Human Pathogen Sequencing in Austria

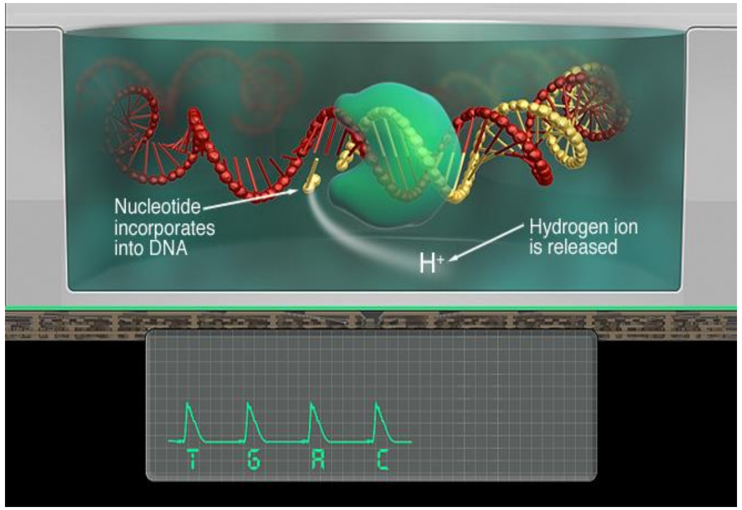
Alexander Indra

WHOLE GENOME SEQUENCING
The present

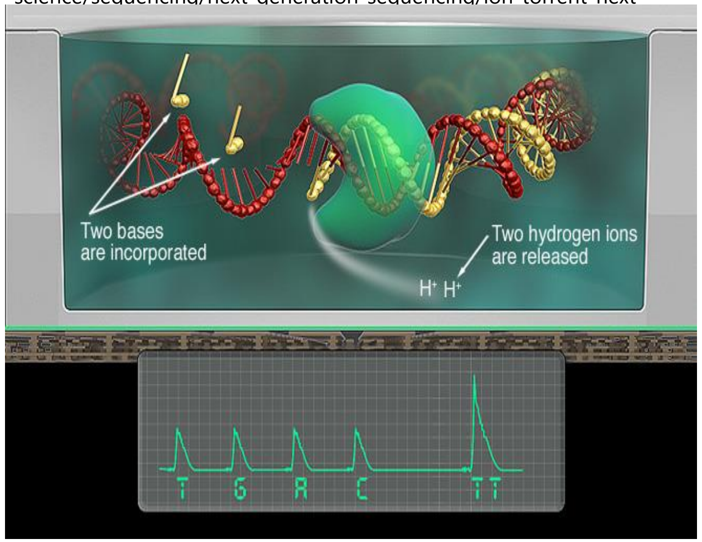
http://upload.wikimedia.org/wikipedia/commons/7/77/Illumina_Hiseq_2000_sequencers,_BGI_Hong_Kong_sequencing_room.JPG

Current generation sequencing: Limitations

- **Read length**
  - Between 35bp-1000bp
  - Quality reduces the longer the read length
- **Homopolymer-problem**
  - a sequence of identical bases
  - brightness of the light or the pH increase indicates the length of the homopolymer, because each homopolymer sequence is called in a single flow
    - difficult to distinguish e.g. AAAAA from AAAA or AAAAAA



Source: https://www.thermofisher.com/at/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-



Source: https://www.thermofisher.com/at/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html

Current generation sequencing: Limitations

- Indel-Errors
- "indel" a combination of **in**sertion and **del**etion
- **Insertions**
  - An additional nucleotide base is inserted into the Sequence
  - due to sequencing errors
- **Deletion**
  - An nucleotide base is left out of the Sequence
  - due to sequencing errors
- Difficult interpretation

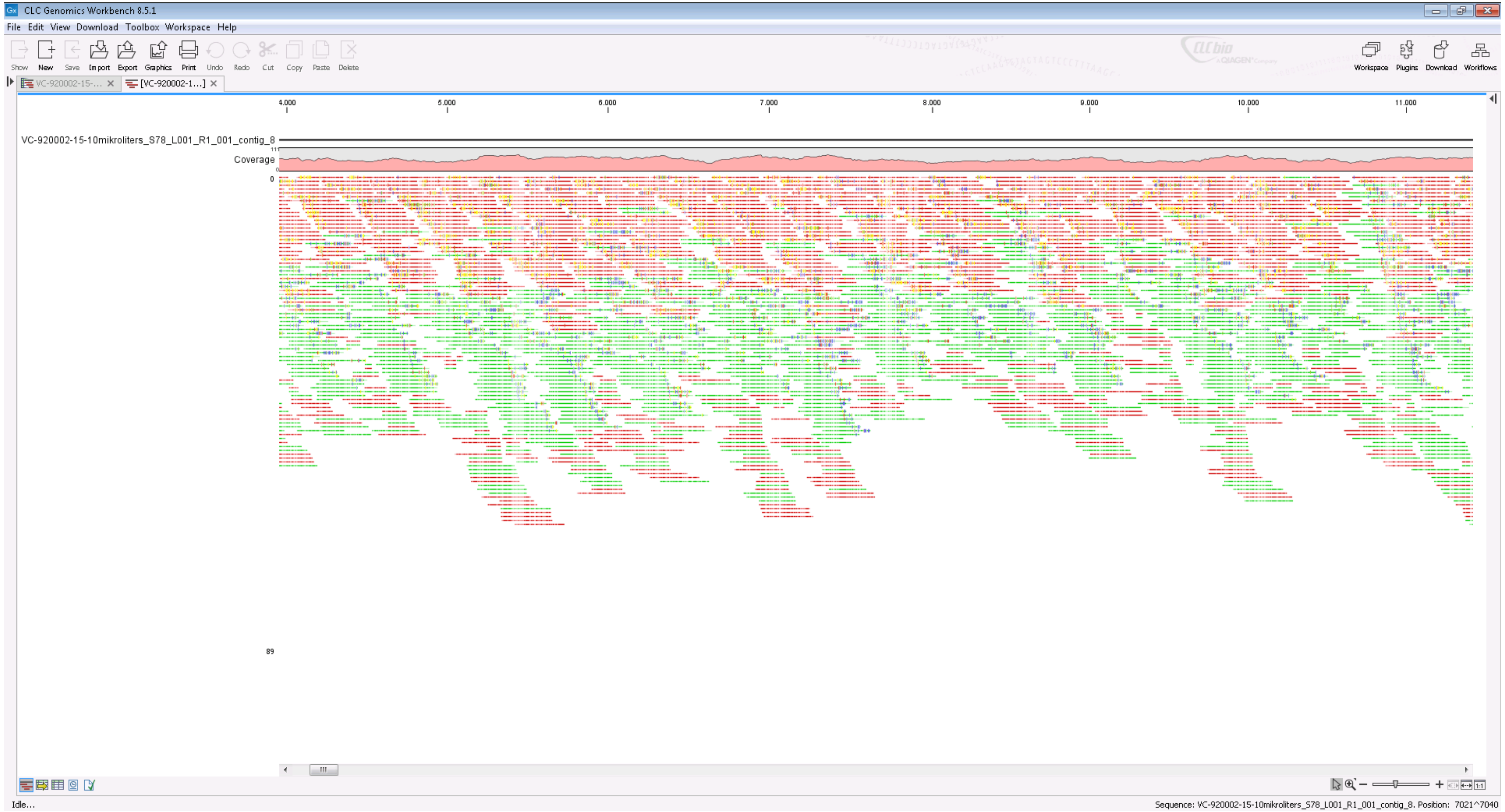
Current generation sequencing

- **Applications**
  - **De novo sequencing and Resequencing**
  - **Metagenomics**
  - **Amplicon-sequencing**
  - **Transcriptome Sequencing**

Current generation sequencing: De novo sequencing/Resequencing

- **De-novo-sequencing**
- **Assambly of an unknown strain without an reference Genome**
- **Various software tools are used**
  - **Velvet (http://www.ebi.ac.uk/~zerbino/velvet/)**
  - **SPAdes (https://en.wikipedia.org/wiki/SPAdes_%28software%29)**

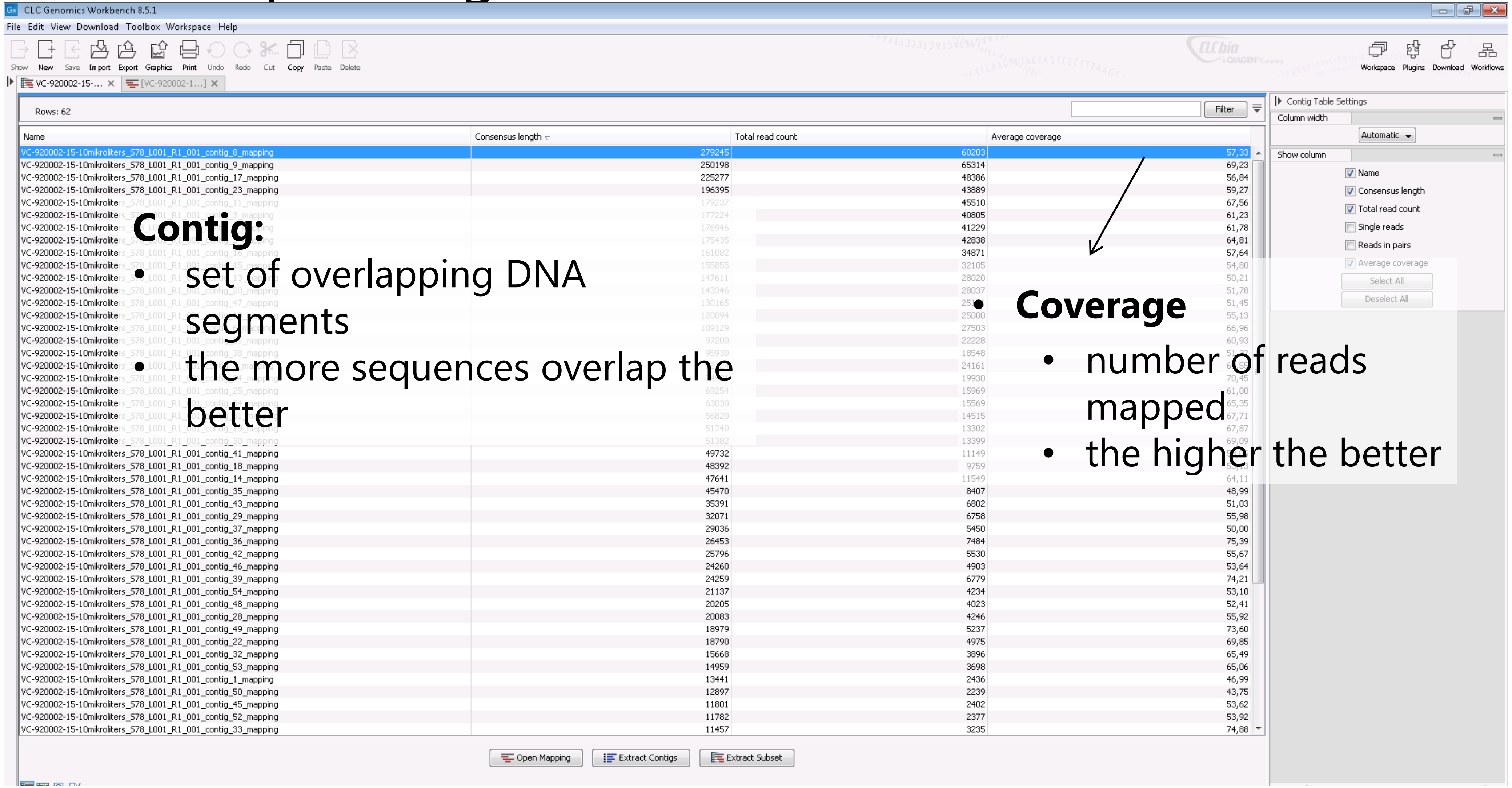
Current generation sequencing:
De novo sequencing



Idle...

Sequence: VC-920002-15-10mikroliters_S78_L001_R1_001_contig_8. Position: 7021^7040

Current generation sequencing: De novo sequencing



Rows: 62

Name	Consensus length	Total read count	Average coverage
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_8_mapping	279245	60203	57,33
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_9_mapping	250198	65314	69,23
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_17_mapping	225277	48386	56,84
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_23_mapping	196395	43889	59,27
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_11_mapping	179237	45510	67,56
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_3_mapping	177224	40805	61,23
VC-920002-15-10mikroliters_[illegible]	176946	41229	61,78
VC-920002-15-10mikroliters_[illegible]	175435	42838	64,81
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_16_mapping	161082	34871	57,64
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_15_mapping	155855	32105	54,80
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_13_mapping	147611	28020	50,21
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_20_mapping	143346	28037	51,78
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_47_mapping	130165	[illegible]	51,45
VC-920002-15-10mikroliters_S78_L001_R1_[illegible]	120094	25000	55,13
VC-920002-15-10mikroliters_S78_L001_R1_[illegible]	109129	27503	66,96
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_[illegible]_mapping	97200	22228	60,93
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_38_mapping	95930	18548	[illegible]
VC-920002-15-10mikroliters_S78_L001_R1_[illegible]	[illegible]	24161	[illegible]
VC-920002-15-10mikroliters_S78_L001_R1_[illegible]_mapping	[illegible]	19930	70,45
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_25_mapping	69254	15969	61,00
VC-920002-15-10mikroliters_S78_L001_R1_[illegible]_mapping	63030	15569	65,35
VC-920002-15-10mikroliters_S78_L001_R1_[illegible]	56820	14515	67,71
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_19_mapping	51740	13302	67,87
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_30_mapping	51382	13399	69,09
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_41_mapping	49732	11149	[illegible]
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_18_mapping	48392	9759	[illegible]
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_14_mapping	47641	11549	64,11
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_35_mapping	45470	8407	48,99
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_43_mapping	35391	6802	51,03
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_29_mapping	32071	6758	55,98
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_37_mapping	29036	5450	50,00
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_36_mapping	26453	7484	75,39
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_42_mapping	25796	5530	55,67
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_46_mapping	24260	4903	53,64
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_39_mapping	24259	6779	74,21
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_54_mapping	21137	4234	53,10
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_48_mapping	20205	4023	52,41
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_28_mapping	20083	4246	55,92
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_49_mapping	18979	5237	73,60
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_22_mapping	18790	4975	69,85
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_32_mapping	15668	3896	65,49
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_53_mapping	14959	3698	65,06
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_1_mapping	13441	2436	46,99
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_50_mapping	12897	2239	43,75
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_45_mapping	11801	2402	53,62
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_52_mapping	11782	2377	53,92
VC-920002-15-10mikroliters_S78_L001_R1_001_contig_33_mapping	11457	3235	74,88

Contig Table Settings

Column width: Automatic

Show column
- ☑ Name
- ☑ Consensus length
- ☑ Total read count
- ☐ Single reads
- ☐ Reads in pairs
- ☑ Average coverage

Select All

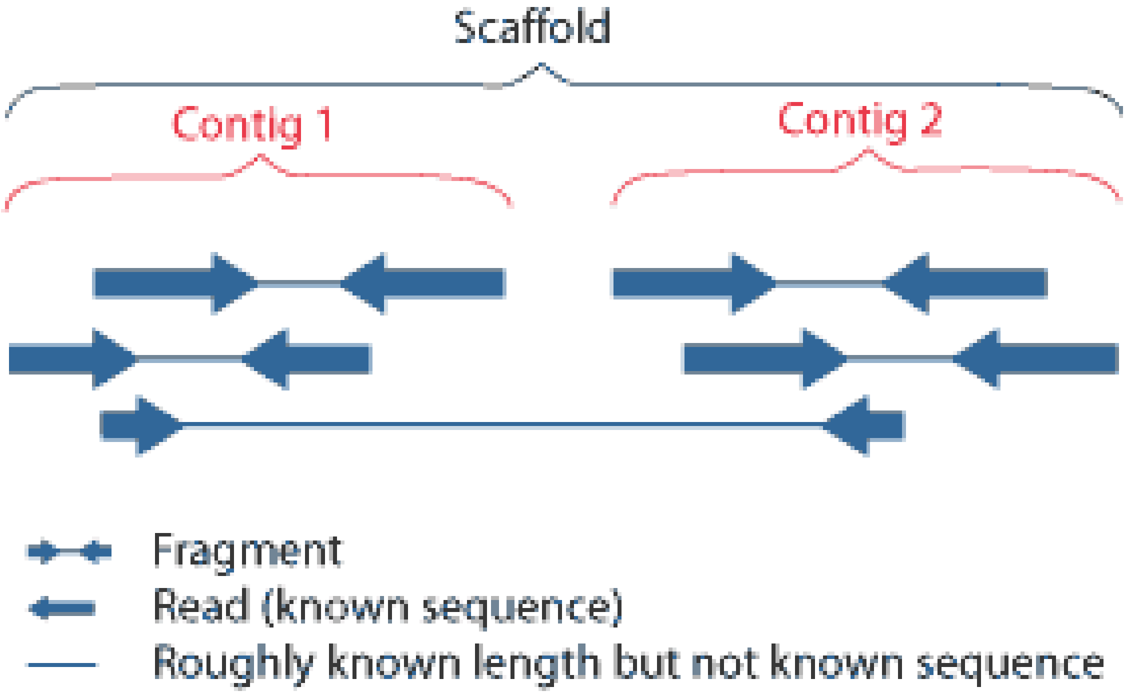
Deselect All

Contig:
- set of overlapping DNA segments
- the more sequences overlap the better

- **Coverage**
  - number of reads mapped
  - the higher the better

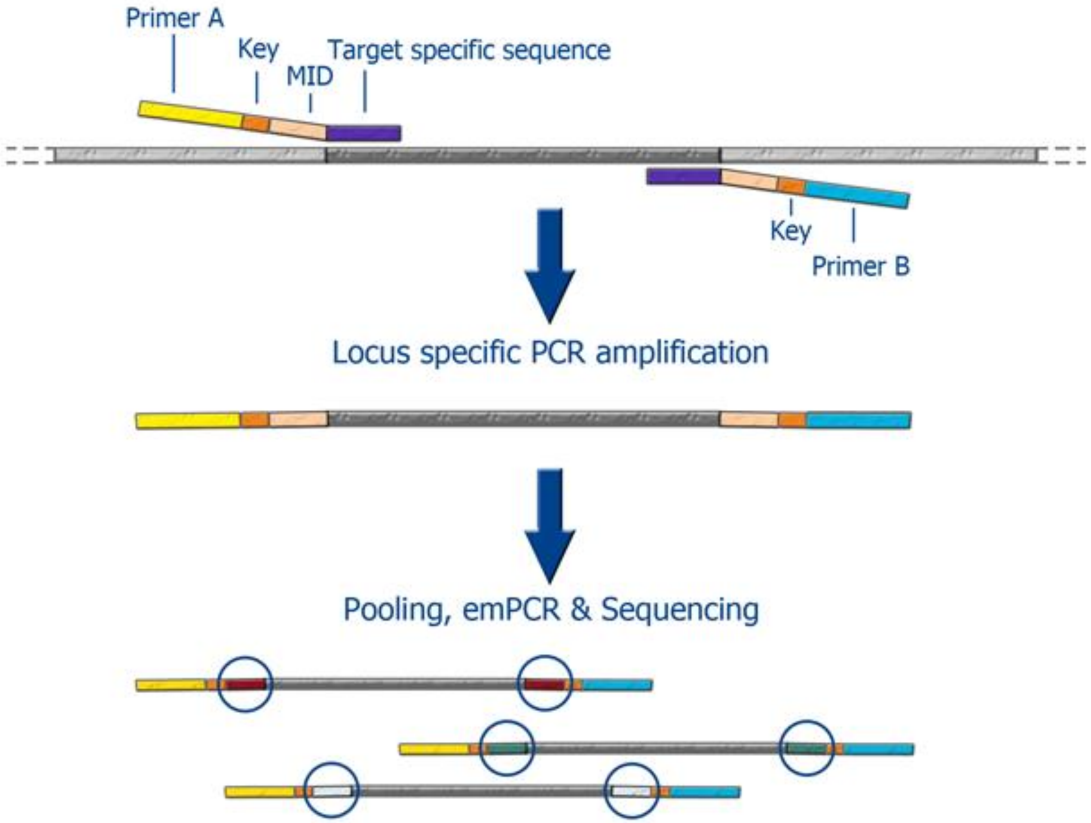
Current generation sequencing: De novo sequencing

- **Scaffolding**
- **overlappin**



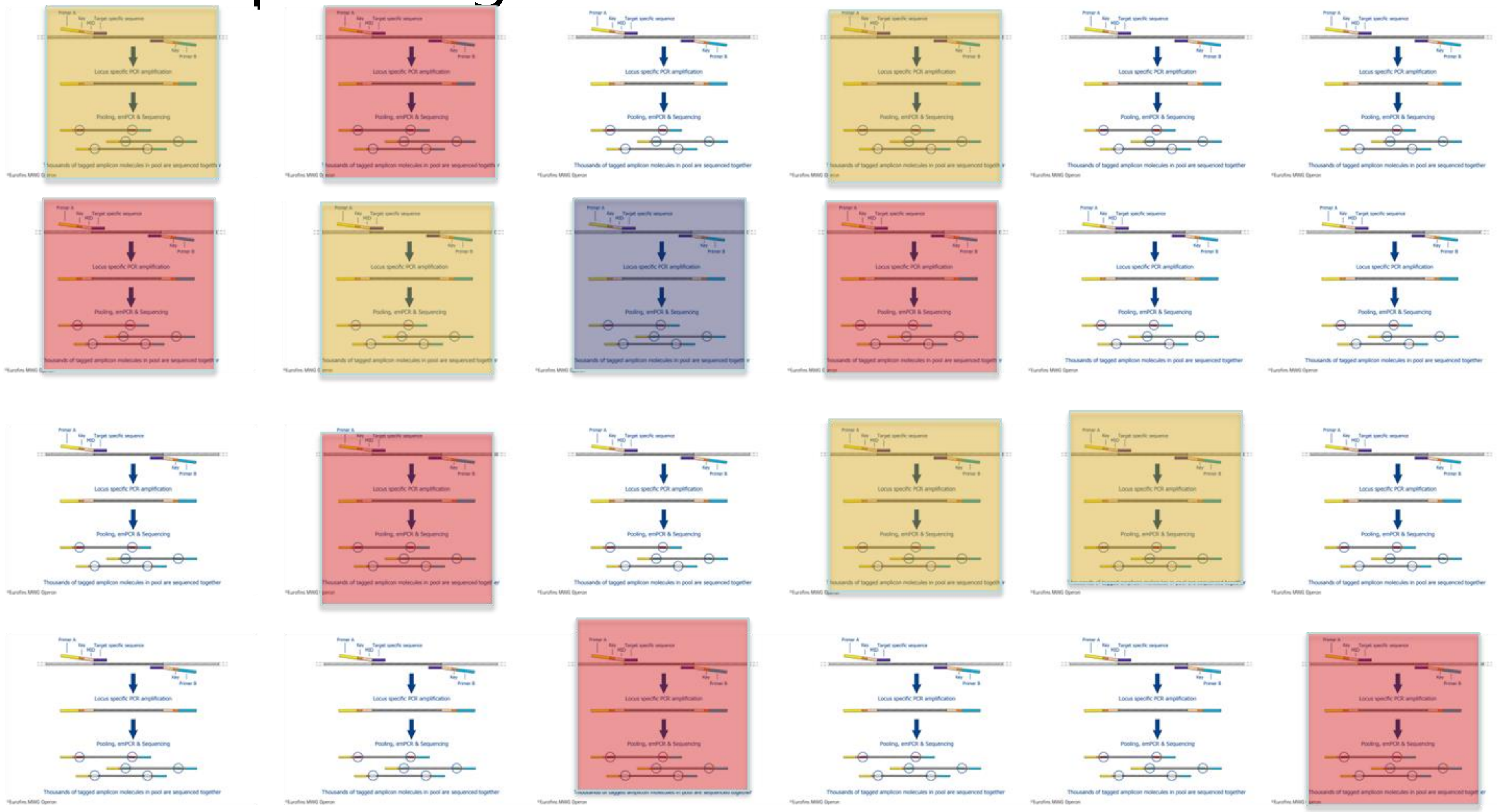
Amplicon-Sequencing

Amplicon Sequencing

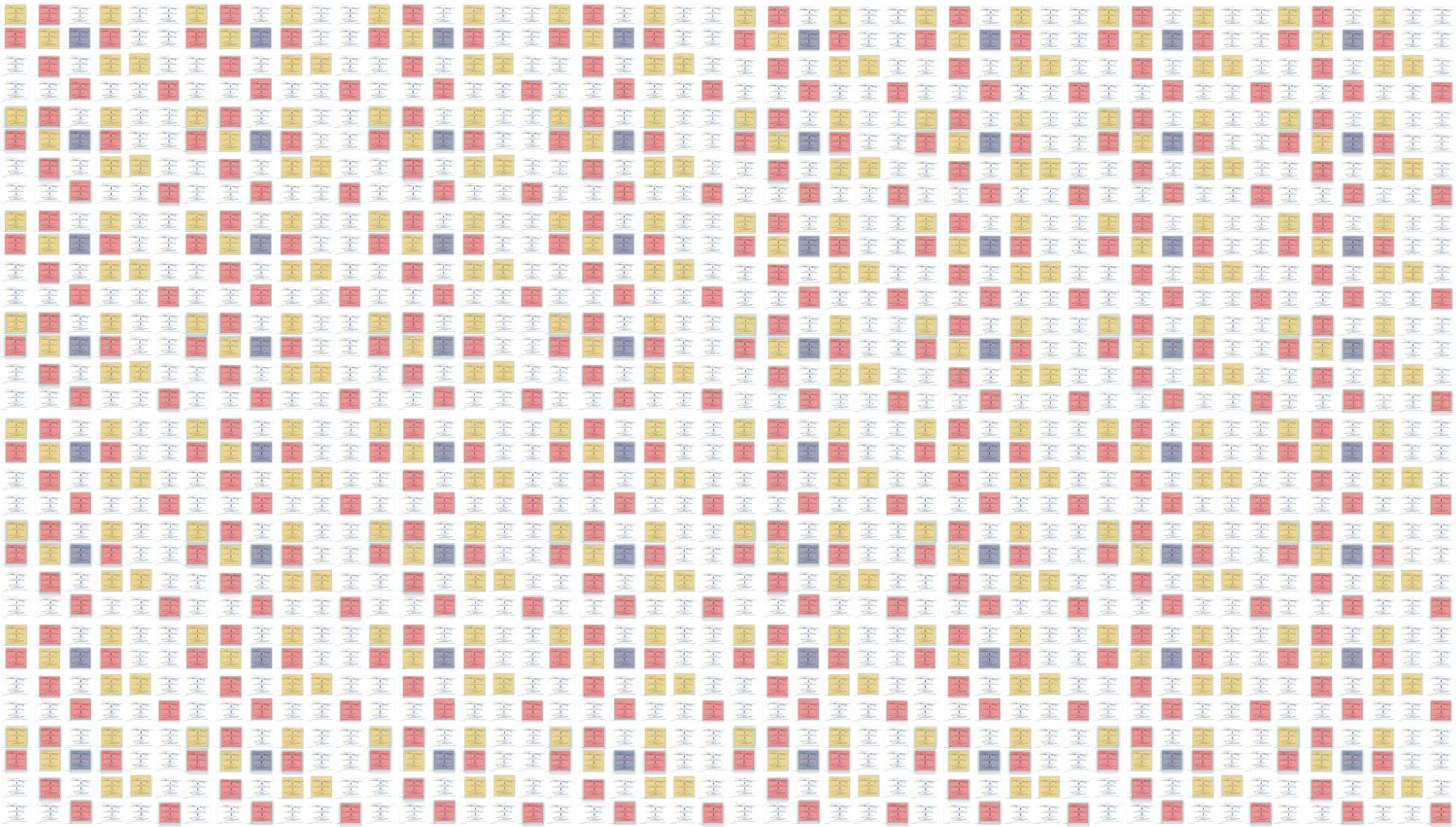


©Eurofins MWG Operon

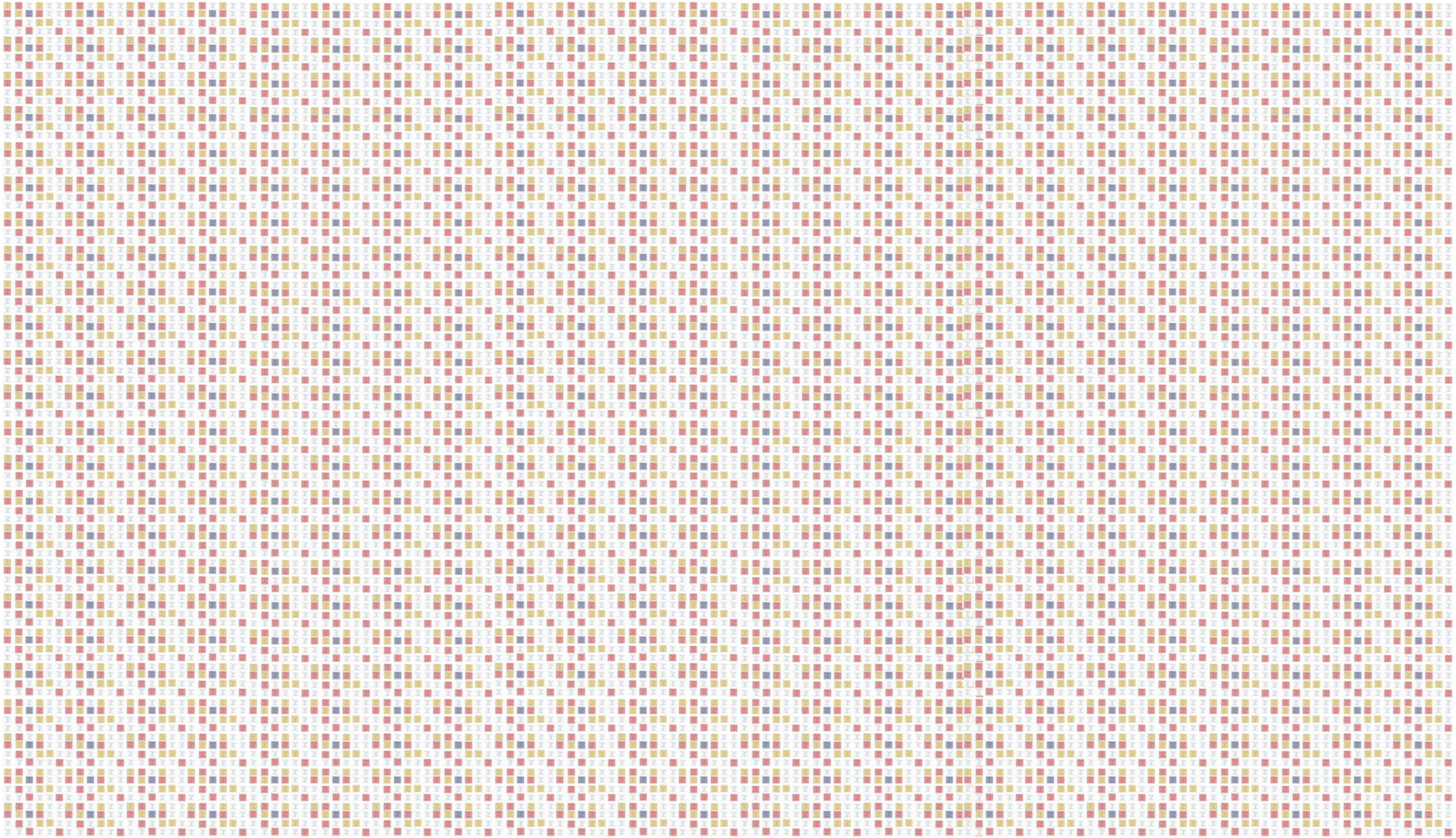
Amplicon Sequencing



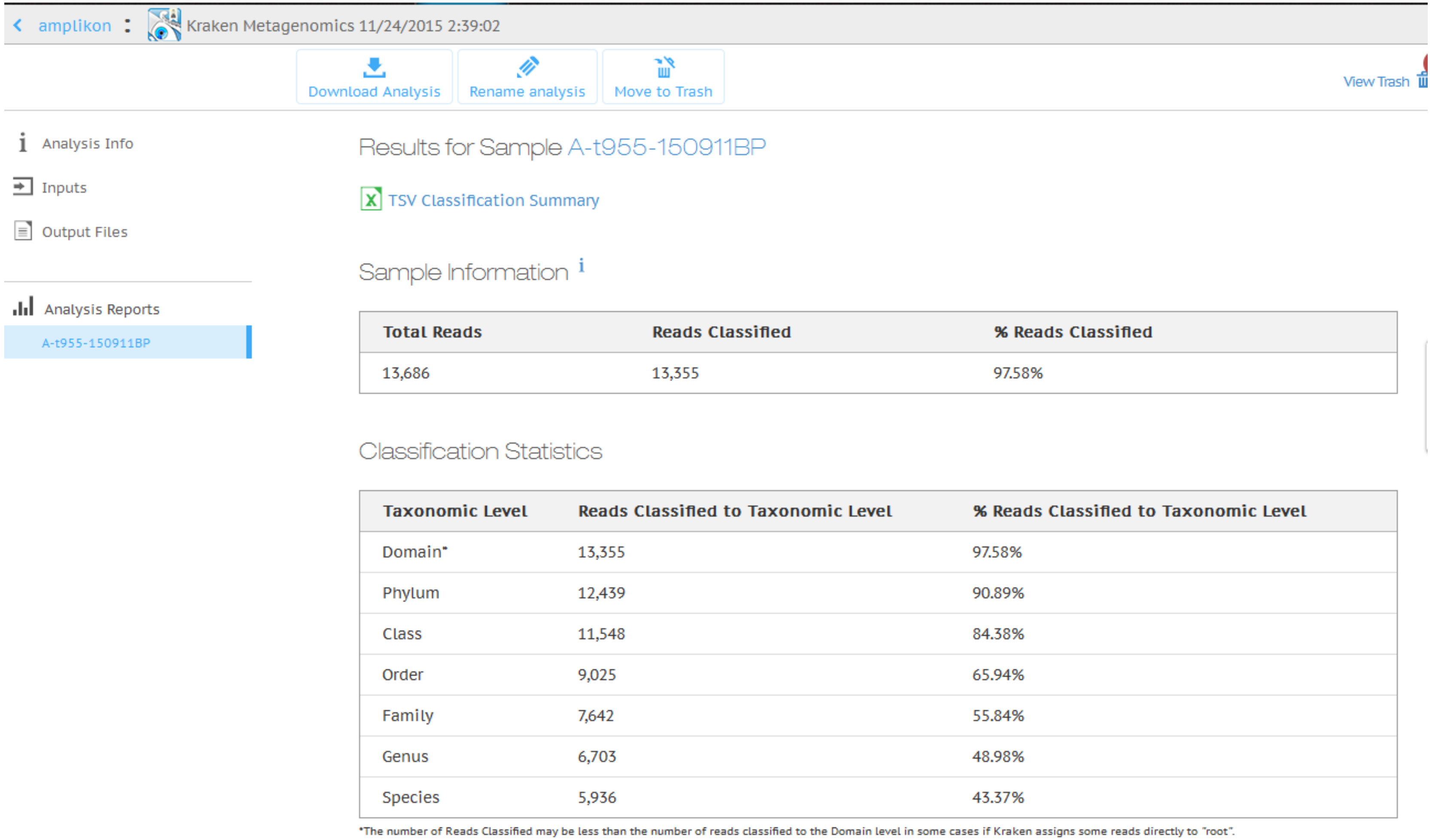
Amplicon Sequencing



Amplicon Sequencing



Amplicon Sequencing



Results for Sample A-t955-150911BP

Sample Information

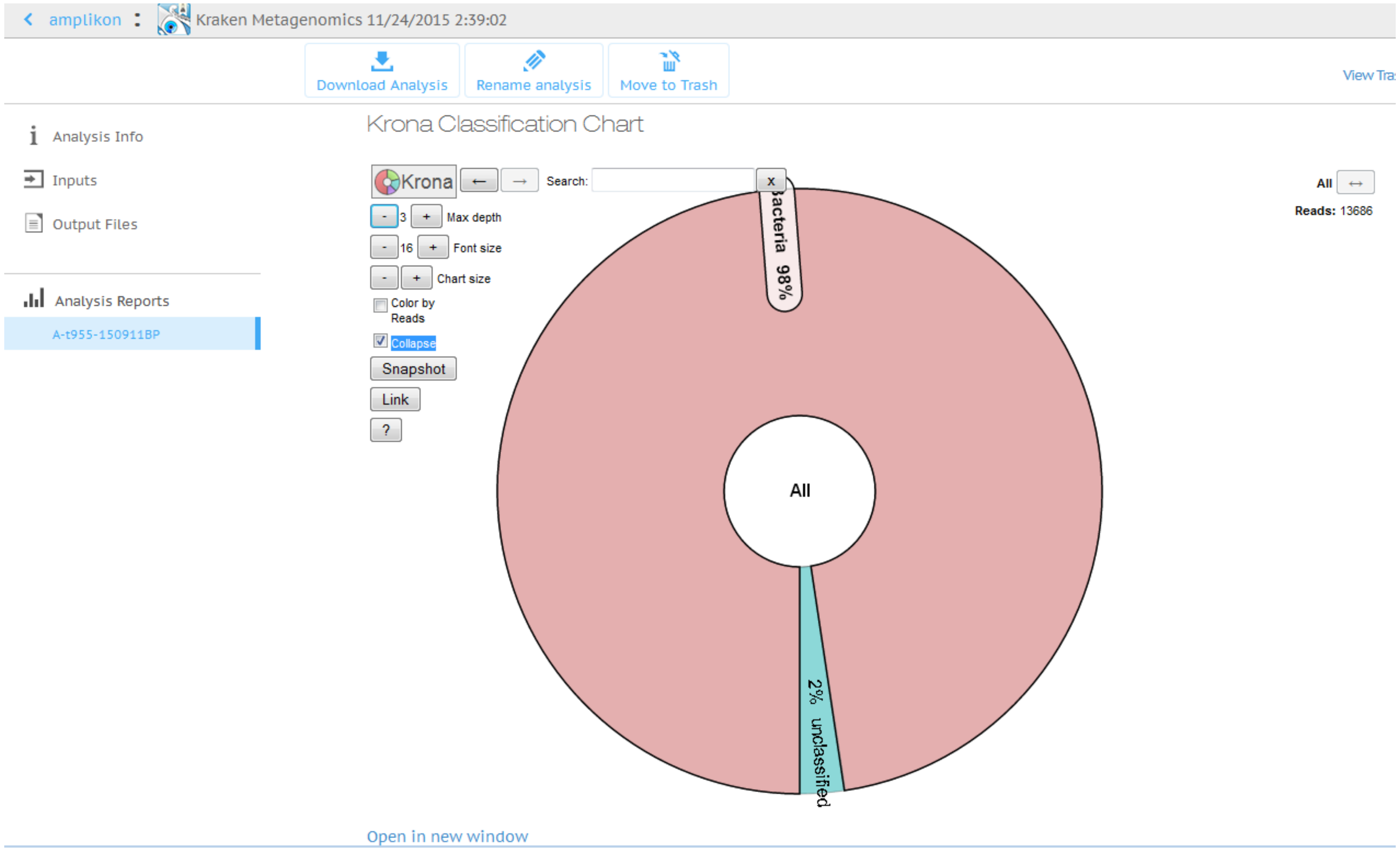
Total Reads	Reads Classified	% Reads Classified
13,686	13,355	97.58%

Classification Statistics

Taxonomic Level	Reads Classified to Taxonomic Level	% Reads Classified to Taxonomic Level
Domain*	13,355	97.58%
Phylum	12,439	90.89%
Class	11,548	84.38%
Order	9,025	65.94%
Family	7,642	55.84%
Genus	6,703	48.98%
Species	5,936	43.37%

*The number of Reads Classified may be less than the number of reads classified to the Domain level in some cases if Kraken assigns some reads directly to "root".

Amplicon Sequencing



amplikon : Kraken Metagenomics 11/24/2015 2:39:02

Krona Classification Chart

3 Max depth

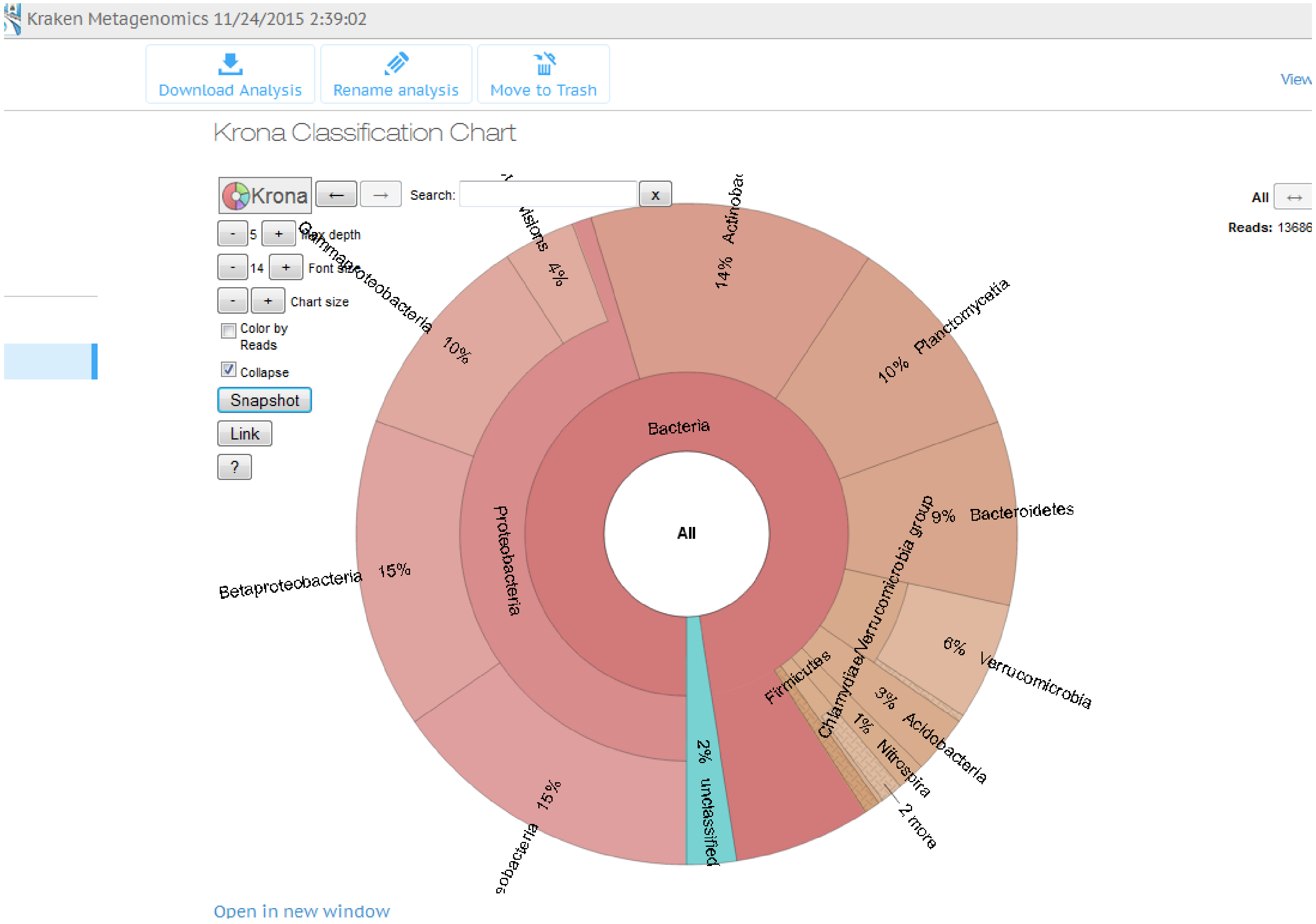
16 Font size

☐ Color by Reads

☑ Collapse

Reads: 13686

Amplicon Sequencing



Kraken Metagenomics 11/24/2015 2:39:02

Krona Classification Chart

5 Max depth

14 Font size

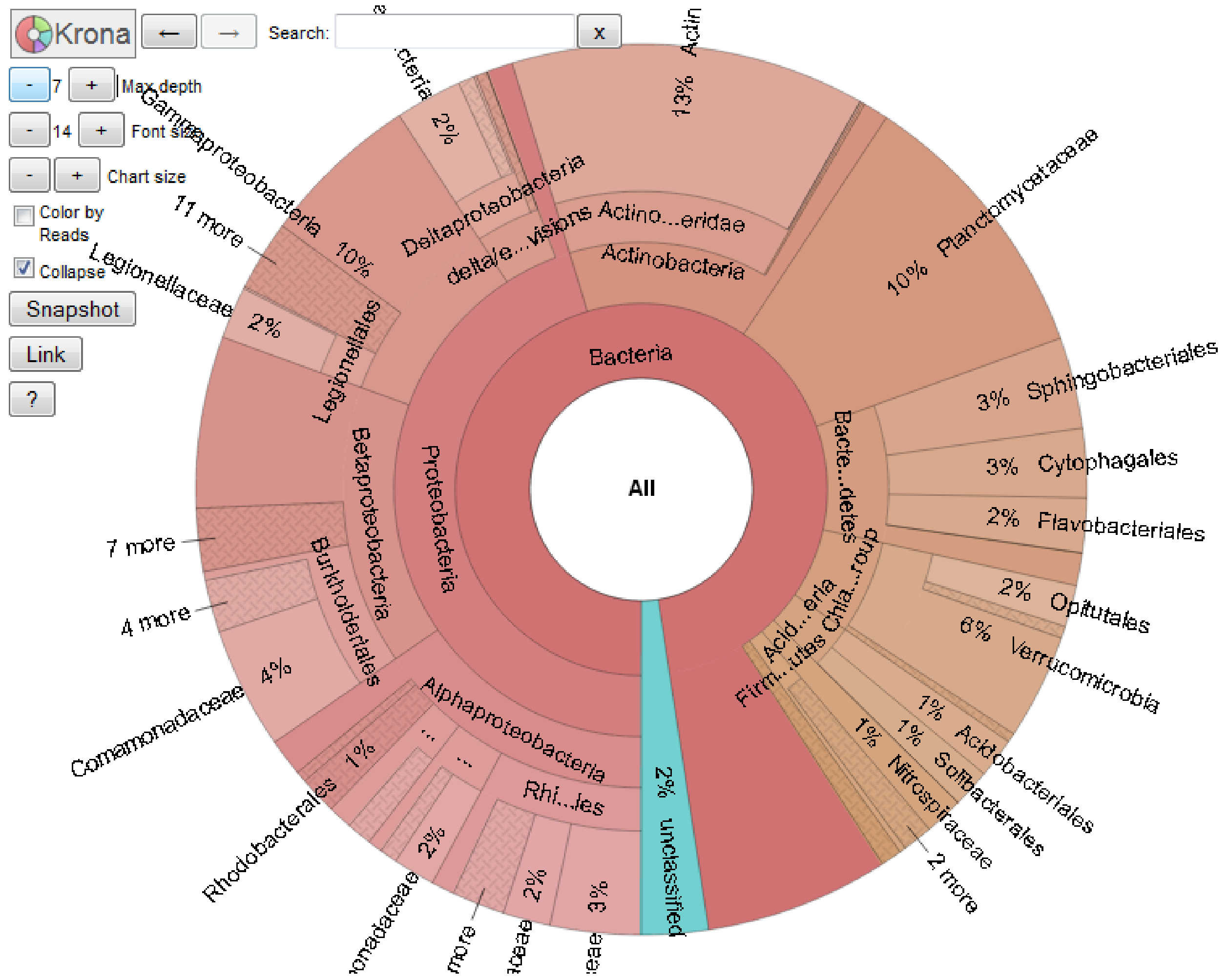
☐ Color by Reads

☑ Collapse

Reads: 13686

Amplicon Sequencing

Krona Classification Chart



Open in new window

Amplicon Sequencing

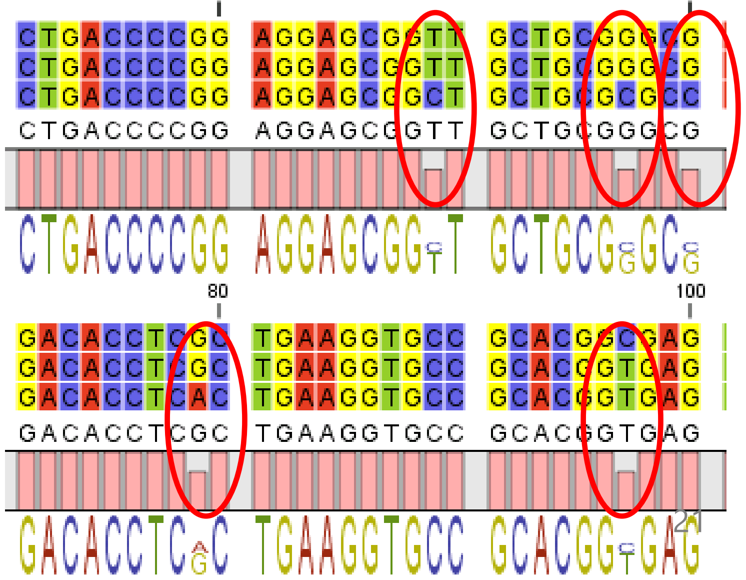
Current generation sequencing: Resequencing

- **Mapping of NGS reads to a known reference genome sequence**
- **As in de-novo-sequencing coverage is important**
- **Allelic variations (Gene variants) can be found for genes calling**
  - **Single nucleotide polymorphism (variants) (SNP or SNV)**
  - **Deletions/Insertions**

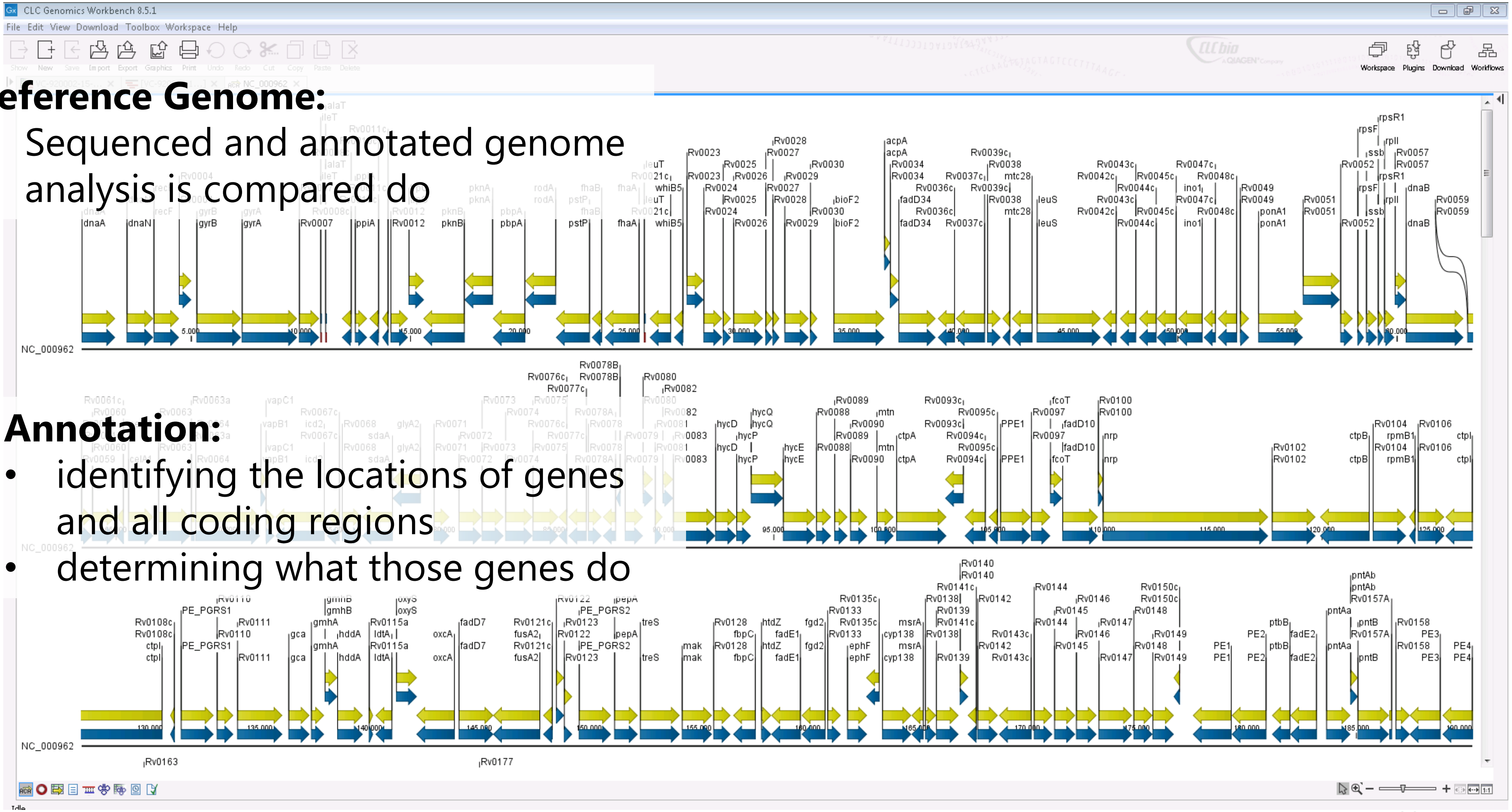
Current generation sequencing: Resequencing

- **Genome wide mutation calling (SNP-calling)**
  - Single nucleotide polymorphism (variants) (SNP or SNV)
    - two or more chromosomes differ by one or more single nucleotide DNA sequence variations
  - Best applicable within clonal populations
  - One reliable complete genome needed
  - Comparison has do be done against the same chromosome
    - SNP occur with every cell cycle
    - Every isolate is different

Source: A. Indra



Current generation sequencing: Resequencing



Reference Genome:

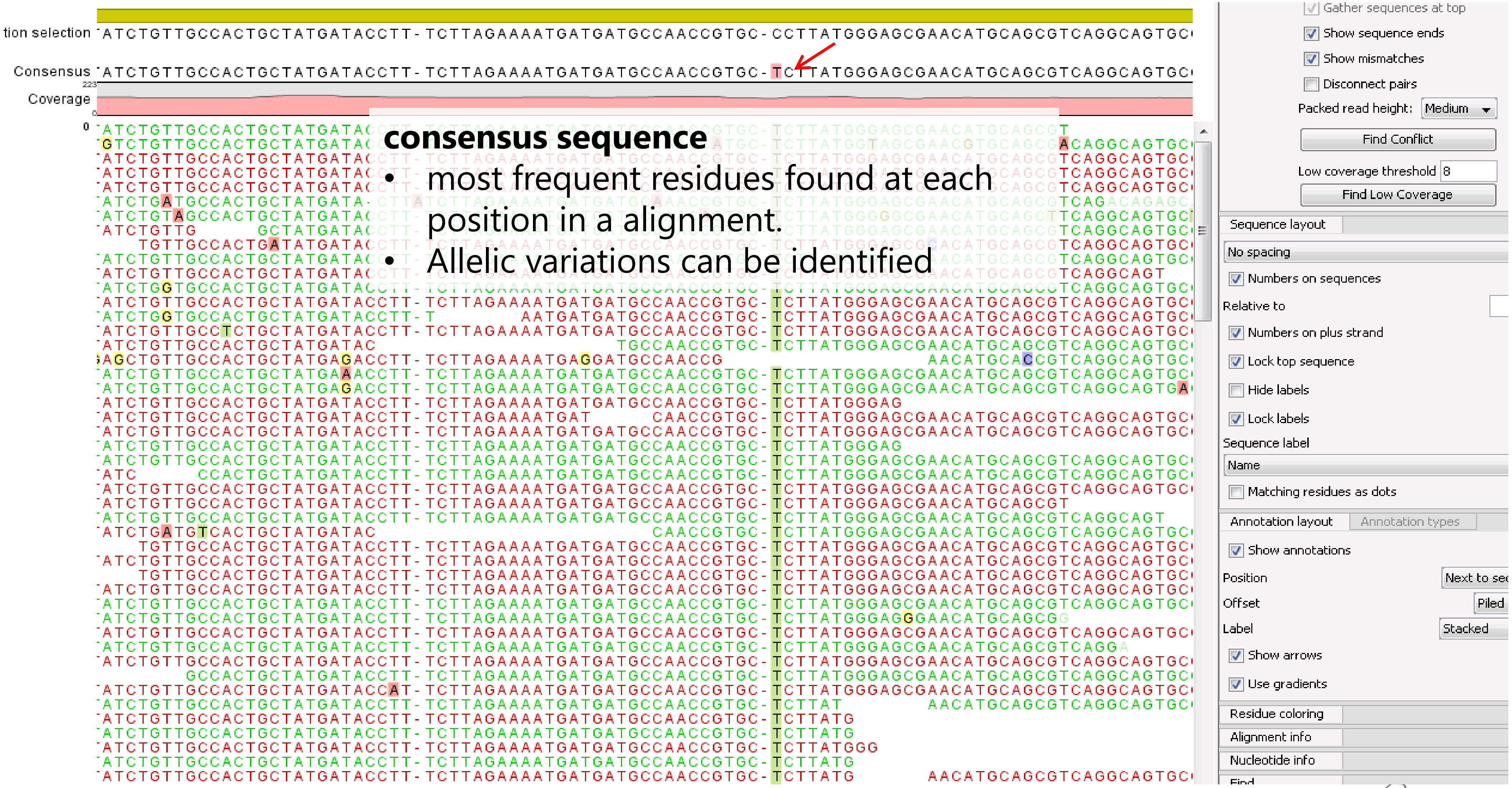
- Sequenced and annotated genome analysis is compared do

Annotation:

- identifying the locations of genes and all coding regions
- determining what those genes do

Source: A. Indra

Current generation sequencing: Resequencing



```
tion selection  ATCTGTTGCCACTGCTATGATACCTT-TCTTAGAAAATGATGATGCCAACCGTGC-CCTTATGGGAGCGAACATGCAGCGTCAGGCAGTGC
Consensus       ATCTGTTGCCACTGCTATGATACCTT-TCTTAGAAAATGATGATGCCAACCGTGC-TCTTATGGGAGCGAACATGCAGCGTCAGGCAGTGC
Coverage
```

☑ Gather sequences at top
☑ Show sequence ends
☑ Show mismatches
☐ Disconnect pairs
Packed read height: Medium

Low coverage threshold 8

No spacing
☑ Numbers on sequences
Relative to
☑ Numbers on plus strand
☑ Lock top sequence
☐ Hide labels
☑ Lock labels
Sequence label: Name
☐ Matching residues as dots

☑ Show annotations
Position: Next to se
Offset: Piled
Label: Stacked
☑ Show arrows
☑ Use gradients

consensus sequence

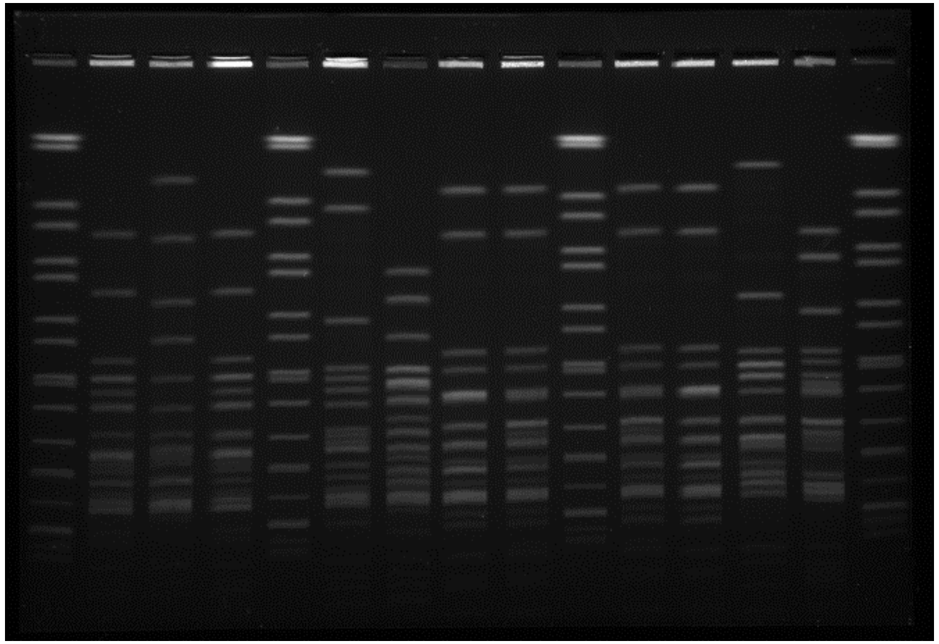
- most frequent residues found at each position in a alignment.
- Allelic variations can be identified

Source: A. Indra

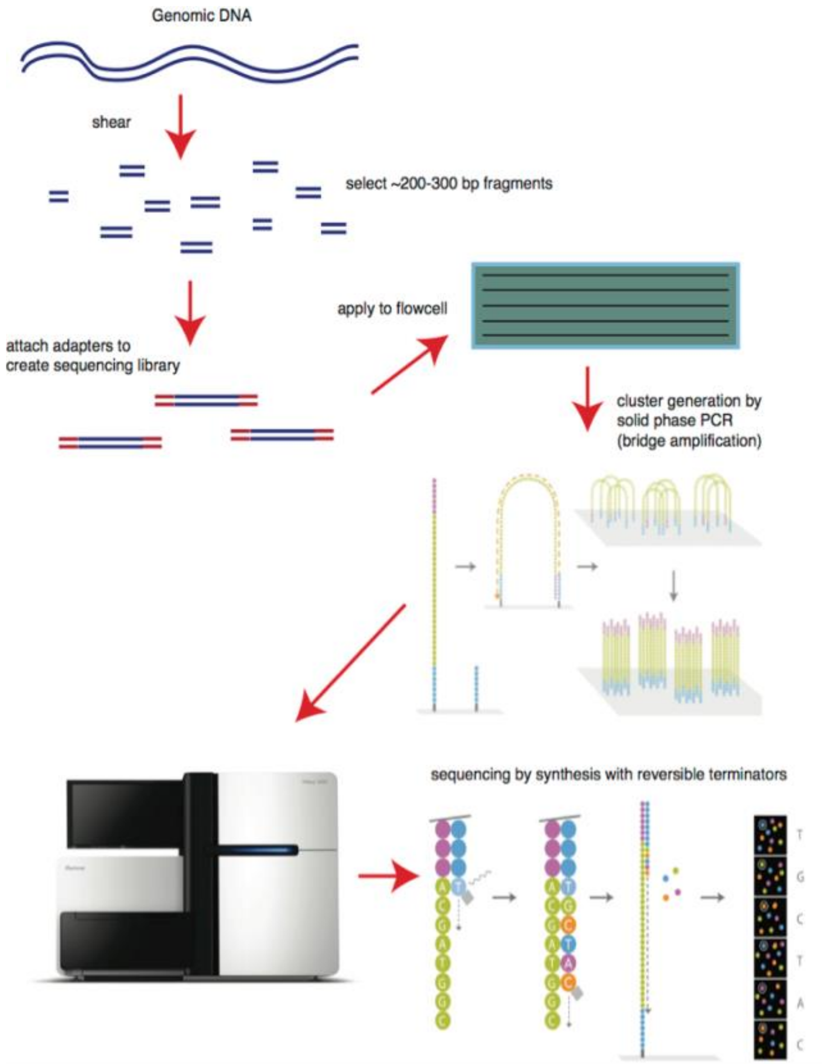
cgMLST Listeria monocytogenes

cgMLST Listeria monocytogenes

- **Source of food bourne outbreaks**
- **PFGE currently standard**
  - Problematic discrematory power



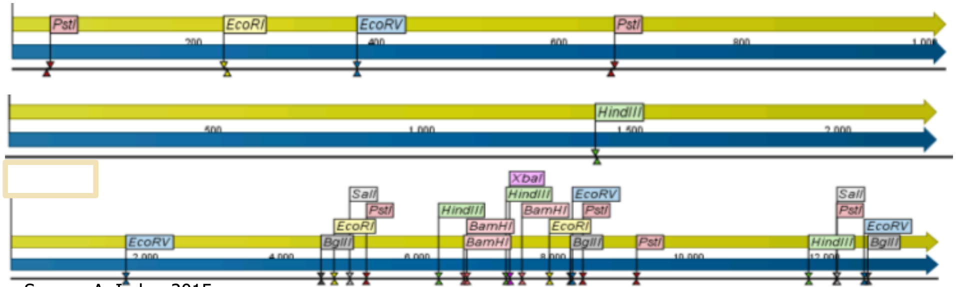
cgMLST
Listeria monocytogenes



http://bitesizebio.com/13546/sequencing-by-synthesis-explaining-the-illumina-sequencing-technology/

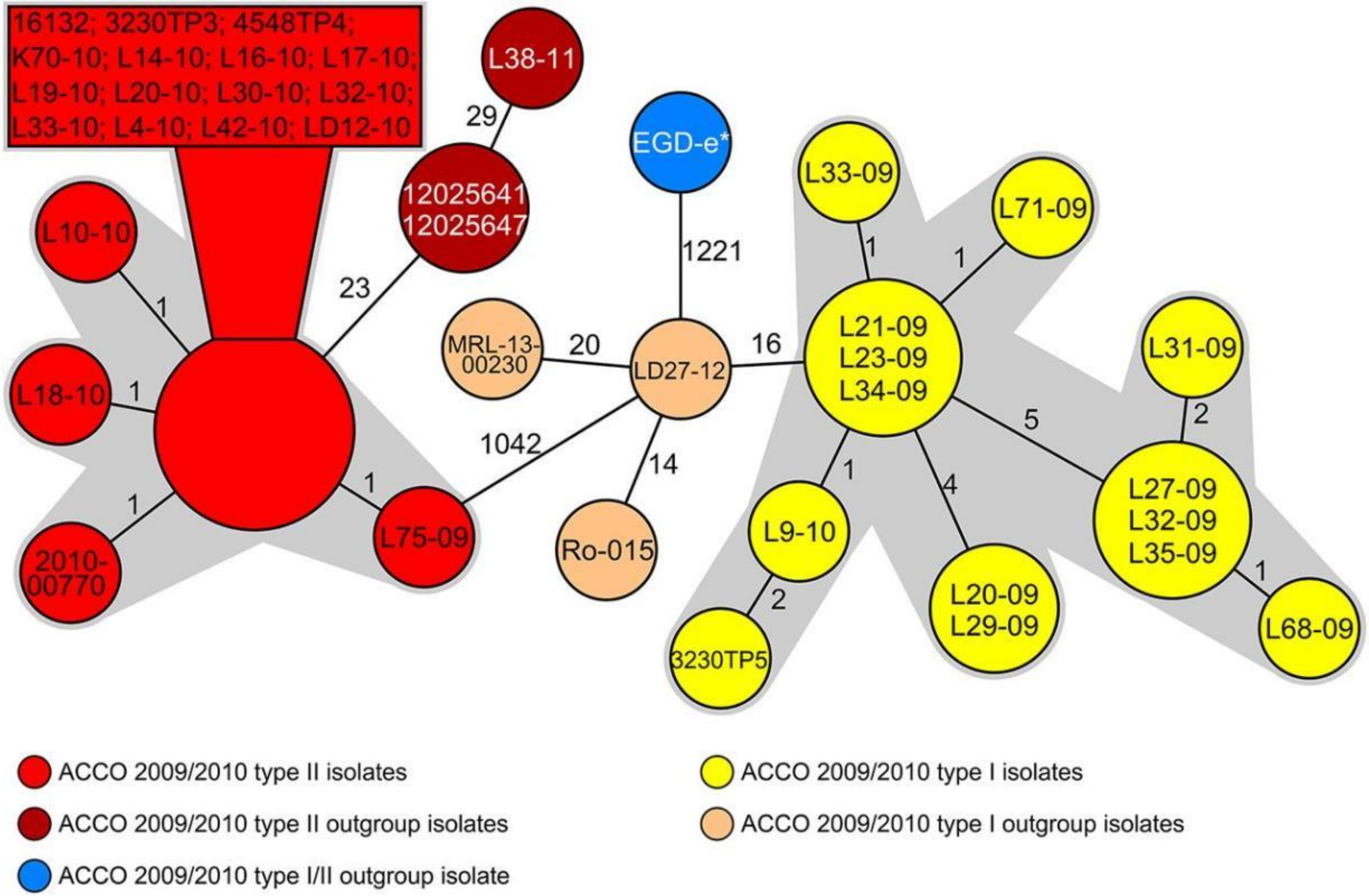
cgMLST
Listeria monocytogenes

- **Allele by allele comparison**
  - Basicly a large MLST
  - Gene variants (allele) are compared against each other
    - Regardless of the number of SNP´s within the gene its is only taken as one new allele
  - A “core” group of genes – defining isolates of the same species – are compared
    - Gives the broader picture
    - Lacks the discriminatory power of SNP and Chromosome comparison
    - Enables a simple nomenclature



Source: A. Indra, 2015

Minimum-spanning tree illustrating the phylogenetic relationship based on the cgMLST allelic profiles of 33 L. monocytogenes isolates from the outbreak associated with acid curd cheese (ACCO) (22, 23) consisting of two clones (ACCO I and ACCO II).



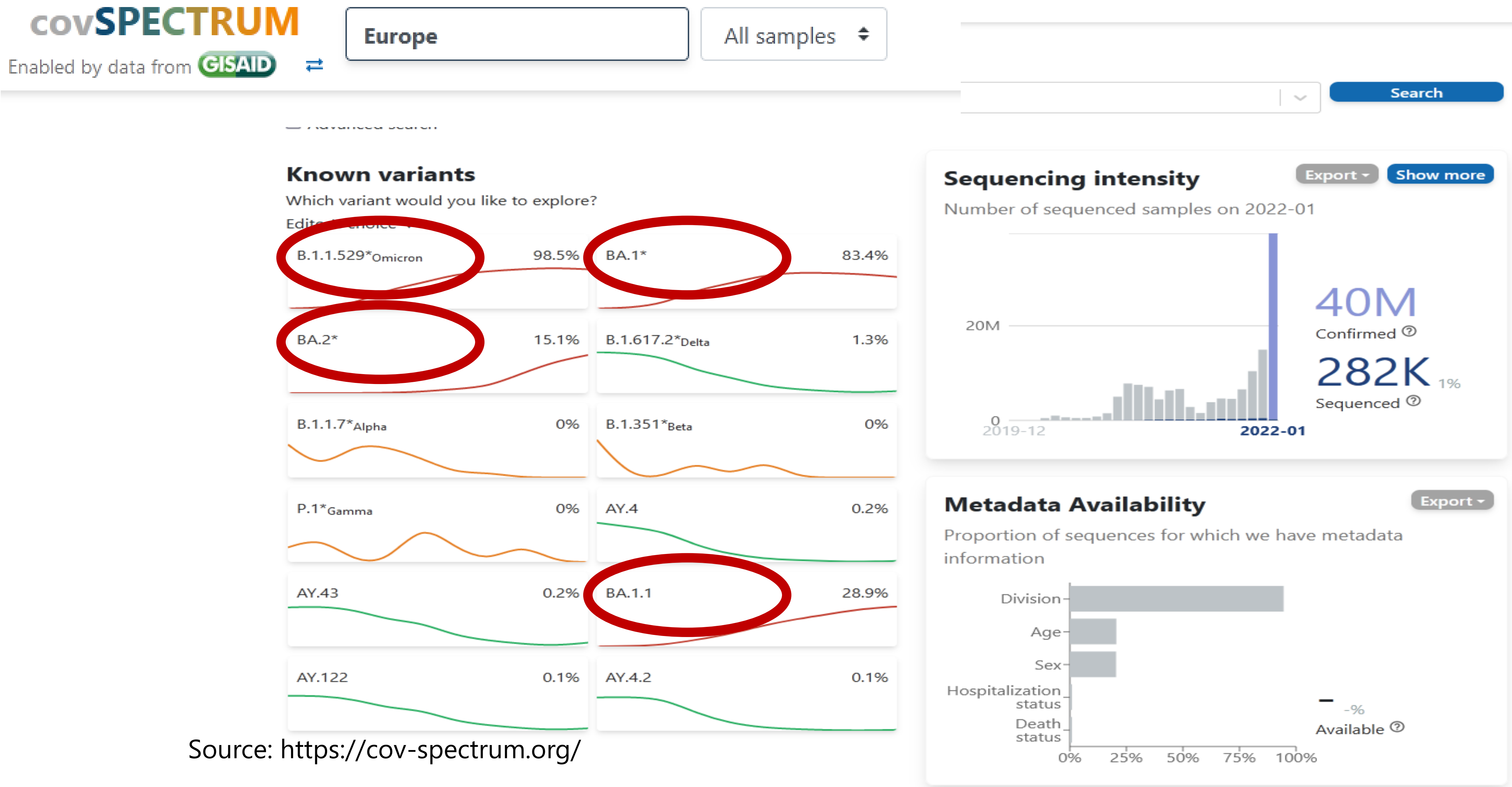
Werner Ruppitsch et al. J. Clin. Microbiol. 2015;53:2869-2876

Journal of Clinical Microbiology

Journals.ASM.org

Future generation Sequencing?

MinION MkI: portable, real-time biological analyses



Known variants

Which variant would you like to explore?

Edit

Variant	Share
B.1.1.529* Omicron	98.5%
BA.1*	83.4%
BA.2*	15.1%
B.1.617.2* Delta	1.3%
B.1.1.7* Alpha	0%
B.1.351* Beta	0%
P.1* Gamma	0%
AY.4	0.2%
AY.43	0.2%
BA.1.1	28.9%
AY.122	0.1%
AY.4.2	0.1%

Sequencing intensity

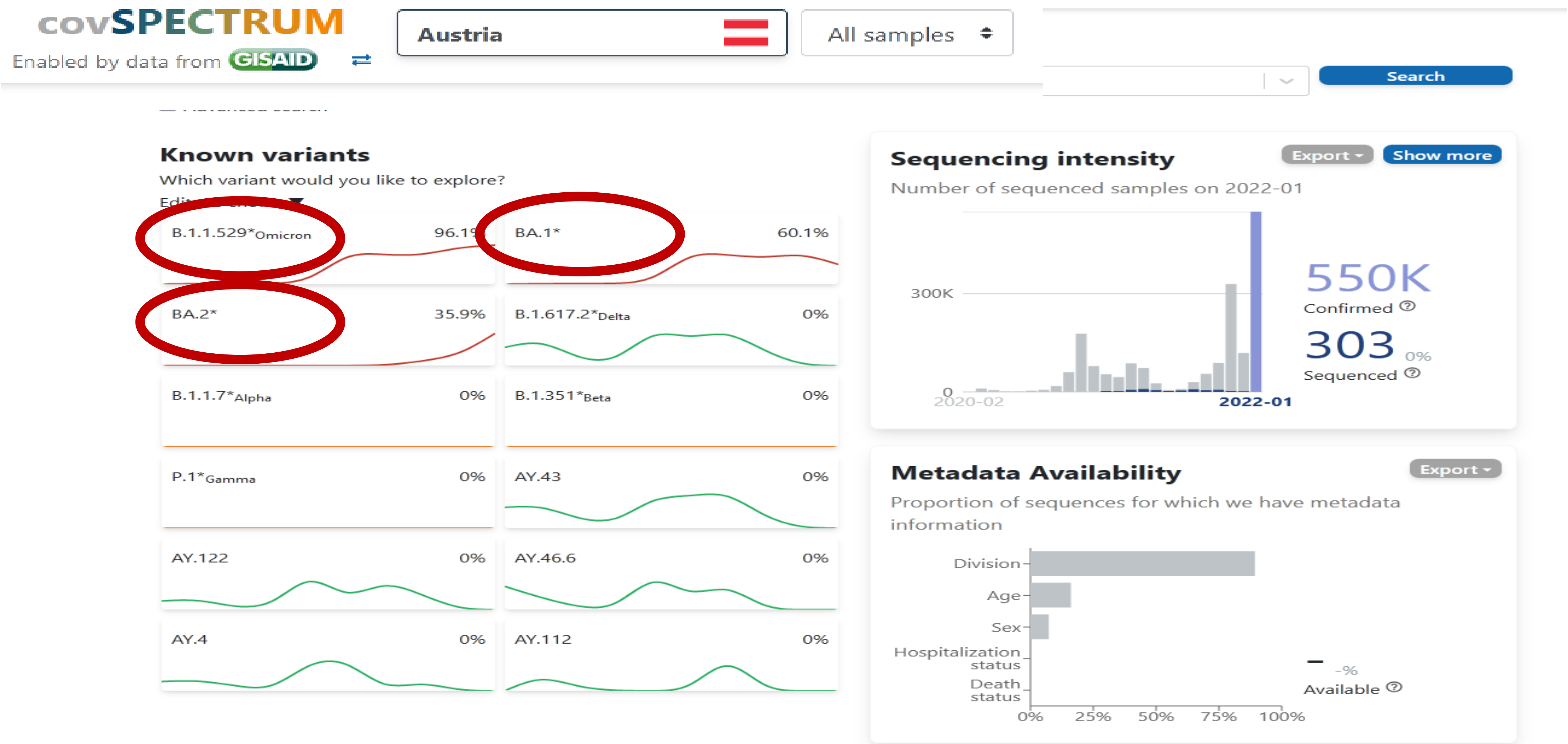
Number of sequenced samples on 2022-01

Metadata Availability

Proportion of sequences for which we have metadata information

Source: https://cov-spectrum.org/

covSpectrum Austria



Austria

All samples

Known variants

Which variant would you like to explore?

B.1.1.529*Omicron 96.1

BA.1* 60.1%

BA.2* 35.9%

B.1.617.2*Delta 0%

B.1.1.7*Alpha 0%

B.1.351*Beta 0%

P.1*Gamma 0%

AY.43 0%

AY.122 0%

AY.46.6 0%

AY.4 0%

AY.112 0%

Sequencing intensity

Number of sequenced samples on 2022-01

550K Confirmed

303 0% Sequenced

Metadata Availability

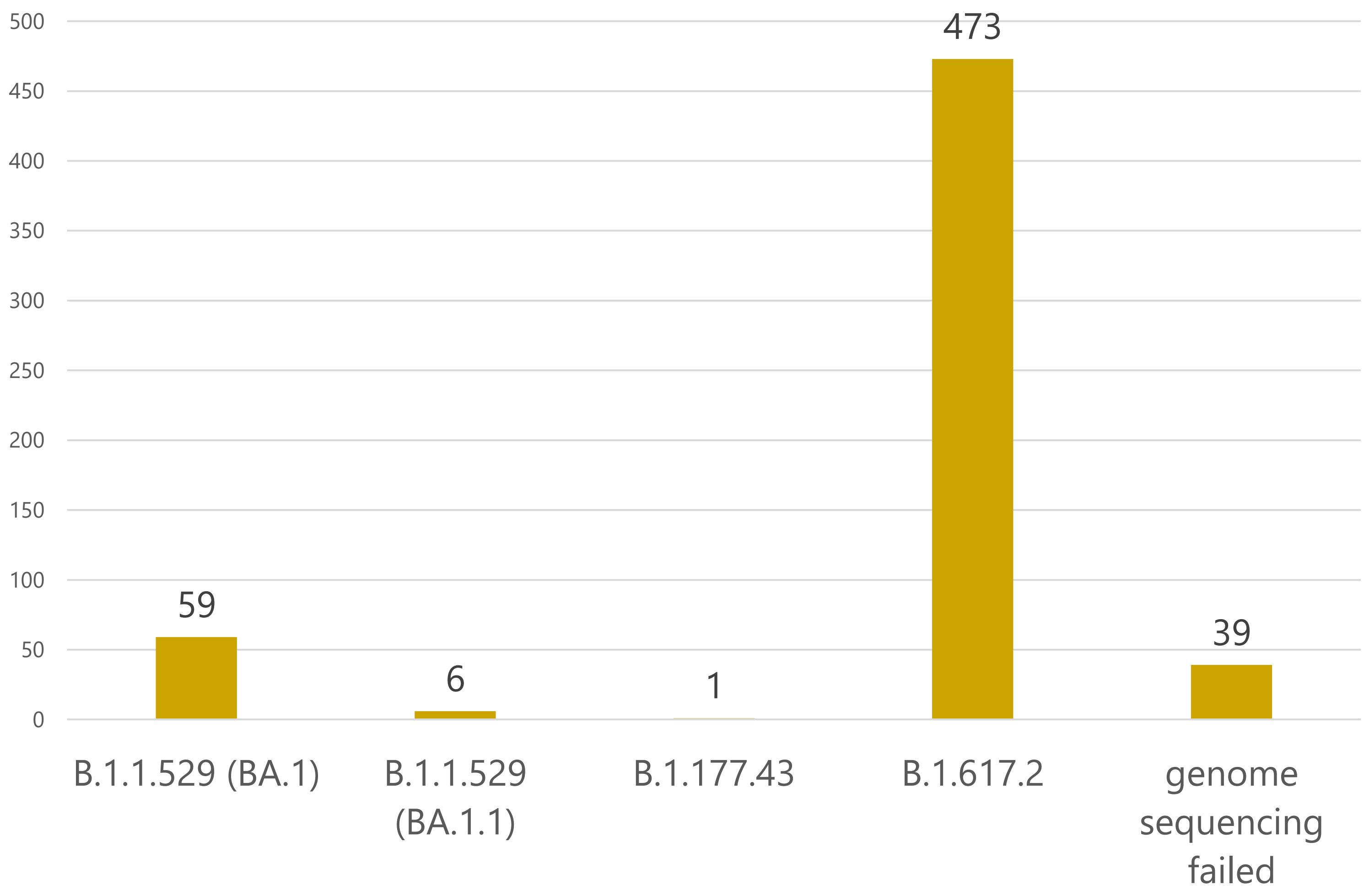
Proportion of sequences for which we have metadata information

– -% Available

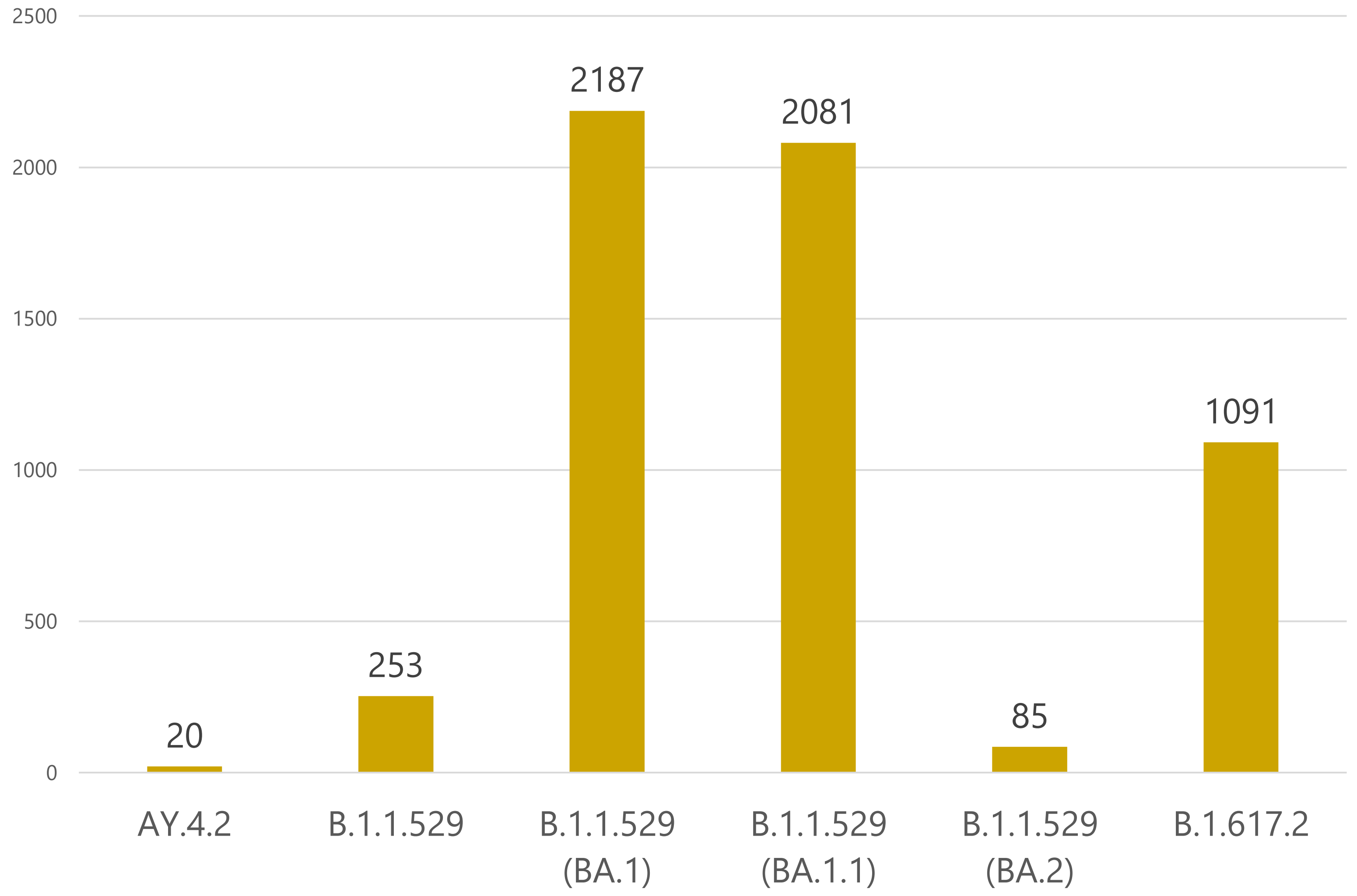
Sentinel System Austria

- 10 samples per 9 Districts and 1 Pool from the Vienna Airport
  - 100 samples per Week
  - Sequenced in cooperation with CeMM
- Samples are collected in participating lab and send to us in the next week
  - Sequenzing is done afterward using Illumina System at CeMM
  - Results finished 6 to 10 days later

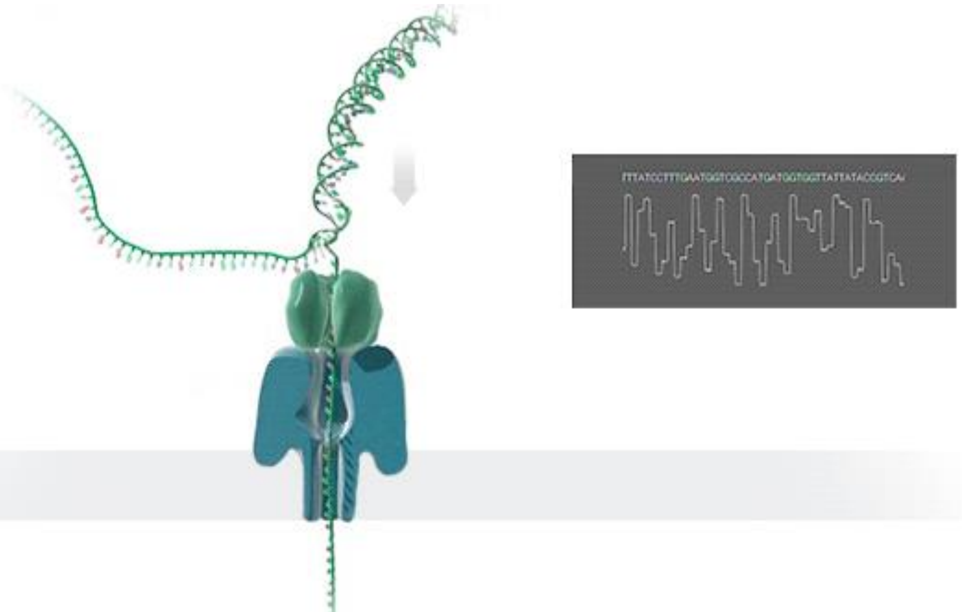
Sentinel Results week 49/2021 to 2/2022



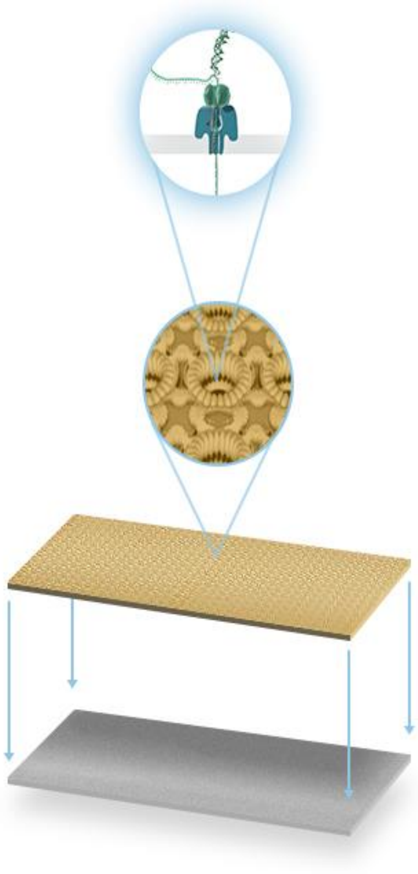
SARS-Seq Results 2022



Nanopore Pipeline AGES



https://www.nanoporetech.com/science-technology/how-it-works



https://www.nanoporetech.com/science-technology/how-it-works

Nanopore Pipeline AGES

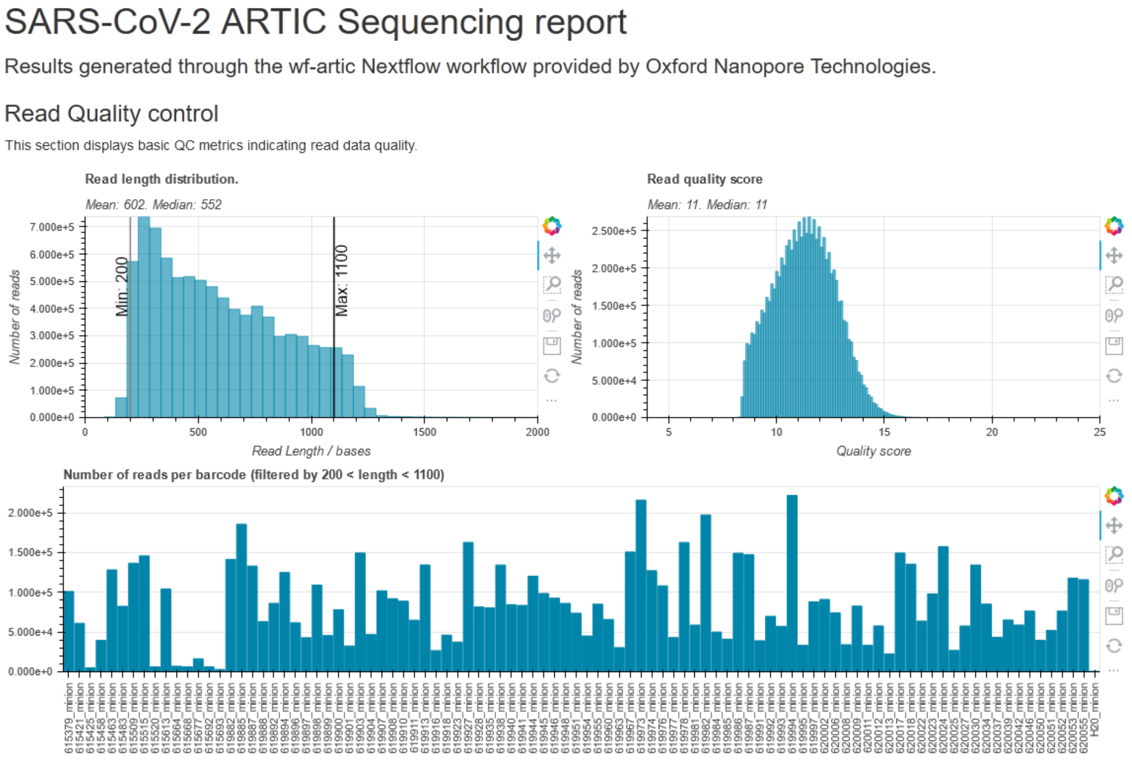
- 2 Oxford Nanopore Minion Mk1c
- Updated Workflow using for seamless analysis
  - Repeated Software errors (Bascalling during analysis) within original application made System unstable
  - errors AGES had to establish new analysis Workflow within Mk1c
  - Since then no analysis error was encounterd
- Nanopore Midnight Kit
  - 96 samples/Flowcell
  - 8 houres Labtime + 24 hours analysis
  - Material costs per sample 22€ including RNA-extraction
  - Lillimina costs ar 2-4 times higher

Source: https://www.nanoporetech.com

Nanopore Pipeline AGES

- Analysis Pipeline
  - Adaped from Nanopore suggested pipeline
  - ARTIC SARS-CoV-2 Workflow
    - Primer Trimming
    - Variant Calling
    - Consensus Sequenz (Medaka)
    - Lineage Assinment (Pangolin)
    - HTML Report
  - Additional AGES specific
    - SARS-CoV-2 Analysis Pipeline
    - VariantCalling (Ivar Version 1.3.1)
    - Coverage (Samtools Version1.12)
    - additional reports are generated

Nanopore Pipeline AGES

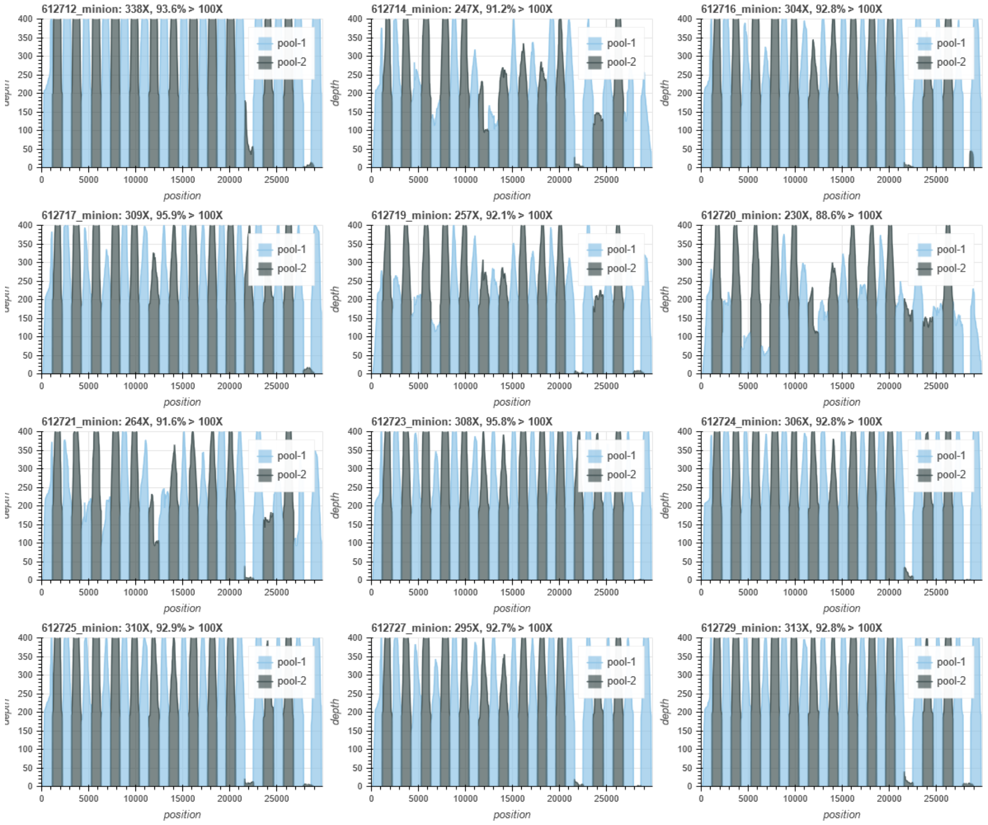


SARS-CoV-2 ARTIC Sequencing report

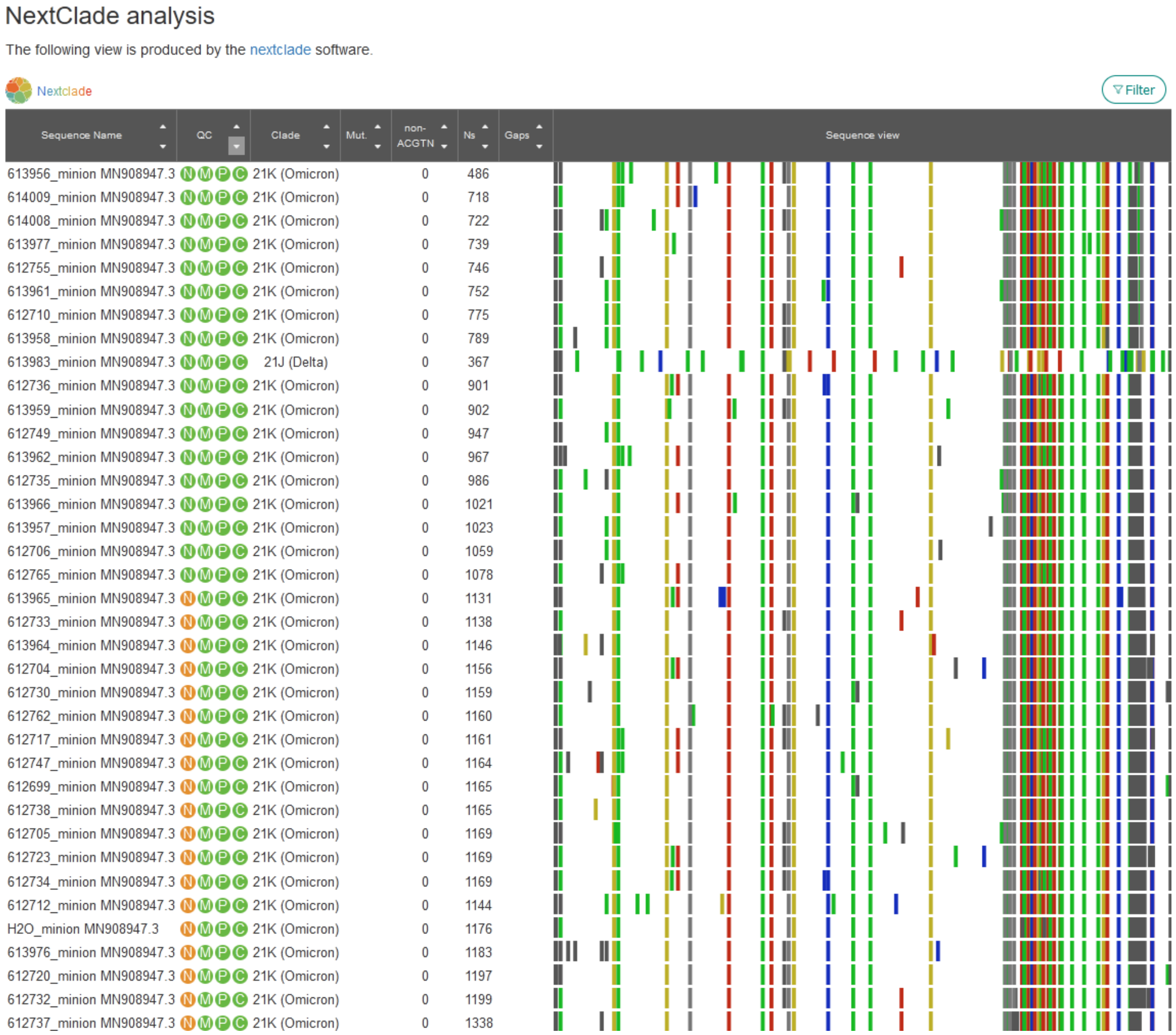
Results generated through the wf-artic Nextflow workflow provided by Oxford Nanopore Technologies.

Read Quality control

This section displays basic QC metrics indicating read data quality.



Nanopore Pipeline AGES

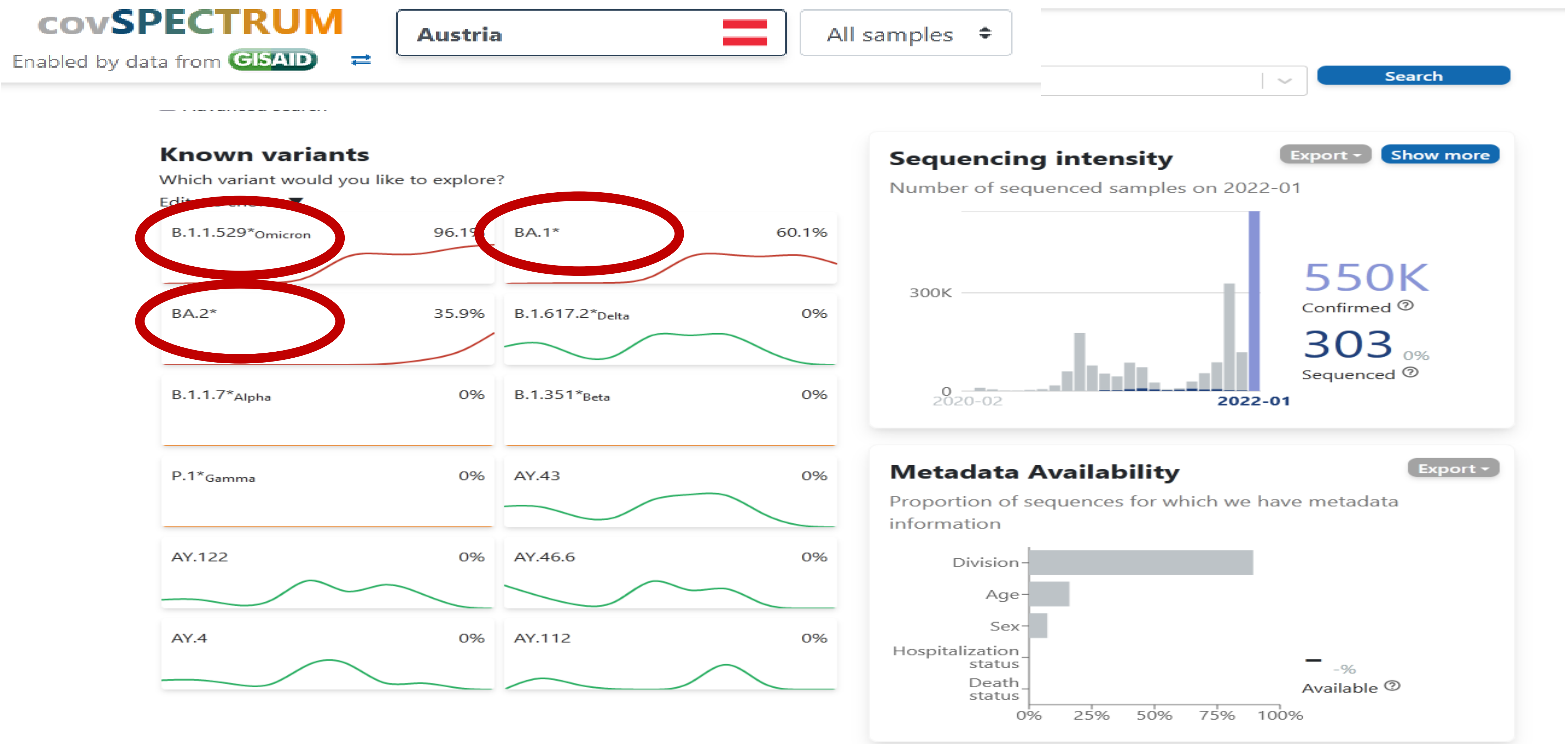


NextClade analysis

The following view is produced by the nextclade software.

Sequence Name	QC	Clade	Mut.	non-ACGTN	Ns	Gaps	Sequence view
613956_minion MN908947.3	N M P C	21K (Omicron)		0	486		
614009_minion MN908947.3	N M P C	21K (Omicron)		0	718		
614008_minion MN908947.3	N M P C	21K (Omicron)		0	722		
613977_minion MN908947.3	N M P C	21K (Omicron)		0	739		
612755_minion MN908947.3	N M P C	21K (Omicron)		0	746		
613961_minion MN908947.3	N M P C	21K (Omicron)		0	752		
612710_minion MN908947.3	N M P C	21K (Omicron)		0	775		
613958_minion MN908947.3	N M P C	21K (Omicron)		0	789		
613983_minion MN908947.3	N M P C	21J (Delta)		0	367		
612736_minion MN908947.3	N M P C	21K (Omicron)		0	901		
613959_minion MN908947.3	N M P C	21K (Omicron)		0	902		
612749_minion MN908947.3	N M P C	21K (Omicron)		0	947		
613962_minion MN908947.3	N M P C	21K (Omicron)		0	967		
612735_minion MN908947.3	N M P C	21K (Omicron)		0	986		
613966_minion MN908947.3	N M P C	21K (Omicron)		0	1021		
613957_minion MN908947.3	N M P C	21K (Omicron)		0	1023		
612706_minion MN908947.3	N M P C	21K (Omicron)		0	1059		
612765_minion MN908947.3	N M P C	21K (Omicron)		0	1078		
613965_minion MN908947.3	N M P C	21K (Omicron)		0	1131		
612733_minion MN908947.3	N M P C	21K (Omicron)		0	1138		
613964_minion MN908947.3	N M P C	21K (Omicron)		0	1146		
612704_minion MN908947.3	N M P C	21K (Omicron)		0	1156		
612730_minion MN908947.3	N M P C	21K (Omicron)		0	1159		
612762_minion MN908947.3	N M P C	21K (Omicron)		0	1160		
612717_minion MN908947.3	N M P C	21K (Omicron)		0	1161		
612747_minion MN908947.3	N M P C	21K (Omicron)		0	1164		
612699_minion MN908947.3	N M P C	21K (Omicron)		0	1165		
612738_minion MN908947.3	N M P C	21K (Omicron)		0	1165		
612705_minion MN908947.3	N M P C	21K (Omicron)		0	1169		
612723_minion MN908947.3	N M P C	21K (Omicron)		0	1169		
612734_minion MN908947.3	N M P C	21K (Omicron)		0	1169		
612712_minion MN908947.3	N M P C	21K (Omicron)		0	1144		
H2O_minion MN908947.3	N M P C	21K (Omicron)		0	1176		
613976_minion MN908947.3	N M P C	21K (Omicron)		0	1183		
612720_minion MN908947.3	N M P C	21K (Omicron)		0	1197		
612732_minion MN908947.3	N M P C	21K (Omicron)		0	1199		
612737_minion MN908947.3	N M P C	21K (Omicron)		0	1338		

covSpectrum Austria



Austria 🟥

All samples

Known variants

Which variant would you like to explore?

B.1.1.529*Omicron 96.1

BA.1* 60.1%

BA.2* 35.9%

B.1.617.2*Delta 0%

B.1.1.7*Alpha 0%

B.1.351*Beta 0%

P.1*Gamma 0%

AY.43 0%

AY.122 0%

AY.46.6 0%

AY.4 0%

AY.112 0%

Sequencing intensity

Number of sequenced samples on 2022-01

550K Confirmed

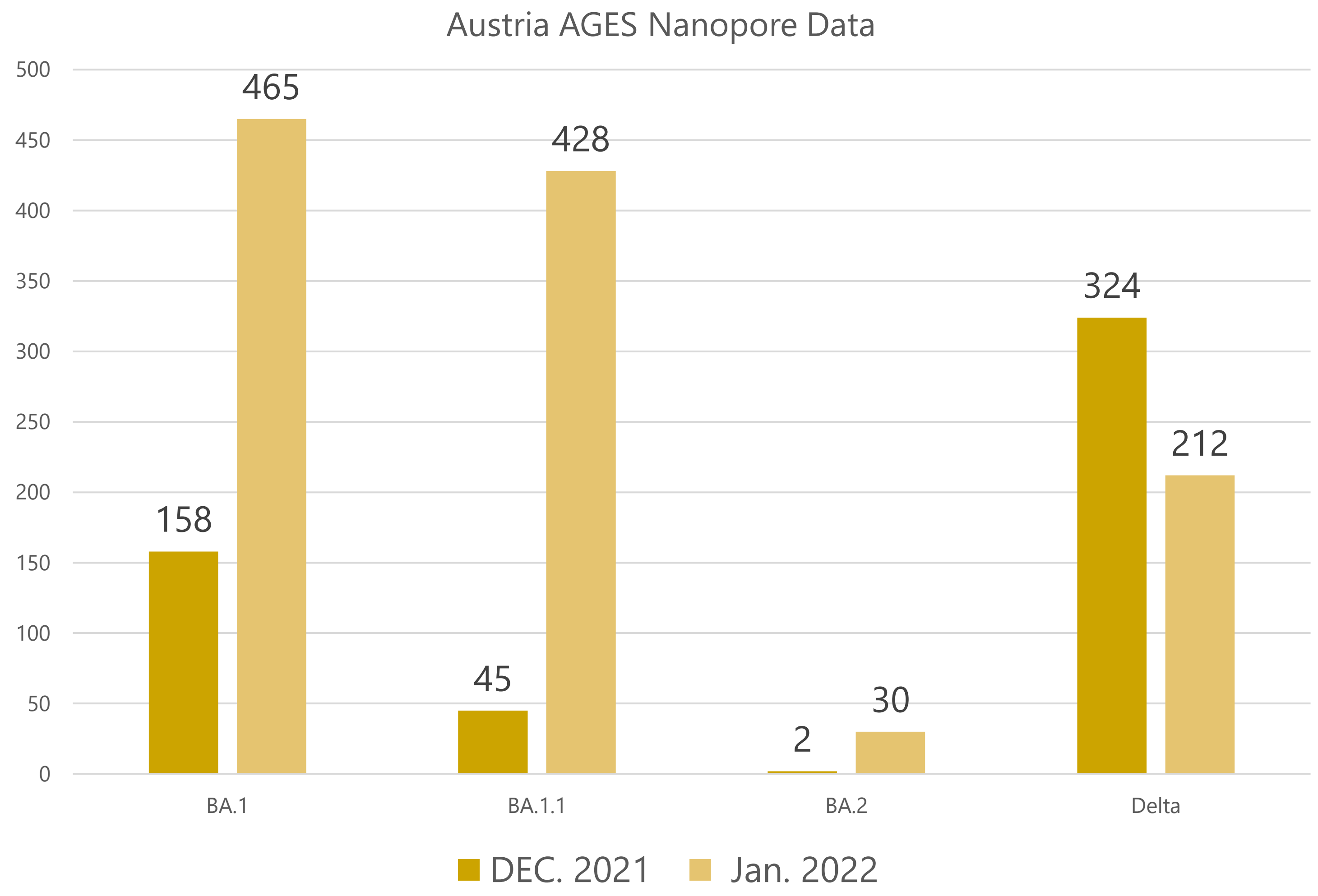
303 0% Sequenced

Metadata Availability

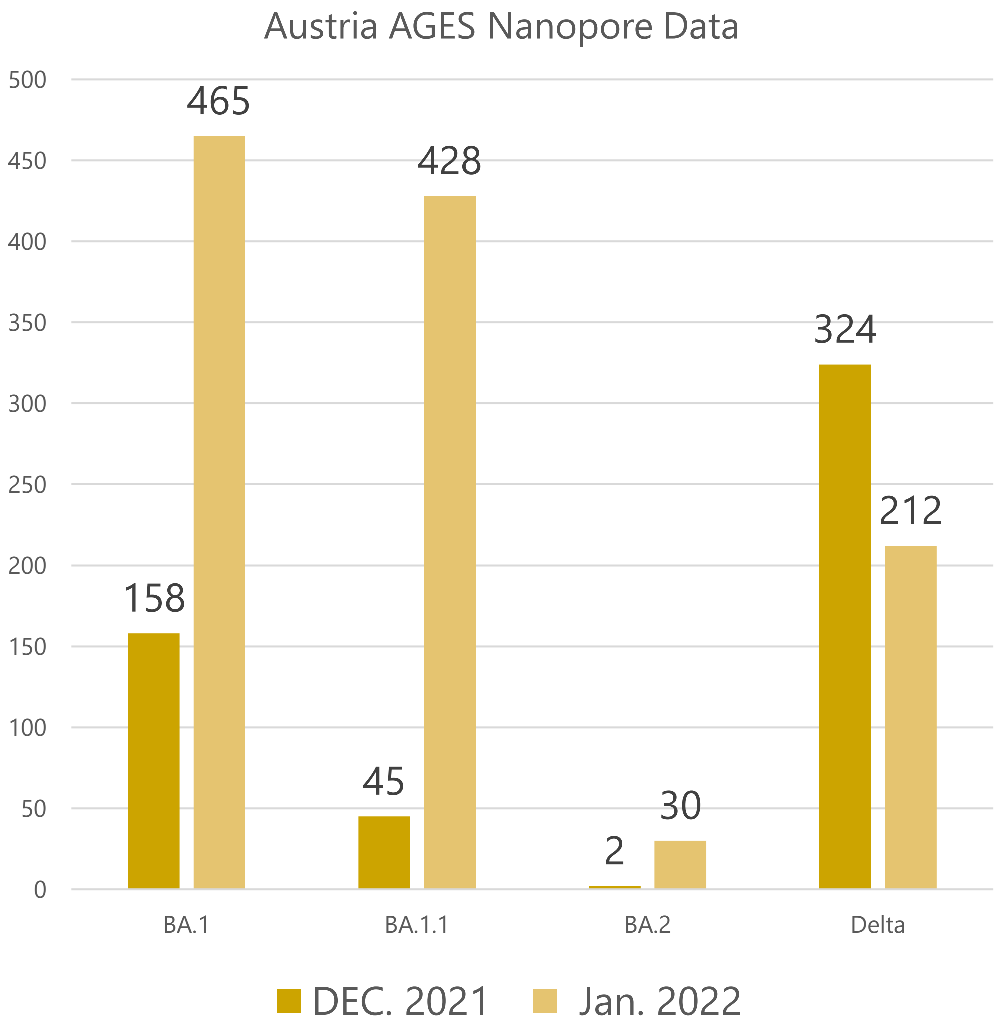
Proportion of sequences for which we have metadata information

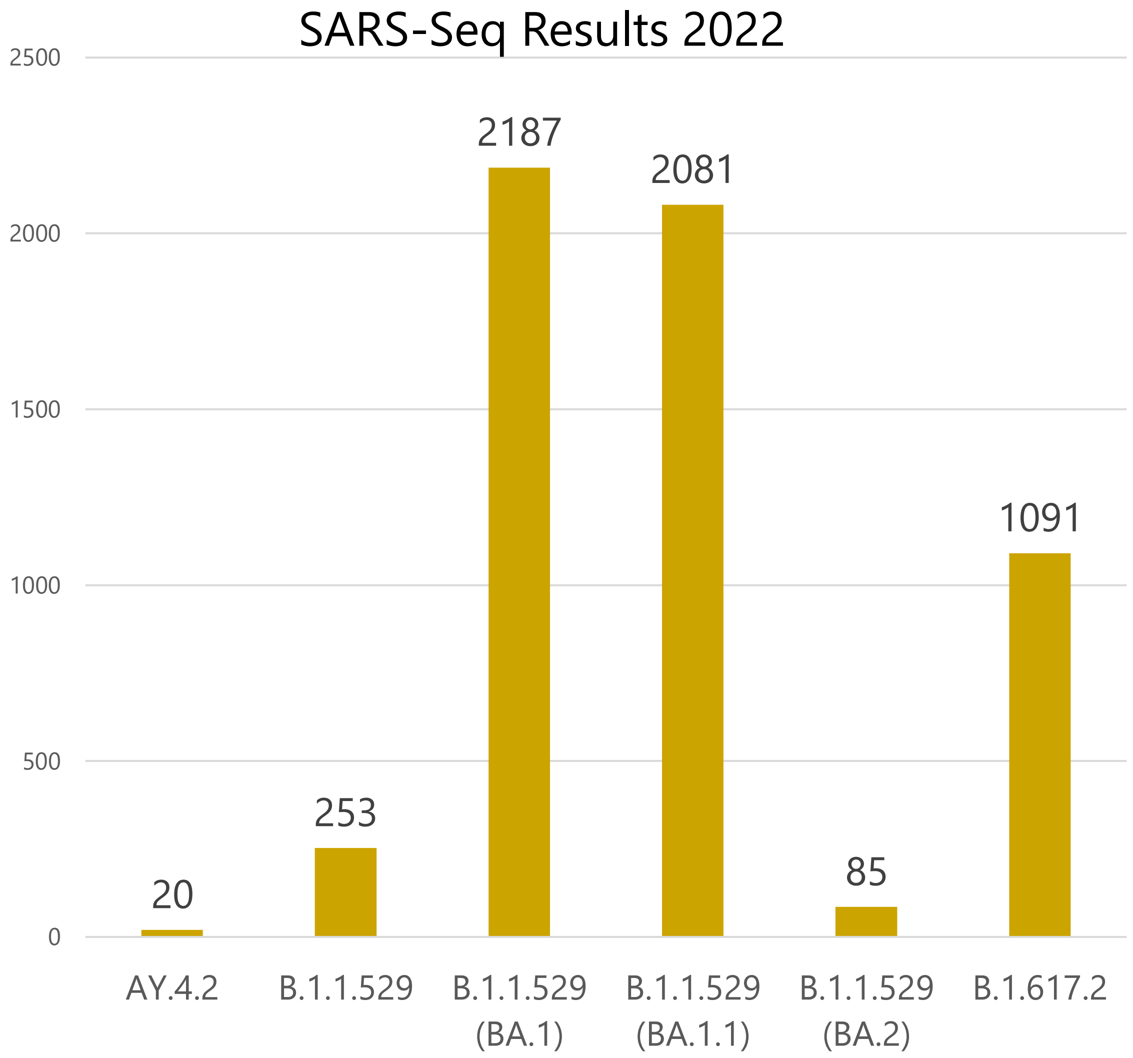
– -% Available

AGES Nanopore Data Week 50/2021 to 4/2022



AGES Nanopore Data Week 50/2021 to 4/2022





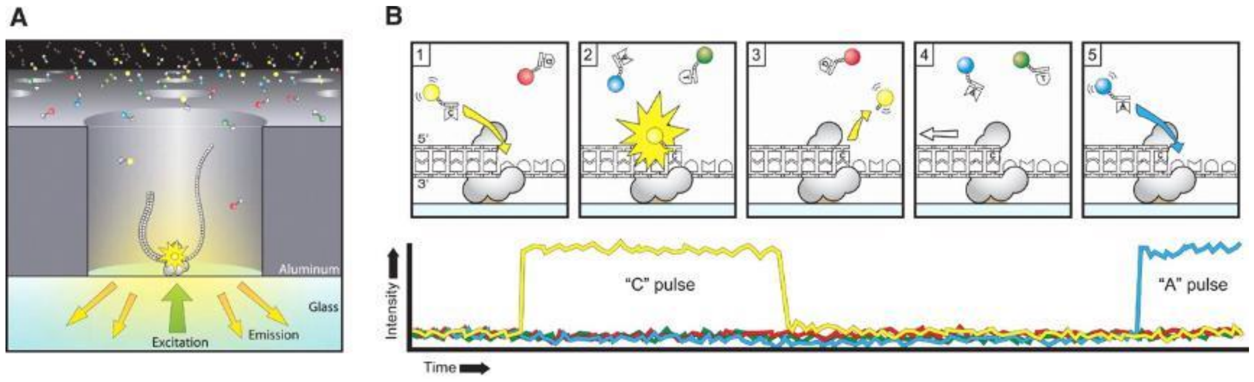
Future generation Sequencing

- **Single molecule sequencing**
- **No amplification step**
- **Shorter analysis time**
- **Cheaper reagents**
- **Less laboratory intensive**
- **Several methods out there**

Future generation Sequencing?

Single molecule real time sequencing (SMRT)

- A single DNA polymerase enzyme is affixed at the bottom of a detector well
- a single DNA molecule as a template
- The well is small enough to observe a single nucleotide being incorporated by DNA polymerase



Source: http://decodingdna.yolasite.com/single-molecule-real-time-sequencing.php

Source: www.pacb.com

Alexander Indra

AGES – Österreichische Agentur für Gesundheit und Ernährungssicherheit GmbH

Waehringersstrass 25a
A-1090 Vienna
T +43 (0) 50 555-37111 |

Alexander.indra@ages.at
www.ages.at