

The Global Pathogen Analysis System: making processing SARS-CoV-2 genetic data cheaper, simpler and more consistent

Derrick Crook

University of Oxford

Developed by:



UNIVERSITY OF
OXFORD

Powered by:

ORACLE

Disclosures

- Donation by Oracle of 10 years cloud resources to University of Oxford for pathogen genome sequence analysis
- Director of GPAS Ltd, a not-for-profit company delivering sequence data analysis and supporting genomic surveillance developed by the University of Oxford

Bottlenecks for delivering diagnostic and global surveillance services



Access to sequencing platforms across the world

- i. This is being increasingly addressed



Limited bioinformatics resources across the world

- i. A challenge even in High Income Settings
- ii. Training up many bioinformaticians isn't a quick fix
- iii. Needs a solution



Global interoperability and integration of surveillance data for genomic surveillance

- i. Work in progress with GISAID providing one successful model
- ii. G20 and WHO addressing the route to global genomic surveillance

Aims for the GPAS SARS-CoV-2 service

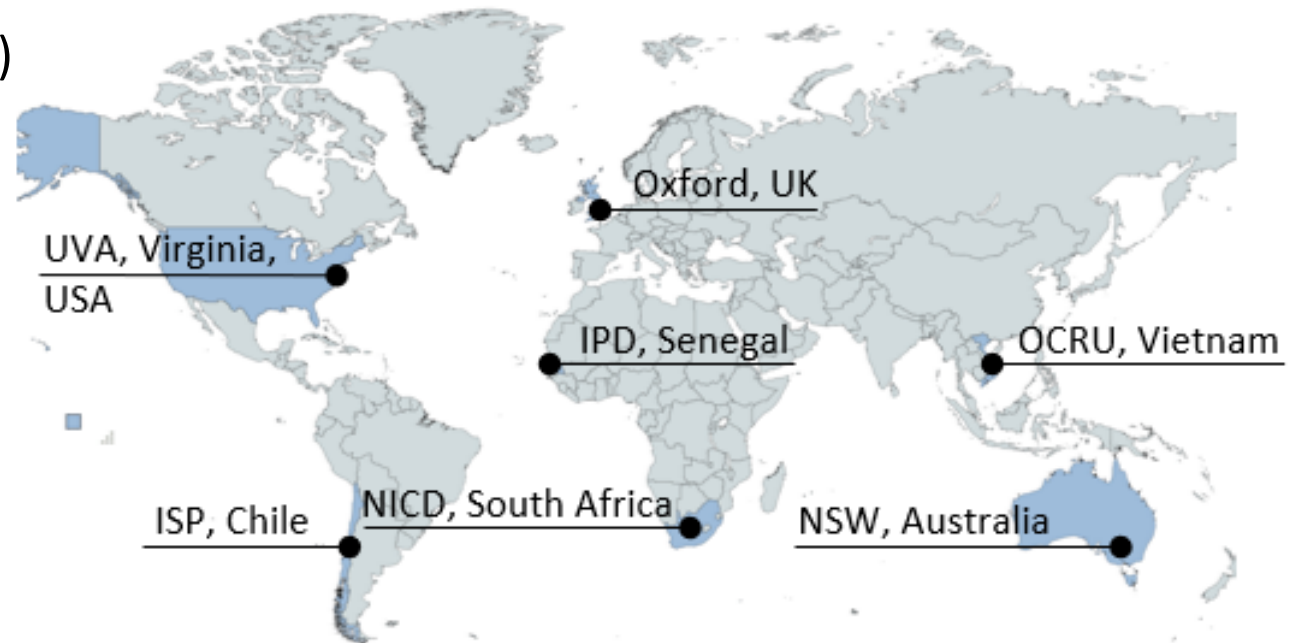
- Turnkey solution operated by routine Lab Scientists anywhere in the world
- Agnostic to sequencer and amplification primer choice
- Simple upload (free of PII), fast assembly, analysis and reporting:
 - In a GUI: lineage, completeness, missed amplicons and failed assemblies
- Simple download
 - JSON, CSV, BAM, VCF, variable and fixed length FASTA files
 - Automated submission to ENA; FASTA ready for GISAID upload
- Rigorously validated vs reference genomes and assemblers

Operational aims

- Access anywhere in the world at any time through the Oracle cloud
- Sovereign control over the data. Individual users control access:
 - Sharing is only authorised by the user through tagging
- Highly scalable:
 - Processes 250,000 ONT or 100,000 Illumina FASTQs/day
- Fast relatedness tool (for reporting per sample):
 - Identifies samples within 0, 1, 2 or 3 SNIPs

Initial deployment and evaluative testing

- Implemented in February as a test in the Oxford routine NHS clinical micro lab. Reports 100 – 700 samples a week to the UKHSA national COVID genomic surveillance programme
- Test set of 40,000 ENA FASTQs downloaded and processed in mid-July
- Seven country two week (Mid July – August) pilot evaluation



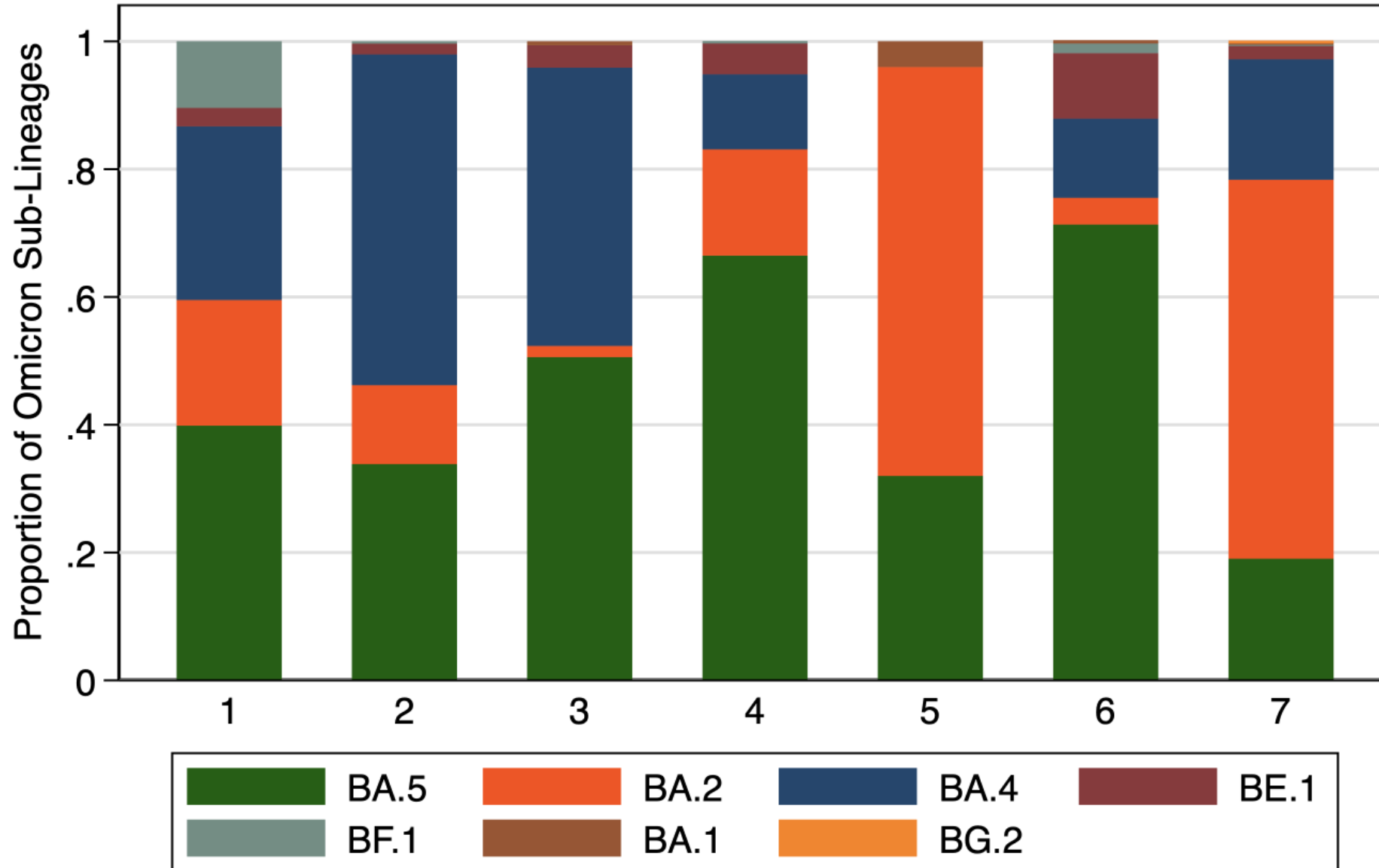
Seven country pilot

- All sites shared data using GPAS tags to grant Access
- 5,424 sequenced sampled from as early as January by 1 centre
- 4,946 with completeness of > 50% and 4,798 > 70%
- **Post June all samples were Omicron**

Full lineage data:

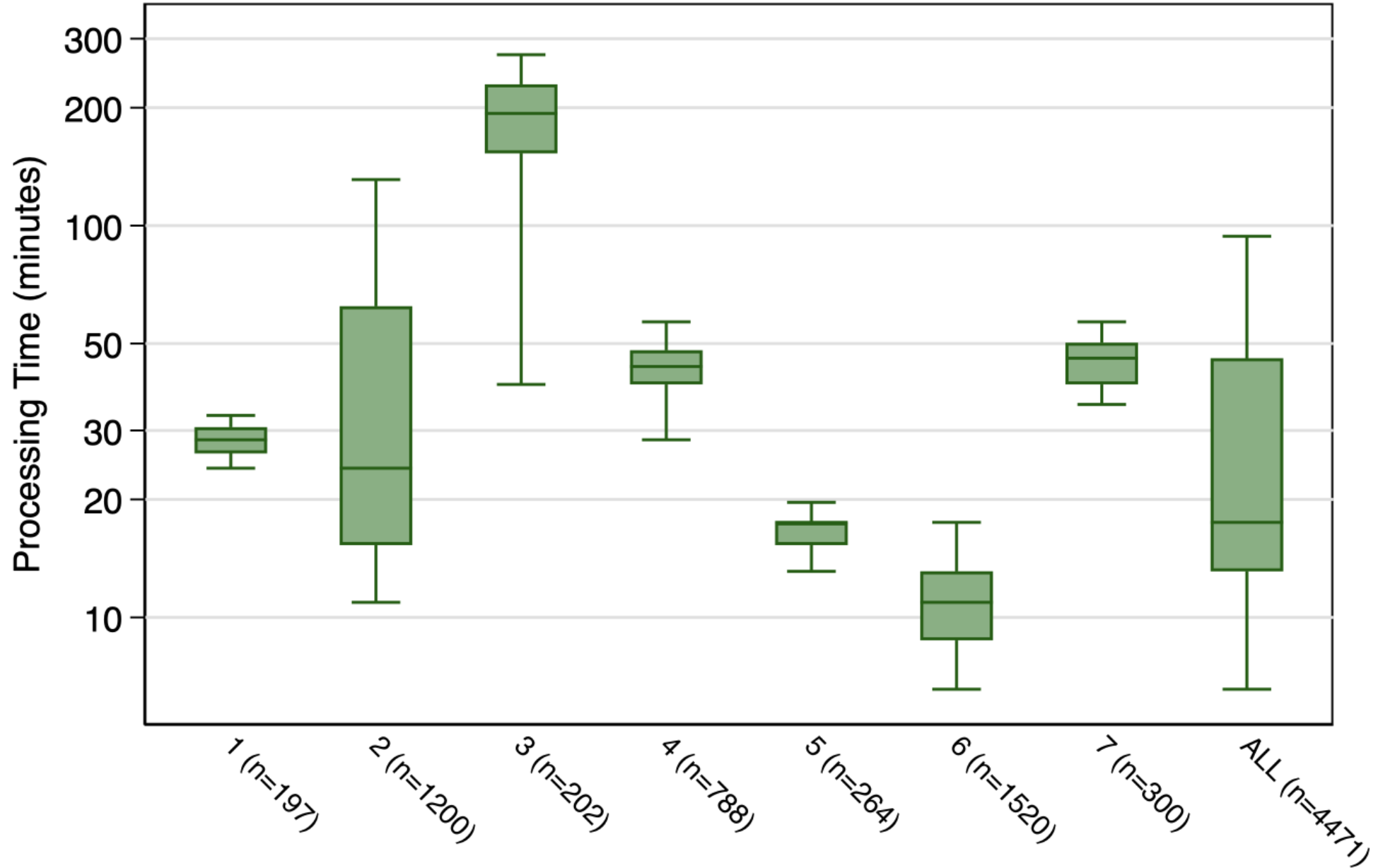
Centre	VOC	Omicron							Delta		Total
	BA.5	BA.4	BA.2	BE.1	BF.1	BA.1	BG.2	B.28	Delta		
1	69	47	34	5	18	0	0	0	0	173	
2	397	607	145	20	4	0	0	0	0	1,173	
3	89	86	3	6	0	2	0	0	3	189	
4	736	130	184	54	3	0	0	0	0	1,107	
5	8	0	256	0	0	4	0	0	22	290	
6	974	169	57	140	23	2	0	1	0	1,366	
7	95	94	296	11	1	1	1	0	1	500	
Total	2,368	1,133	975	236	49	9	1	1	26	4,798	

Omicron Sub-Lineages by Sequencing Centre

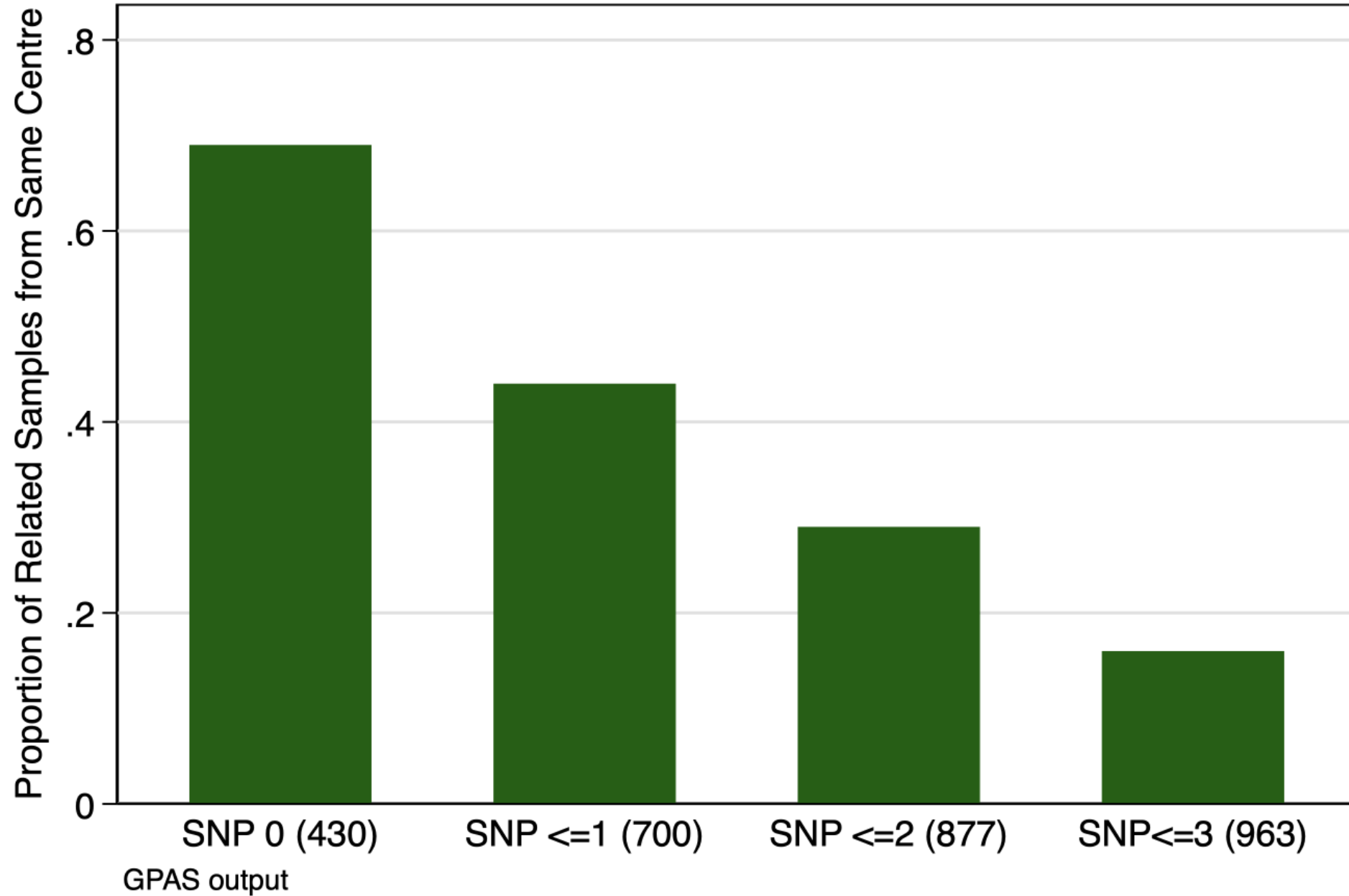


4512 Omicron isolates sequenced - June/July 2022

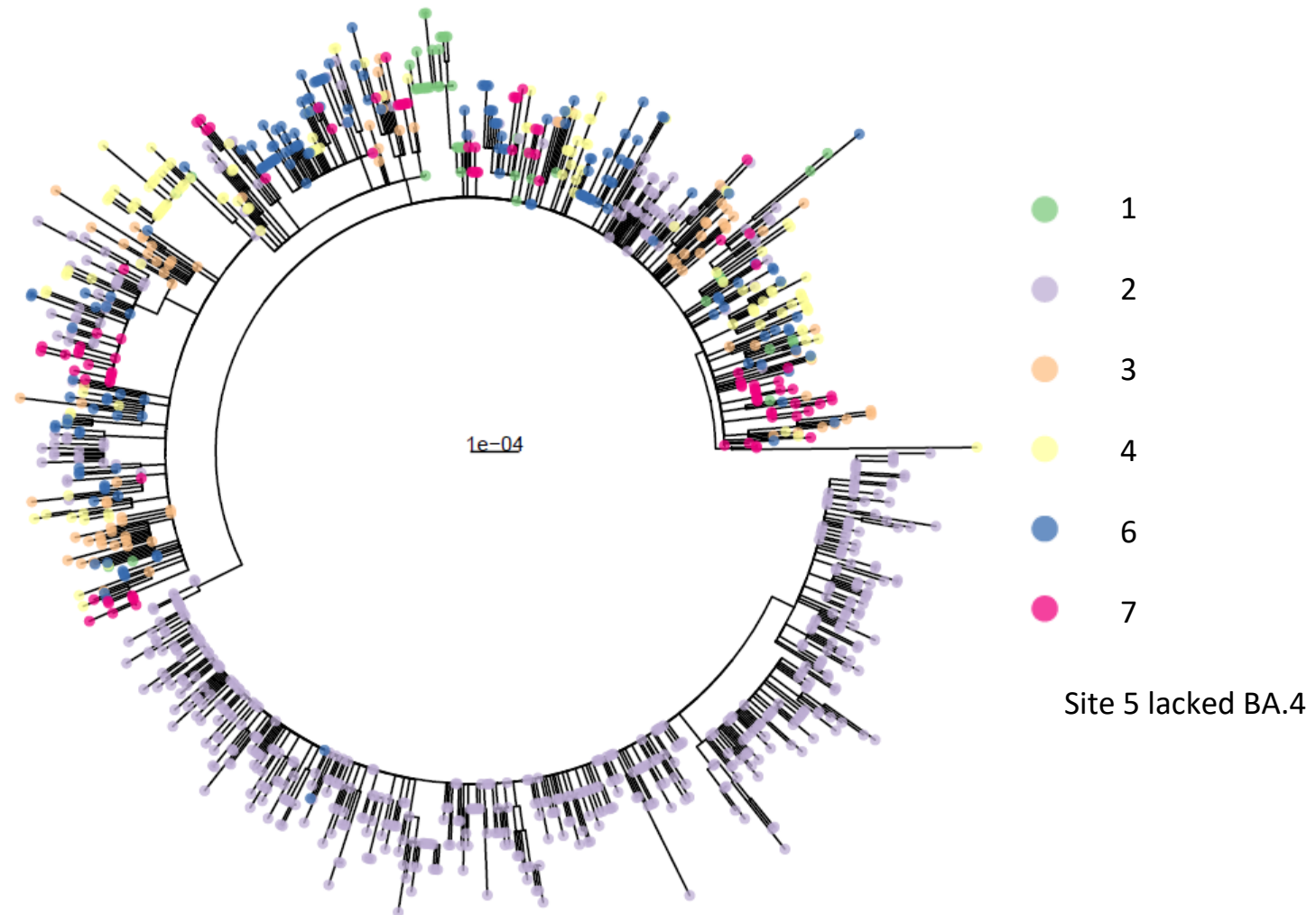
Processing Time by Sequencing Centre



Related Samples in Pangolin Lineage BA.4



BA.4 phylogenetic analysis



Conclusions

- A **simple, stable, scalable and fast** cloud hosted global services for SARS-CoV-2 analysis yielding a standardised output
- Offers validation tools for **harmonisation of assemblers**
- Pilot study revealed potential for **synchronous data processing and analysis**
- Also revealed issues with upload, which have been pinpointed and **remedied for further user-testing**
- Data download can be used for **local aggregated analysis** which conforms to current observations
- Building a service for Mycobacterium tuberculosis to be followed by others

Collaborators and Contributors

- The Institut Pasteur de Dakar (IPD)
- El Instituto de Salud Pública de Chile (ISP)
- The National Institute for Communicable Diseases (NICD) South Africa
- Microbial Genomics Reference Laboratory, New South Wales Health Pathology, Sydney, Australia
- Oxford University Clinical Research Unit OUCRU, Vietnam
- University of Virginia, Charlottesville
- Oxford University Hospitals NHS Trust
- New Variant Assessment Platform, UKHSA and NIHR HPRU Oxford
- Nuffield Department of Medicine, University of Oxford
- Global Health Security Consortium

Further info: info@gpas.global

Funders

Oracle team build simple informative dashboards



Batch View

<input type="checkbox"/>	D-8cb3e642	26-JUL-22	1	96	New (94) Control (2)	BA.5.1 (52), BA.5.2.1 (16), +10	Error (16) Unreleased (86)
<input type="checkbox"/>	D-62015c57	25-JUL-22	1	96	New (94) Control (2)	BA.5.1 (25), BA.5.2.1 (18), +12	Error (18) Unreleased (1) Released (77)

