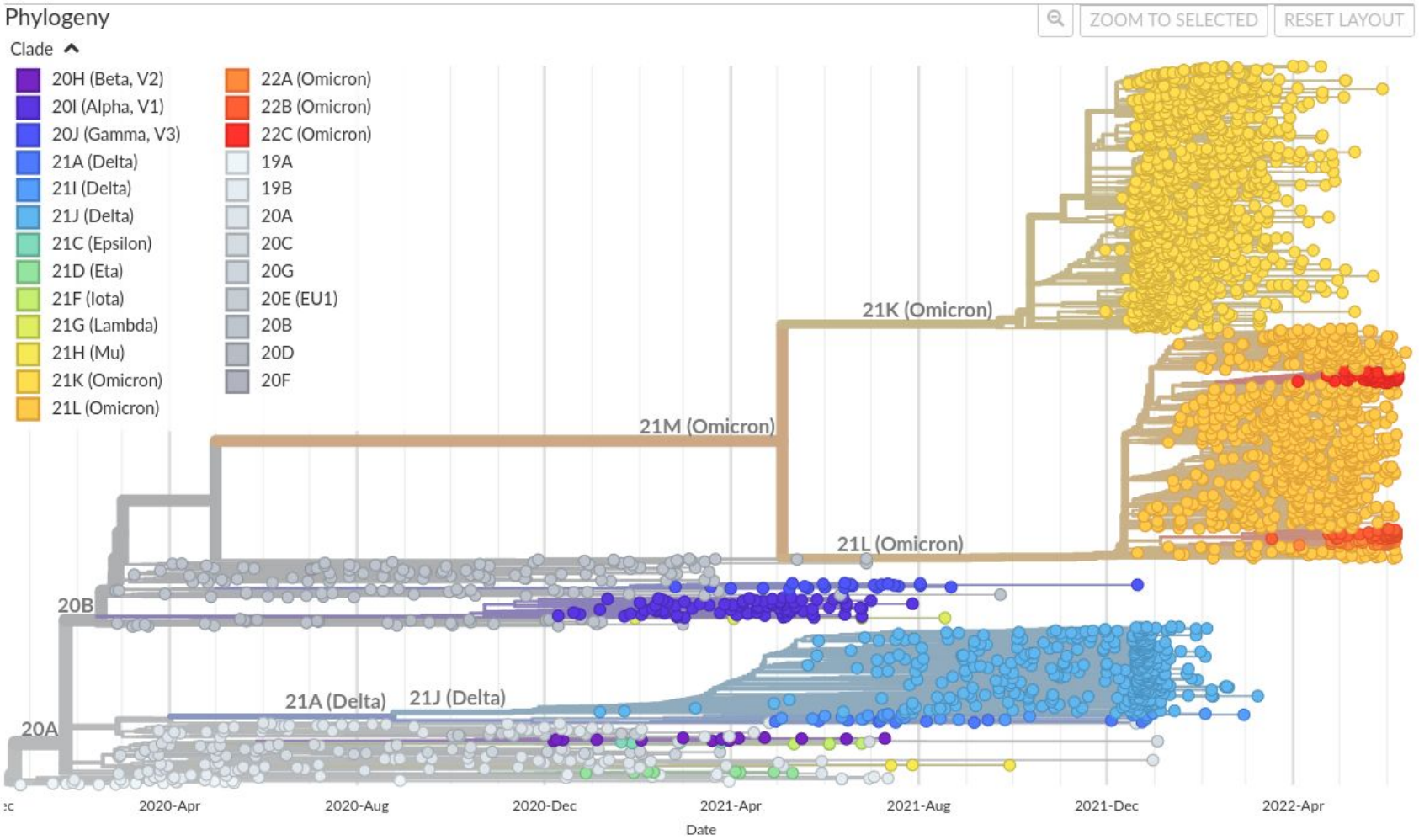
barcode05 | 54264 processed reads | 1863.0 avg read length | 54828 mapped reads | Omicron



variant: Omicron (A2832G, T5386G, G8393A, C10029T, C10449A, A11537G, T13195C, C15240T, A18163G, C21762T, G22578A, T22673C, C22674T, T22679C, G22813T, T22882G, G22898A, G22992A, C22995A, A23013C, A23040G, G23048A, A23055G, A23063T, C23202A, C23525T, T23599G, C23604A, C23854A, G23948T, C24130A, A24424T, T24469A, C24503T, C25000T, C25584T, C26270T, C26577G, G26709A, A27259C, C27807T, A28271T, C28311T)

---variant: BA.1 (A2832G, T5386G, G8393A, A11537G, C15240T, C21762T, C21846T, T22673C, C22674T, G22898A, G23048A, C23202A, C24130A, C24503T)

variant	#samples	samples
Omicron	47	barcode01 , barcode02 , barcode03 , barcode04 , barcode05 , barcode06 , barcode07 , barcode08 , barcode09 , barcode10 , barcode11 , barcode12 , barcode13 , barcode14 , barcode15 , barcode47 , barcode48 , barcode53 , barcode54 , barcode55 , barcode56 , barcode57 , barcode59 , barcode60 , barcode61 , barcode62 , barcode63 , barcode64 , barcode65 , barcode68 , barcode69 , barcode70 , barcode71 , barcode73 , barcode74 , barcode75 , barcode77 , barcode78 , barcode79 , barcode80 , barcode81 , barcode82 , barcode83 , barcode85 , barcode86 , barcode87 , barcode88
BA.1	*39*	*barcode02 , barcode03 , barcode05 , barcode06 , barcode07 , barcode08 , barcode09 , barcode10 , barcode11 , barcode12 , barcode13 , barcode14 , barcode53 , barcode54 , barcode55 , barcode56 , barcode57 , barcode59 , barcode60 , barcode61 , barcode64 , barcode65 , barcode68 , barcode69 , barcode70 , barcode73 , barcode74 , barcode75 , barcode77 , barcode78 , barcode79 , barcode80 , barcode81 , barcode82 , barcode83 , barcode85 , barcode86 , barcode87 , barcode88*
BA.2	*3*	*barcode62 , barcode63 , barcode71*
Delta	45	barcode16 , barcode17 , barcode18 , barcode19 , barcode20 , barcode21 , barcode22 , barcode24 , barcode25 , barcode26 , barcode28 , barcode29 , barcode30 , barcode31 , barcode32 , barcode33 , barcode34 , barcode35 , barcode36 , barcode37 , barcode39 , barcode40 , barcode41 , barcode42 , barcode43 , barcode44 , barcode45 , barcode46 , barcode49 , barcode50 , barcode51 , barcode52 , barcode58 , barcode66 , barcode67 , barcode72 , barcode76 , barcode89 , barcode90 , barcode91 , barcode92 , barcode93 , barcode94 , barcode95 , barcode96
AY.43	*28*	*barcode16 , barcode18 , barcode19 , barcode21 , barcode22 , barcode24 , barcode25 , barcode29 , barcode32 , barcode33 , barcode34 , barcode35 , barcode36 , barcode40 , barcode41 , barcode42 , barcode44 , barcode46 , barcode50 , barcode51 , barcode58 , barcode72 , barcode76 , barcode89 , barcode91 , barcode92 , barcode95 , barcode96*



https://nextstrain.org/ncov/gisaid/global

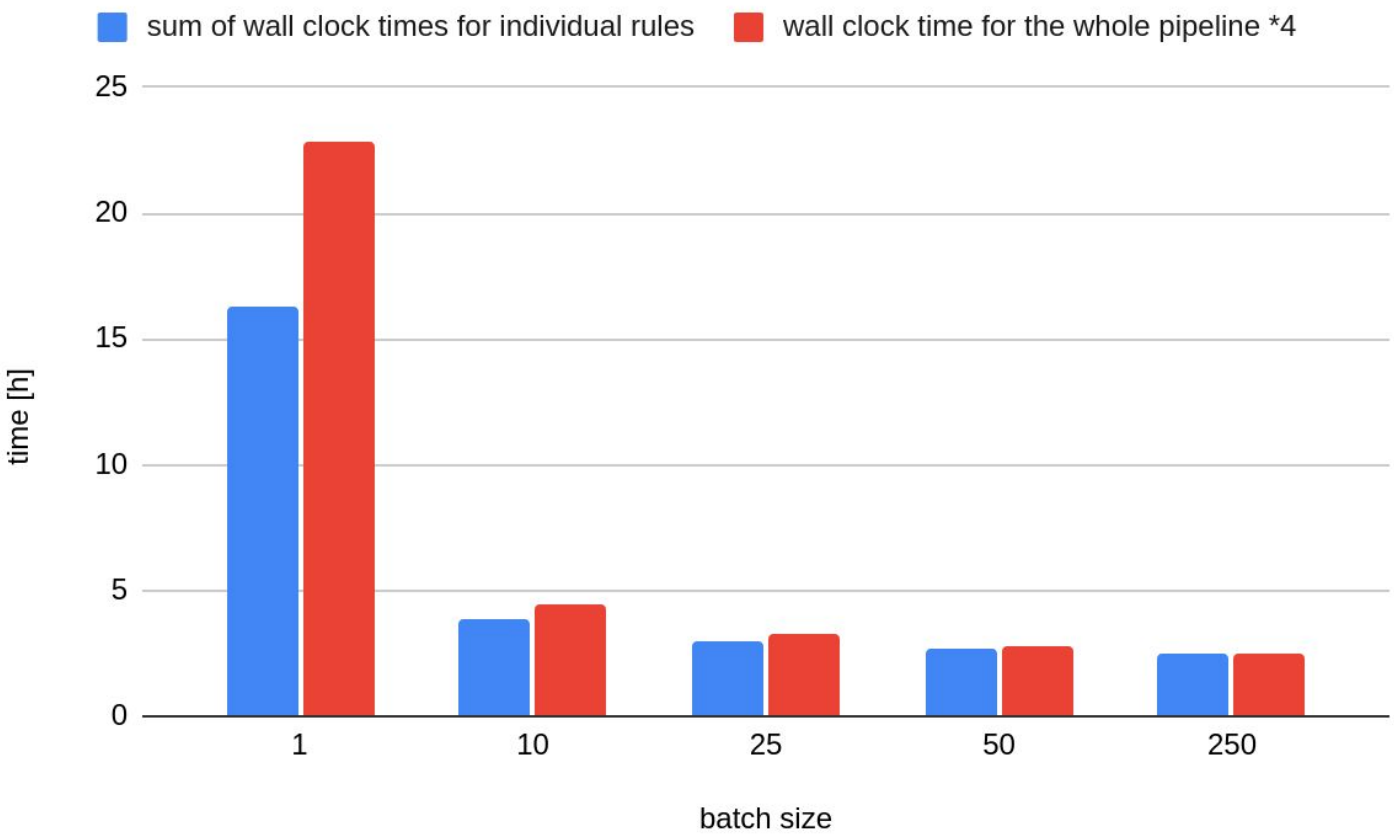
Zhrnutie

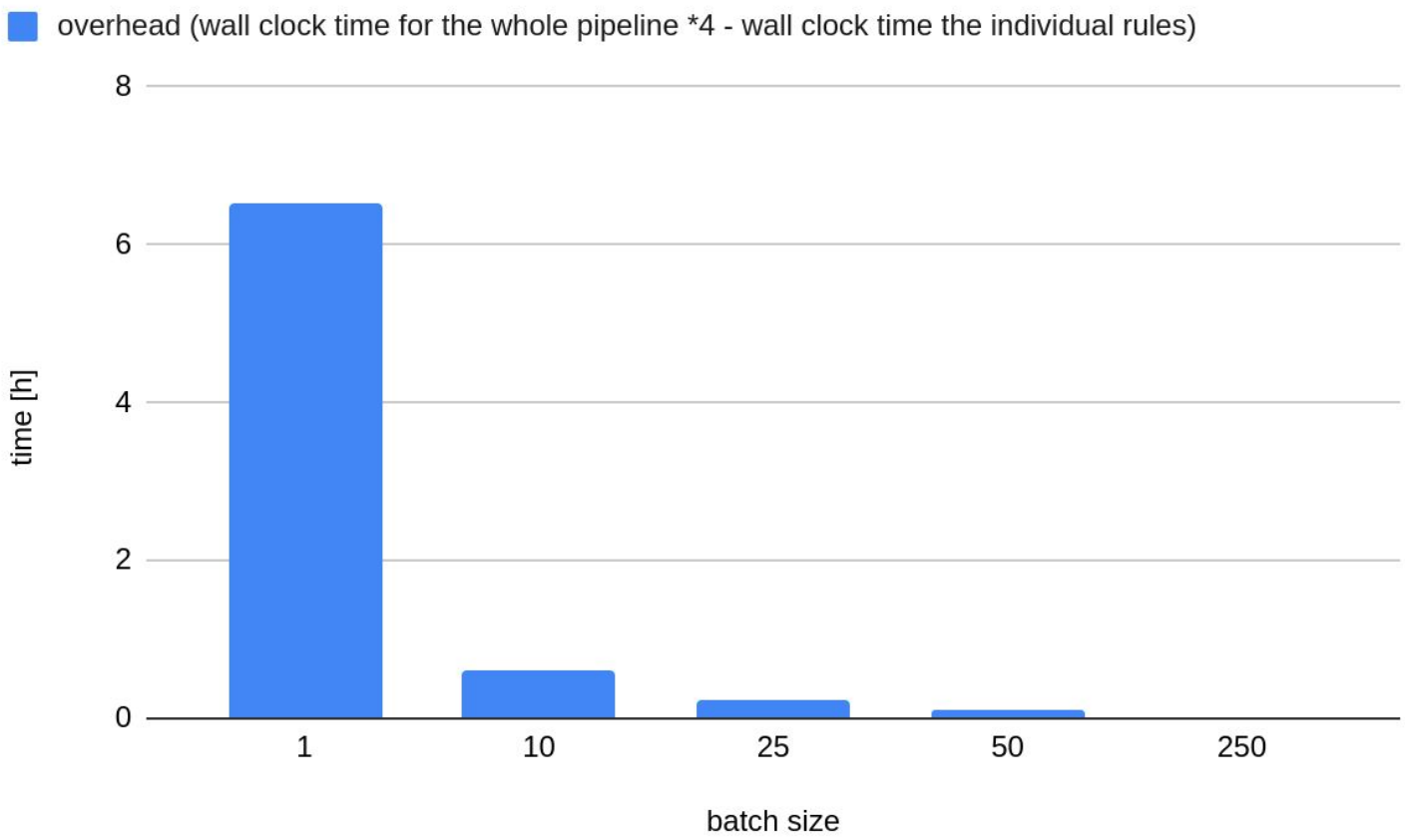
- riešili sme problém monitorovania sekvenovania SARS-CoV-2 vírusu
- navrhli a naprogramovali sme riešenie v ktorom sme sa sústredili na odstránenie bežných problémov podobných nástrojov
- úspešne sme otestovali naše riešenie pri monitorovaní sekvenovania vzoriek zo SAV

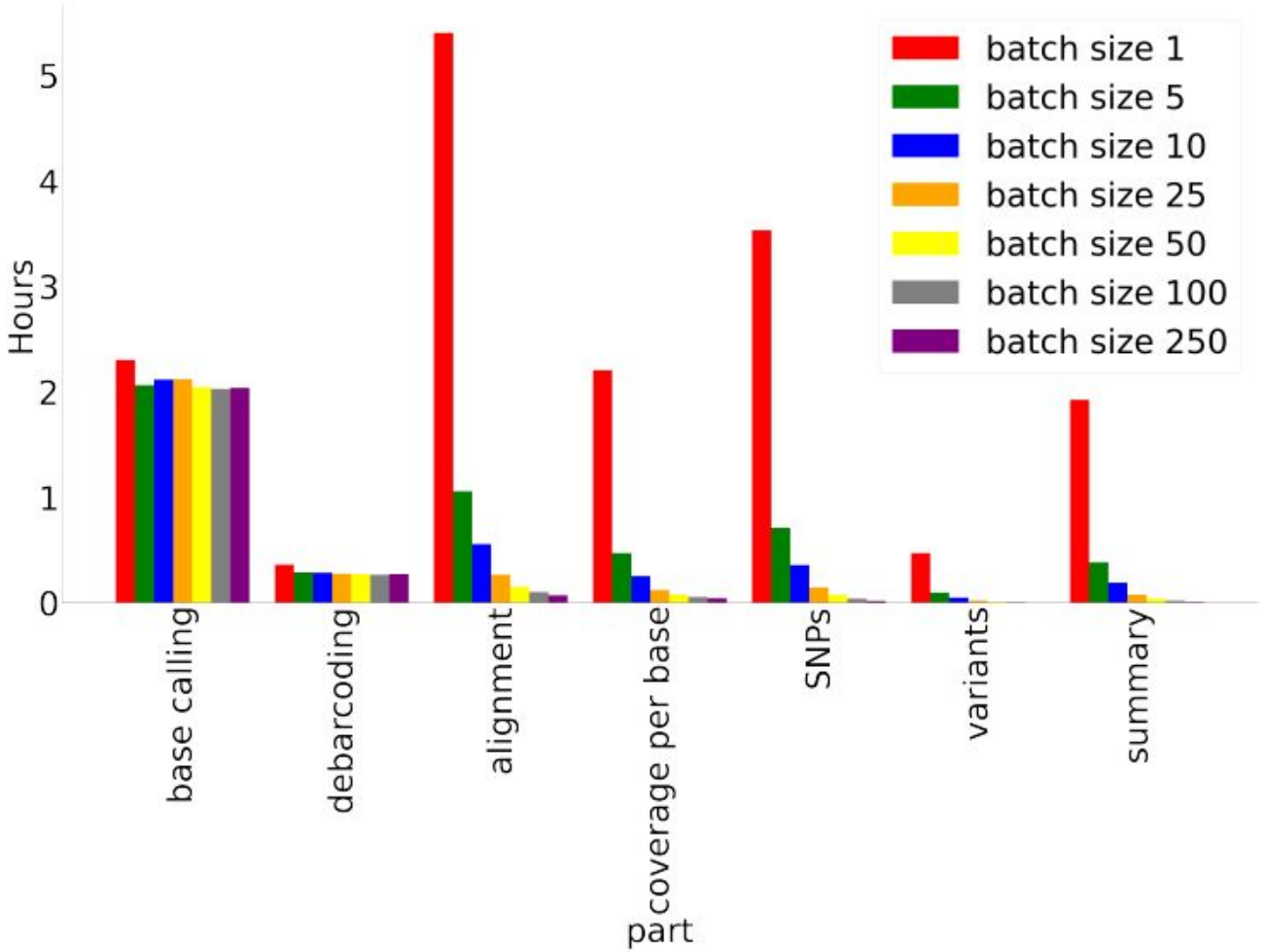
Ďakujem za pozornosť

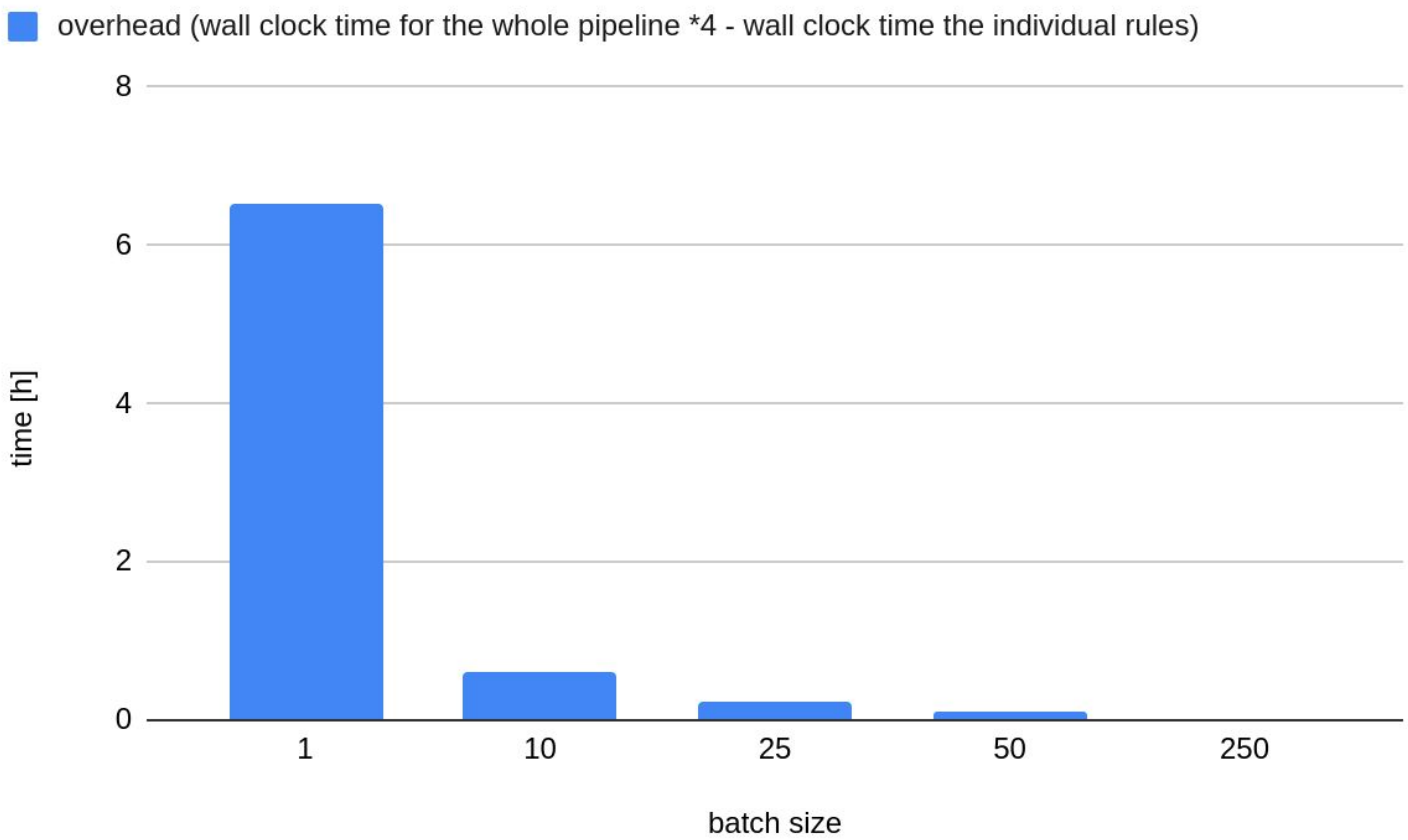
...

1. Viete odhadnúť vplyv veľkosti "batchov" na množstvo času, ktorý potrebuje Snakemake na plánovanie výpočtov? Koľko výpočtového času celkovo zaberá plánovanie výpočtov týmto nástrojom (t.j. je to zanedbateľné množstvo alebo nie)?
2. V prípade, že by výpočtový čas, použitý Snakemake-om bol signifikantný, viete si predstaviť nejakú alternatívu (napr. vlastný skript, alebo iný nástroj na plánovanie úloh)?


sum of wall clock times for individual rules
wall clock time for the whole pipeline *4
25
20
15
10
5
0
time [h]
1
10
25
50
250
batch size

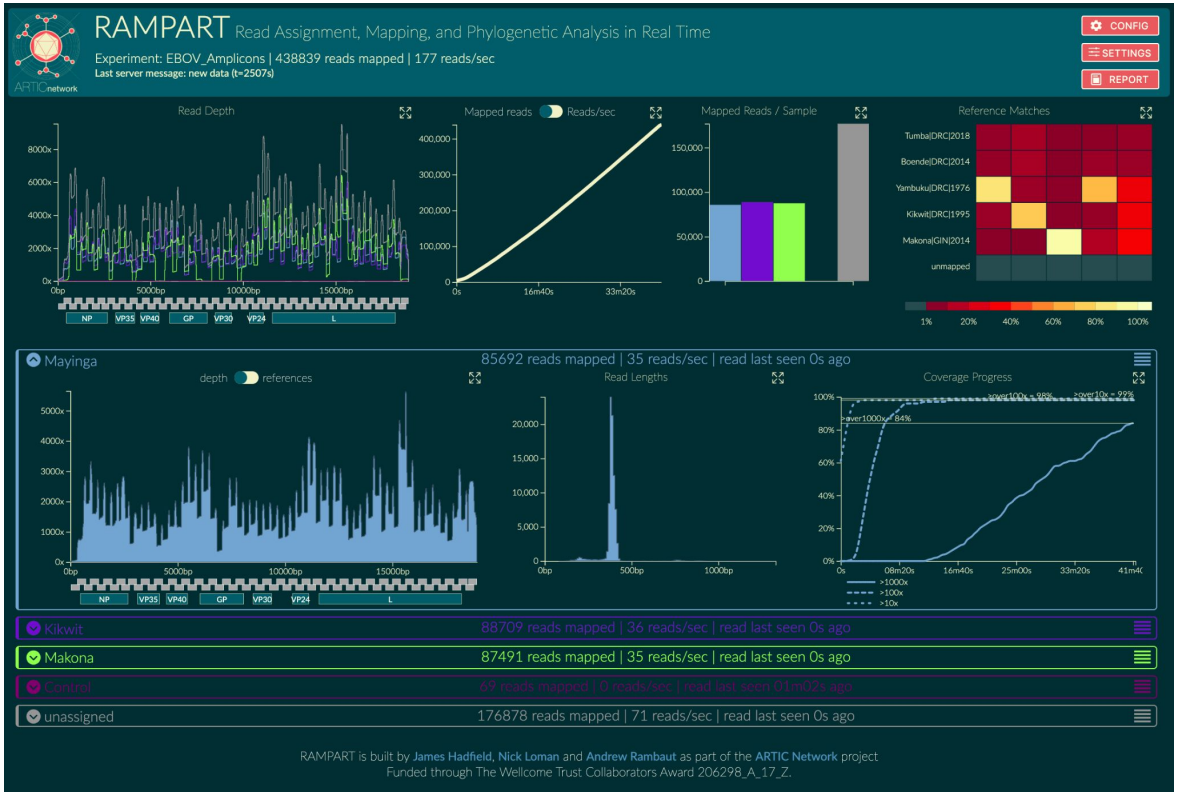

overhead (wall clock time for the whole pipeline *4 - wall clock time the individual rules)
8
6
4
2
0
time [h]
1
10
25
50
250
batch size


batch size 1
batch size 5
batch size 10
batch size 25
batch size 50
batch size 100
batch size 250
Hours
0
1
2
3
4
5
base calling
debarcoding
alignment
coverage per base
SNPs
variants
summary
part


overhead (wall clock time for the whole pipeline *4 - wall clock time the individual rules)
8
6
4
2
0
time [h]
1
10
25
50
250
batch size

...

RAMPART

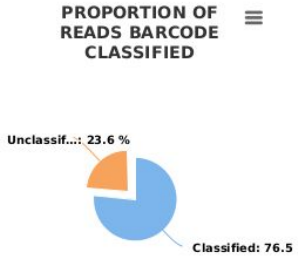


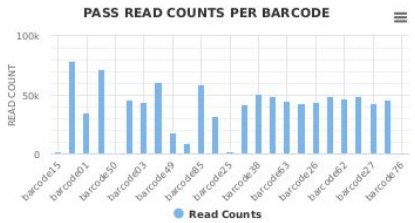
https://github.com/artic-network/rampart

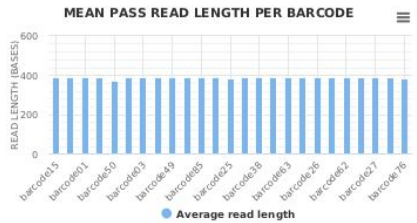
minoTour

8th September 2021 14:18

Report for artic output of FAQ01023_CV197_25_M1 generated 8th September 2021 14:18

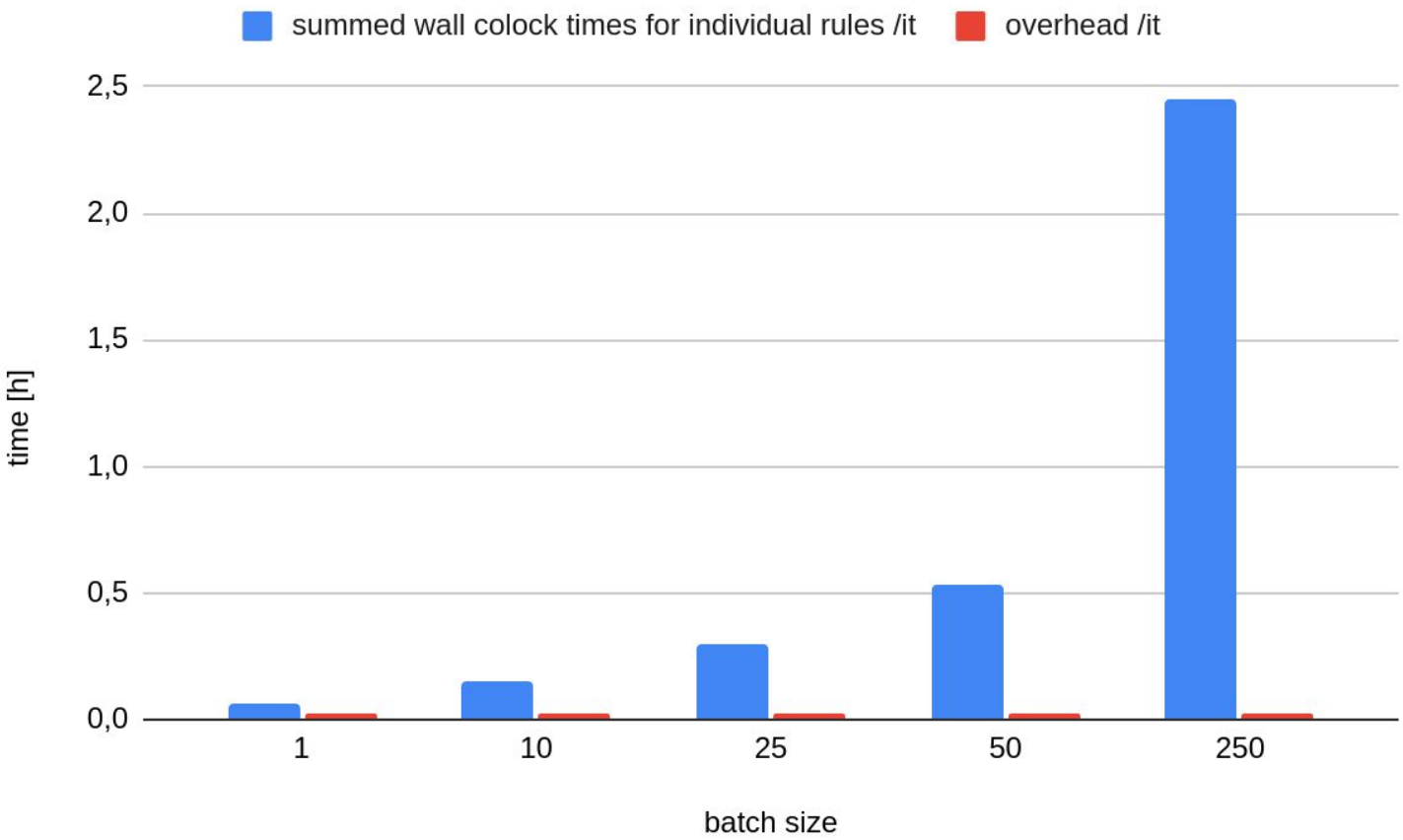


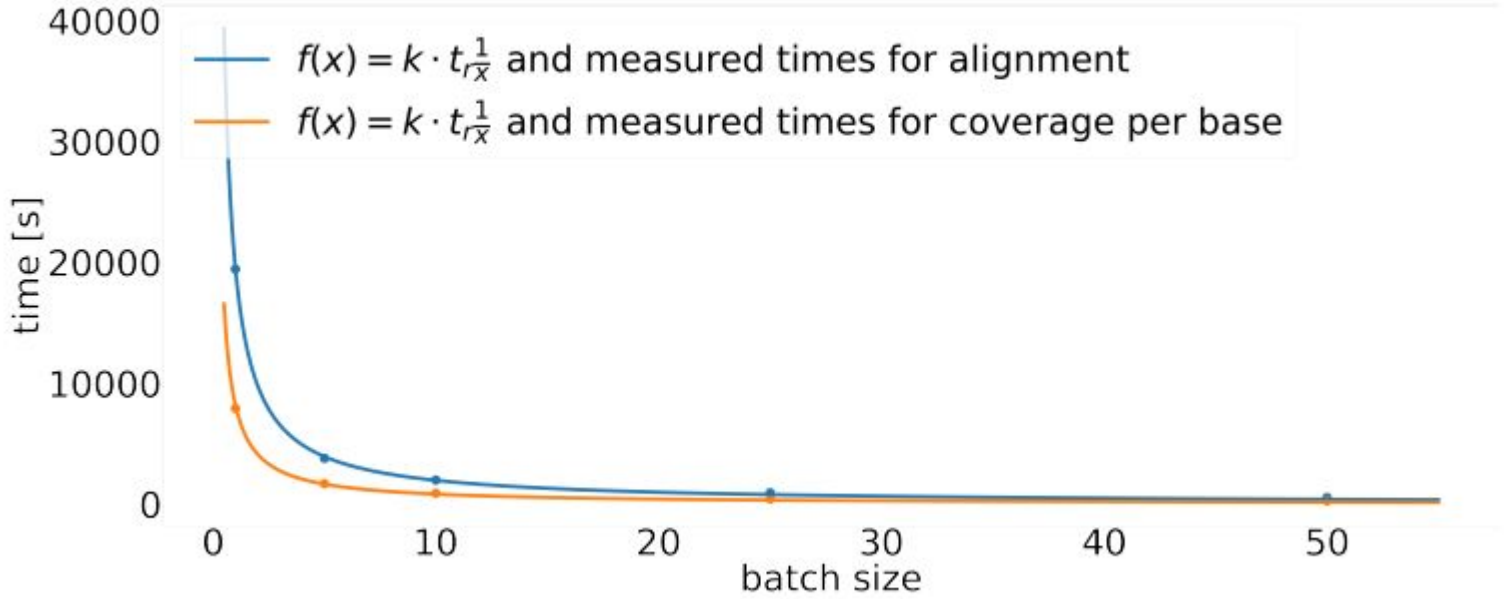




Barcode Name	Chromosome	Read Len	Read Count	Yield	Coverage	# Success Amplicons	# Partial Amplicons	# Failed Amplicons	Lineage	VoC found
barcode01	MN908947.3	388	34,489	13,393,773	438.88	76	4	19	Currently unknown	Not Tested
barcode02	MN908947.3	388	49,127	19,051,766	624.2	99	0	0	B.1.1.7	0 VOC-20DEC-01 Name: phe-label, dtype: object
barcode03	MN908947.3	389	44,022	17,106,691	560.4	94	5	0	Currently unknown	Not Tested
barcode13	MN908947.3	388	32,321	12,537,233	410.86	88	9	2	Currently unknown	Not Tested
barcode14	MN908947.3	391	45,820	17,913,802	587.57	99	0	0	B.1.1.7	0 VOC-20DEC-01 Name: phe-label, dtype: object
barcode15	MN908947.3	388	2,415	936,286	30.72	17	5	77	Currently unknown	Not Tested
barcode25	MN908947.3	386	2,348	906,274	29.77	8	2	89	Currently unknown	Not Tested
barcode26	MN908947.3	390	44,063	17,173,790	563.16	99	0	0	B.1.1.7	0 VOC-20DEC-01 Name: phe-label, dtype: object

https://www.biorxiv.org/content/10.1101/2021.09.13.459777v2

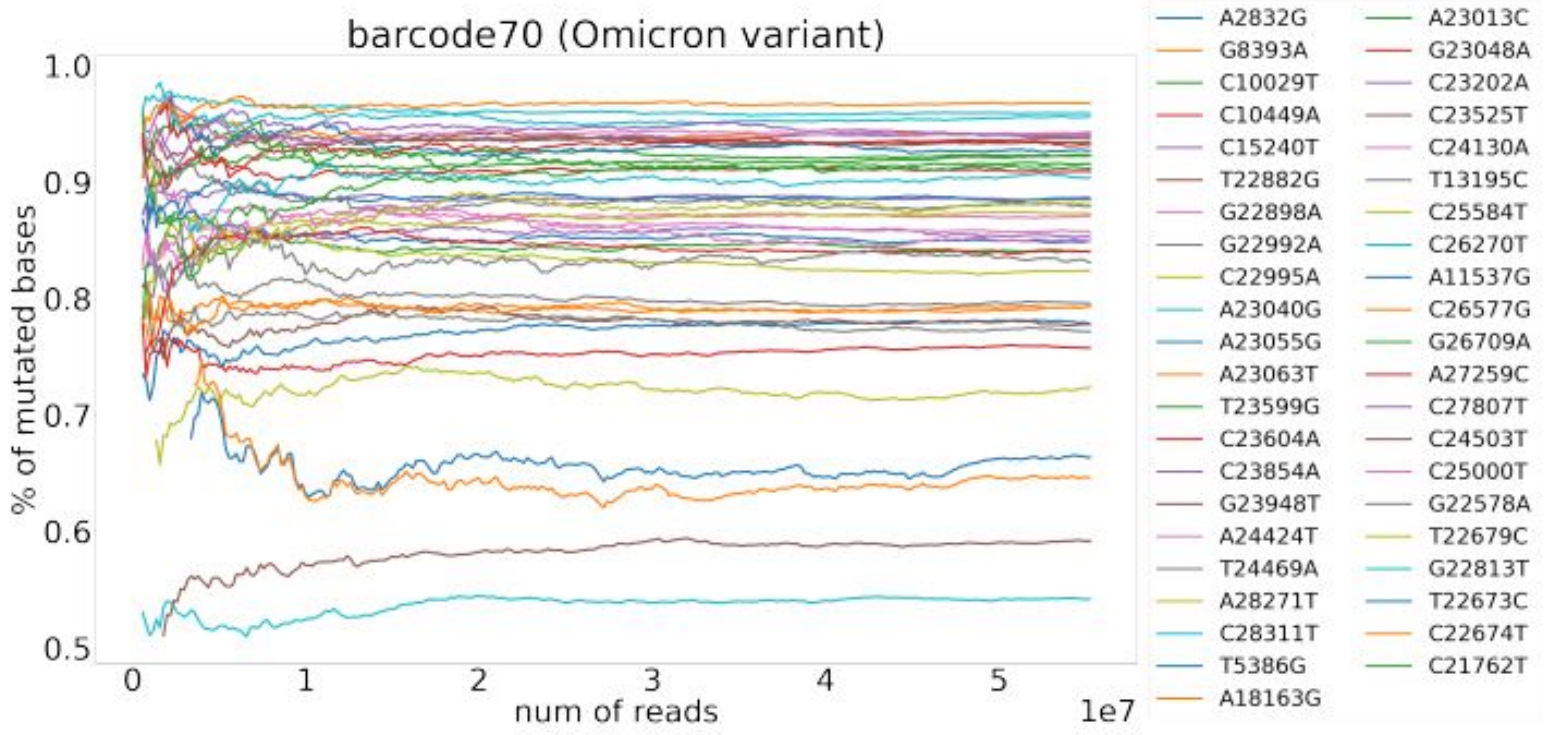

summed wall colock times for individual rules /it
overhead /it
2,5
2,0
1,5
1,0
0,5
0,0
time [h]
1
10
25
50
250
batch size



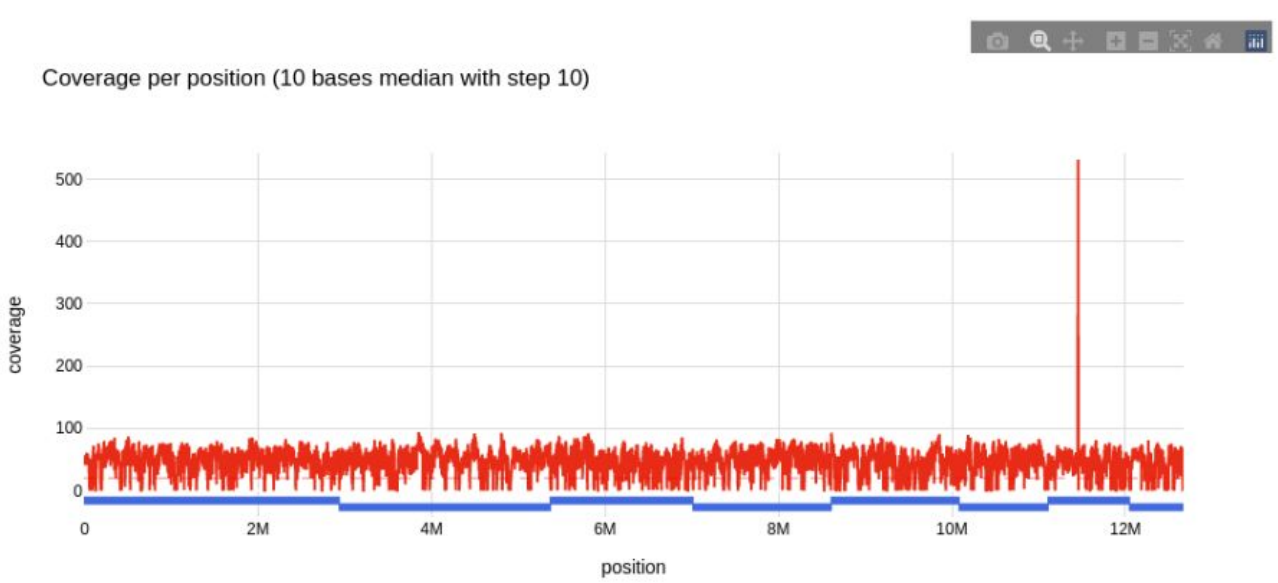
k - num of files to be analysed
t_r - average time* for the rule for a batch
x - batch size

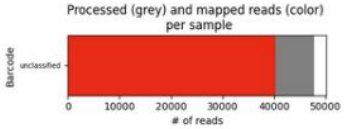
*(calculated as the sum of all the measured wall clock times for the rule divided by the total number of batches)

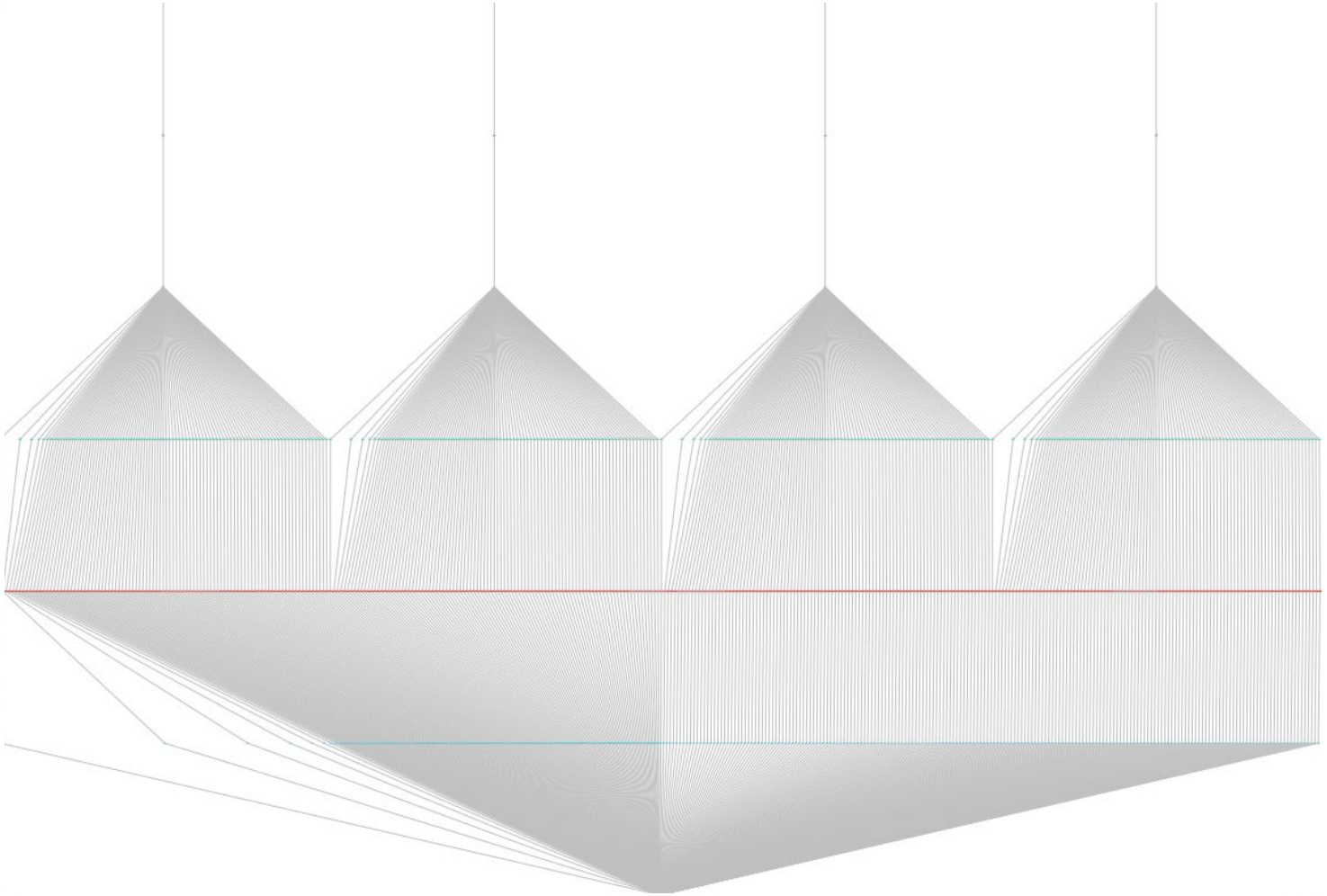
Presnosť určenia variantu vírusu



Testovanie na sekvenačnom behu na kvasinke Candida orthopsilosis








basecalling
debarcoding
debarcoding_summary
alignment
count_bases_in_alignment
find_SNPs
summarize_coverage_per_base
summarize_SNPs
summarize_basecalled
summarize_barcode_summary
summarize_aligned
variant_calling
summarize_batch
all

a

PRIMAL SCHEME

Scheme name

Scheme name

Email

Email

Fasta

Choose File No file chosen

One or more viral reference genomes in FASTA format

Amplicon length

400

Overlap

75

Generate my scheme!

Want to try it out but no genomes to hand?
Download CHIKV_demo.fa

b

Job Name CHIKV

Download as CSV

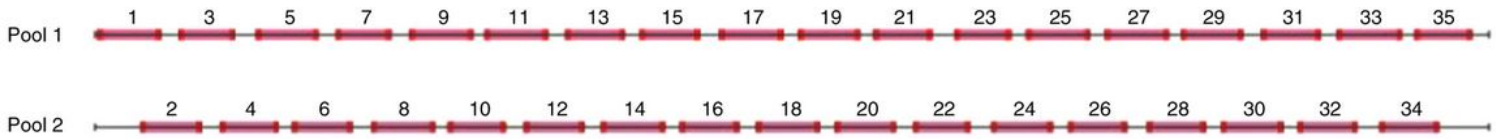
FASTA file uploads/2016/11/05/CHIKV_demo.fa
Amplicon length 400
Overlap 75

Primer Table

Primer Table

Region No.	Pool	Left Primer Name	Left Primer Sequence	Right Primer Name	Right Primer Sequence
1	1	400_1_LEFT_3	CTTTTTGAAGGCCCTGCAACGT	400_1_RIGHT_3	ACGGCCATCACCTCTTGTAAGT
2	2	400_2_LEFT_1	TGTCGGACAGGAAGTACCACTG	400_2_RIGHT_1	CCGCCCAATTTGTCGAGTATGA
3	1	400_3_LEFT_1	GTACTGGGTAGGGTTCGACACA	400_3_RIGHT_1	TAAAGGCCTGGGCTCATCGTTA
4	2	400_4_LEFT_4	TCGGTGTTCCATCTAAAGGGCA	400_4_RIGHT_4	CTTCATGGTGTTCGTGTTCCGT
5	1	400_5_LEFT_2	TGTGATCAAATGACCGGCATCC	400_5_RIGHT_2	GTCCTCAACGGGATTGACAACC
6	2	400_6_LEFT_0	ACAAGAGGCCTGATACCCAGTC	400_6_RIGHT_0	GCGCCTCTCGGAGTCTCTATTA
7	1	400_7_LEFT_1	TACAGGCAGCACAGGAAGATGT	400_7_RIGHT_1	TTGTACACCATCGTTGCGCTTT
8	2	400_8_LEFT_3	GGAGCAAGTGAAGACGTGTACG	400_8_RIGHT_3	GGACAAGCGGGGCGAATTTTTA
9	1	400_9_LEFT_3	TGCAAGAAGGAAGAAGCTGCAG	400_9_RIGHT_3	GGCAAGTAACGTTCCAGAGTGG
10	2	400_10_LEFT_0	ATCTGCACGTACGGTTGATTCG	400_10_RIGHT_0	GGTTTTGTTGAGCCCGTAGTGT

c



https://www.nature.com/articles/nprot.2017.066