

Public Health Genomic Epidemiology

AMR to SARS-CoV2

Finlay Maguire

finlaymaguire@gmail.com

November 27, 2020

Donald Hill Family Fellow,
Faculty of Computer Science, Dalhousie University

Table of contents

Table of contents

1. Antimicrobial Resistance

- 1.1 Incompatibilities across AMR gene detection tools
- 1.2 Creating a standardised specification
- 1.3 Using this to implement parsers and summary tools
- 1.4 Workflow proof of concept
- 1.5 Current use-cases

Table of contents

1. Antimicrobial Resistance

- 1.1 Incompatibilities across AMR gene detection tools
- 1.2 Creating a standardised specification
- 1.3 Using this to implement parsers and summary tools
- 1.4 Workflow proof of concept
- 1.5 Current use-cases

2. SARS-CoV2

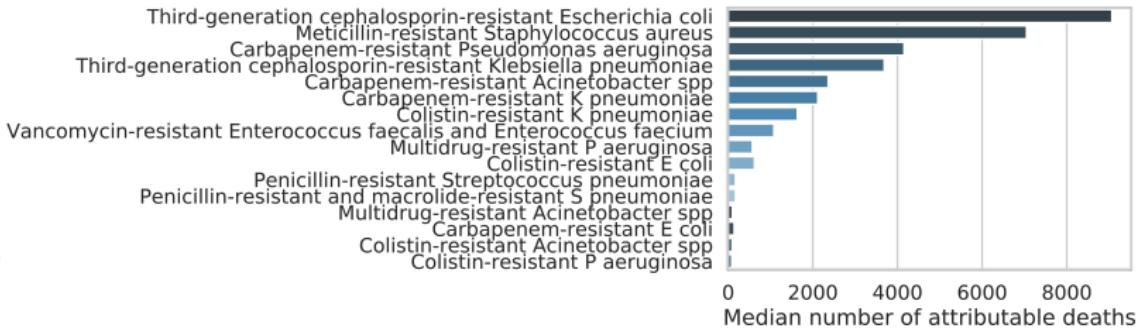
- 2.1 SARS-CoV2 genome
- 2.2 Sequencing
- 2.3 Assembling
- 2.4 Metadata
- 2.5 Phylogeography
- 2.6 On-going public health analyses

Antimicrobial Resistance

Why do we care about AMR?

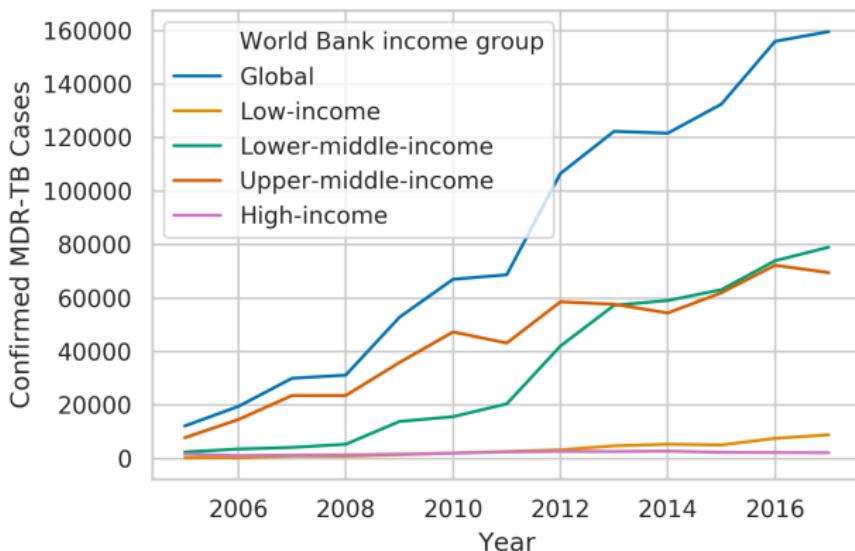
AMR is currently a problem

EU/EAA Resistant Infections



Overall: 33110 (28480-38430), Data from Cassini et al. (2019).

AMR is growing



WHO Global Health Observatory Data Repository.

What can we do about it?

Improve Surveillance

- Locally: information would help improve patient health.

Improve Surveillance

- Locally: information would help improve patient health.
- Nationally: health policies and responses to emergencies.

Improve Surveillance

- Locally: information would help improve patient health.
- Nationally: health policies and responses to emergencies.
- Globally: emerging threats and long-term trends.

Improve Surveillance

- Locally: information would help improve patient health.
- Nationally: health policies and responses to emergencies.
- Globally: emerging threats and long-term trends.
- Scientifically: better understanding of underlying biology.

**Surveillance requires AMR gene
detection**

Which tool?

- Many Stakeholders: Many Tools.

- | | |
|---------------------------|-------------------|
| 1. abricate | 8. kmerresistance |
| 2. ariba | 9. srst2 |
| 3. NCBI AMRFinderPlus | 10. staramr |
| 4. RGI (includes RGI-BWT) | 11. c-sstar |
| 5. resfinder | 12. amrplusplus |
| 6. sraX | 13. resfams |
| 7. deepARG | 14. groot |

Incompatible outputs

```
<=> results/SAMEA634591/aericate/report.csv <=
FILE   SEQUENCE   START   END   STRAND   GENE   COVERAGE   COVERAGE_MAP   GAPS   %COVERAGE   %IDENTITY   DATABASE   ACCESSION   PRODUCT   RESISTANCE
GCF_902827215_1_SBS5881_genomic.fna  NZ_LR792628.1   1330432 1333584 -   oqxB4  1-3153/3153 ===== 0/0 100.00 99.72 ncbi  NG_050453.1 multidrug efflux
subunit OqxB4 PHENICOL:QUINOLONE
GCF_902827215_1_SBS5881_genomic.fna  NZ_LR792628.1   1333608 1334783 -   oqxA  1-1176/1176 ===== 0/0 100.00 99.58 ncbi  NG_048024.1 multidrug efflux
ic adaptor subunit OqxA PHENICOL:QUINOLONE

<=> results/SAMEA634591/ampfinderplus/report.tsv <=
Protein identifier Contig id Start Stop Strand Gene symbol Sequence name Scope Element type Element subtype Class Subclass Method Target length
gth % Coverage of reference sequence % Identity to reference sequence Alignment length Accession of closest sequence Name of closest sequence HMM id
NA_NZ_LR792628.1 1330435 1333584 - oqxB multidrug efflux RND transporter permease subunit OqxB core AMR AMR PHENICOL:QUINOLONE PHENICOL:QUINOLONE BLAST
99.90 1058 WP_019794817.1 multidrug efflux RND transporter permease subunit OqxB14 NFG090937.1 multidrug efflux RND transporter permease subunit OqxB
NA_NZ_LR792628.1 1333611 1334783 - oqxA multidrug efflux RND transporter periplasmic adaptor subunit OqxA core AMR AMR PHENICOL:QUINOLONE PHENICOL:QUINOLONE
9.00 99.49 391 WP_002914189.1 multidrug efflux RND transporter periplasmic adaptor subunit OqxA NF090272.1 multidrug efflux RND transporter periplasmic adaptor

<=> results/SAMEA634591/reffinder/data.refinder.json <=
"reffinder": {"user_input": ["filename": "data/test/SAMEA634591/GCF_902827215_1_SBS5881_genomic.fna"], "method": "blast", "file_format": "fasta"}, "run_info": {"e": "17-03-09", "results": {"Aminoglycoside": {"Aminoglycoside": {"no_hit": false, "Beta-lactam": {"beta-lactam": {"no_hit": false, "Colistin": {"colistin": {"no_hit": false, "Fosfomycin": {"fosA_3_ACW001000079": {"resistance_gene": "fosA", "identity": 97.38, "HSP_length": 420, "template_length": 420, "position_in_ref": "1..420", "contig": "Klebsiella pneumoniae isolate SBS5881 chromosome SBS5881_0mosome", "positions_in_contig": "4862705..4863124", "note": "", "accession": "ACW001000079", "predicted_phenotype": "multidrug resistance"}, "coverage": 100.0, "hit_id": "NZ_LR792628.1_Klebsiella_pneumoniae_isolate_SBS5881_chromosome_SBS5881_0mosome:4862705..4863124[fosA_3_ACW001000079:97.389052]}}, "no_hit": false, "Macrolide": {"macrolide": {"no_hit": false, "Nitroimidazole": {"nitroimidazole": {"no_hit": false, "Quinoxalidinone": {"quinoxalidinone": {"no_hit": false, "Quinolone": {"quinolone": {"no_hit": true, "opxB_1_EU370913": {"resistance_gene": "opxB", "identity": 98.92, "HSP_length": 3153, "template_length": 3153, "contig_name": "NZ_LR792628_1_Klebsiella_pneumoniae_isolate_SBS5881_chromosome_SBS5881_0mosome", "positions_in_contig": "1130432..1333584", "note": "1", "accession": "Quinolone resistance", "coverage": 100.0, "hit_id": "NZ_LR792628_1_Klebsiella_pneumoniae_isolate_SBS5881_chromosome_SBS5881_0mosome:1330432..1333584[opxB_1_EU370913]", "resistance_gene": "opxA", "identity": 99.57, "HSP_length": 1176, "template_length": 1176, "position_in_ref": "1..1176", "contig_name": "NZ_LR792628_1_Klebsiella_pneumoniae_isolate_SBS5881_chromosome_SBS5881_0mosome", "positions_in_contig": "1333608..1334783", "note": "1", "accession": "EU370913", "predicted_phenotype": "Quinolone resistance"}, "hit_id": "NZ_LR792628_1_Klebsiella_pneumoniae_isolate_SBS5881_chromosome_SBS5881_0mosome:1333608..1334783[opxA_1_EU370913:99.574830]"}, "Rifampicin": {"rifampicin": {"sulphonamide": "No hit found", "Tetracycline": {"tetracycline": "No hit found"}, "Trimethoprim": {"trimethoprim": "No hit found"}, "Glycopeptide": {"glycopeptide": "No hit found"}, "Other": {"Other": "No hit found"}}, "Strict": 83.69, "Nudged": 3003579, "Model_ID": "BL_ORD_ID(1977)hsp_nu", "Note": "AM"}, "run_id": "NZ_LR792628_1_511 # 557114 # 557746 # -1 # ID=1..511;partial=99;start_type=ATG;rbx_motif=GAGG/GAGG;rbx_spacer=5..10bp;gc_cont=0..520", "Model": "hsp_nu", "Nudged": 441875, "ID": "441875 # 442859 # 1 # ID=1..400;partial=99;start_type=ATG;rbx_motif=GAGG/GAGG;rbx_spacer=5..10bp;gc_cont=0..604", "Mechanism": "protein homolog model", "Gene": "n/a", "Predicted_DNA": "n/a", "Protein_Sequence": "n/a", "Percentage": 100.0, "Length": 557114, "Resistance": "penam antibiotic efflux resistance-modulation-cell division (RND)"}
```

Why is this a problem?

Hinders public health use of AMR genomics

- Can't easily compare tools

Hinders public health use of AMR genomics

- Can't easily compare tools
- Can't easily change tools in workflows

Hinders public health use of AMR genomics

- Can't easily compare tools
- Can't easily change tools in workflows
- Can't easily generate reports

Hinders public health use of AMR genomics

- Can't easily compare tools
- Can't easily change tools in workflows
- Can't easily generate reports
- Difficult to explain variable output formats to end-users

hAMRonization specification

Interface Label	Required/Optional	Definition	Value Type	Example
Input File Name	Mandatory	<ul style="list-style-type: none">▼ Name or other identifier of an entry from a sample database.	String	ERR3581801
Input Sequence ID	Optional	<ul style="list-style-type: none">▼ The identifier of a molecular sequence being analyzed.	String	DAAGAT0100000411
Input Protein Start	Optional	<ul style="list-style-type: none">▼ The position of the first amino acid in a protein sequence being analyzed (input protein sequence). Also known as the query protein start site in a BLAST search.	Int	6
Input Protein Stop	Optional	<ul style="list-style-type: none">▼ The position of the last amino acid in a protein sequence being analyzed (input protein sequence). Also known as the query protein stop site in a BLAST search.	Int	307
Input Gene Start	Optional	<ul style="list-style-type: none">▼ The position of the first nucleotide in a gene sequence being analyzed (input gene sequence). Also known as the query gene start site in a BLAST search.	Int	18
Input Gene Stop	Optional	<ul style="list-style-type: none">▼ The position of the last nucleotide in a gene sequence being analyzed (input gene sequence). Also known as the query gene stop site in a BLAST search.	Int	921

Mandatory Terms:

- Input File Name
- Gene Symbol
- Gene Name
- Sequence Identity
- Reference Database ID

- Reference Database Version
- Reference Accession
- Reference Sequence Length
- Target Sequence Length
- Analysis Software Name
- Analysis Software Version

hAMRonization parsers

```
>hamronize -h
usage: hamronize <tool> <options>

Convert AMR gene detection tool output to hAMRonization specification format

optional arguments:
  -h, --help            show this help message and exit
  -v, --version         show program's version number and exit

Tools with hAMRonizable reports:
{abricate,amrfinderplus,ariba,rgi,resfinder,srax,deeparg,kmerresistance,srst2,staramr,csstar,
 abricate          hAMRonize abricate's output report i.e., OUTPUT.tsv
 amrfinderplus    hAMRonize amrfinderplus's output report i.e., OUTPUT.tsv
 ariba            hAMRonize ariba's output report i.e., OUTDIR/OUTPUT.tsv
 rgi              hAMRonize rgi's output report i.e., OUTPUT.txt or OUTPUT_bwtoutput.gene
 resfinder        hAMRonize resfinder's output report i.e., data_resfinder.json
 srax             hAMRonize srax's output report i.e., srax_detected_ARGs.tsv
 deeparg          hAMRonize deeparg's output report i.e., OUTDIR/OUTPUT.mapping.ARG
 kmerresistance   hAMRonize kmerresistance's output report i.e., OUTPUT.KmerRes
 srst2            hAMRonize srst2's output report i.e., OUTPUT_srst2_report.tsv
 staramr          hAMRonize staramr's output report i.e., resfinder.tsv
 csstar           hAMRonize csstar's output report i.e., OUTPUT.tsv
 amrplusplus      hAMRonize amrplusplus's output report i.e., gene.tsv
 resfams          hAMRonize resfams's output report i.e., resfamstblout
 groot            hAMRonize groot's output report i.e., OUTPUT.tsv (from `groot report`)
```

github.com/fmaguire/hamronization

hAMRonization parsers

```
>hamronize abricate -h
usage: hamronize abricate <options>

Applies hAMRonization specification to output from abricate (OUTPUT.tsv)

positional arguments:
  report           Path to tool report

optional arguments:
  -h, --help        show this help message and exit
  --format FORMAT   Output format (tsv or json)
  --output OUTPUT    Output location
  --analysis_software_version ANALYSIS_SOFTWARE_VERSION
                     Input string containing the analysis_software_version for abricate
  --reference_database_version REFERENCE_DATABASE_VERSION
                     Input string containing the reference_database_version for abricate
```

github.com/fmaguire/hamronization

hAMRonization reports

```
> hamronize summarize -h
usage: hamronize summarize <options> <list of reports>

Concatenate and summarize AMR detection reports

positional arguments:
    hamronized_reports      list of hAMRonized reports

optional arguments:
    -h, --help            show this help message and exit
    -t {tsv,json,interactive}, --summary_type {tsv,json,interactive}
                          Which summary report format to generate
    -o OUTPUT, --output OUTPUT
                          Output file path for summary
```

Front-end Wizardry: Alex Manuele

hAMRonization reports

The screenshot shows a web-based reporting tool for hAMRonization. At the top left is the logo of the Public Health Alliance for Genomic Epidemiology. To its right is a search bar with a placeholder 'Search' and a button labeled 'Show Only Genomes With Hits'. Below the search bar is a table of search results.

Search Results

Genome ID	abricate: config 0	amrfinderplus: config 0	csstar: config 0	resfinder.py: config 0	staramr: config 0
ERR873305	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873306	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873307	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873308	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873309	16 hits	14 hits	14 hits	6 hits	11 hits
ERR873310	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873311	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873312	11 hits	10 hits	10 hits	6 hits	9 hits

Total hits: 111 (with a link icon)

Genomes with hits: 8 (with a link icon)

Tools with hits: 5 (with a link icon)

Differential results: 0 (with a link icon)

Selected (highlighted in blue)

Compare (highlighted in blue) | **Clear**

Front-end Wizardry: Alex Manuele

hAMRonization reports

Public Health Alliance for Genomic Epidemiology

Show Only Genomes With Hits

	abricate: config 0	amerfinderplus: config 0	csstar: config 0	resfinder.py: config 0	staramr: config 0
ERR873305	11 hits	10 hits	10 hits	6 hits	9 hits
	aac(6')-29a	aac(6')-29a	aac(6')	aac(6')-29a	aac(6')-29a
	aph(3')-Ib	aph(3')-Ib	aph(3')	blaVIM-2	aph(3')-Ib
	blaOXA-395	blaOXA-395	bcr1	catB7	blaOXA-50
	blaPDC-158	blaPDC-3	blaOXA	crrP	blaPAO
	blaPDC-55	blaVIM-2	blaPAO	fosA	blaVIM-2
	blaVIM-2	catB7	blaVIM	sul1	catB7
	catB7	crrP	catB7	crrP	crrP
	crrB1	fosA	crrP	fosA	fosA
	crrP	qacDelta1	fosATR	sul1	sul1
	fosA-354827590	sul1	fosATR		
	sul1		sul1		

Search Results

Total hits: 346

Genomes with hits: 45

Tools with hits: 4

Differential results: 62

Selected

fosA-354827590 from genome 0, tool 0

fosA from genome 0, tool 1

fosATR from genome 0, tool 2

Compare Clear

ERR873306

Front-end Wizardry: Alex Manuele

hAMRonization reports



Public Health Alliance for
Genomic Epidemiology

fosfomycin Show Only Genomes With Hits

	fosA-354827590	fosA	fosATR
	ERR873305	ERR873305	ERR873305
	abricate: config 0	amrfinderplus: config 0	csstar: config 0
input_file_name	ERR873305	ERR873305	ERR873305
gene_symbol	fosA-354827590	fosA	fosATR
gene_name	FosA family fosfomycin resistance glutathione transferase	FosA family fosfomycin resistance glutathione transferase	fosATR
reference_database_id	ncbi	NCBI Reference Gene Database	ResGANNOT
reference_database_version	2020-Apr-19	2020-03-20.1	2020-Nov-05
reference_accession	NG_047883.1	WP_003082280.1	fosATR
analysis_software_name	abricate	amrfinderplus	csstar
analysis_software_version	1.0.1	3.6.10	2.0.0
sequence_identity	98.53	98.52	98.529
config_id	gnl BUGS ERR873305_36	gnl BUGS ERR873305_36	gnl BUGS ERR873305_36
query_start_aa			
query_stop_aa			
query_start_nt	8371	8374	

Front-end Wizardry: Alex Manuele

hAMRonization reports

Public Health Alliance for Genomic Epidemiology

aac(6')-lb

Show Only Genomes With Hits

	abricate: config 0	amrfinderplus: config 0	csstar: config 0	resfinder.py: config 0	starmr: config 0
ERR873305	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873306	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873307	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873308	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873309	16 hits aac(5')-lb'	14 hits aac(6')-lb'	14 hits CARB	6 hits aac(6')-lb-cr	11 hits aac(6')-lb
	aadA2	aadA2	aac(3')	aac(6')-lb3	aac(6')-lb-cr
	aph(3')-lb	aph(3')-lb	aph(3')-lc	blaCARB-2	aadA2
	aph(3')-lc	aph(3')-lc	aadA12	catB7	aph(3')-lb
	aph(6')-Smalt	aph(6)	aadA2	fosA	blaCARB-2
	blaCNDP-2	blaCNDP-2	blaOXA-51	blaOXA-50	blaOXA-50

Search Results

Total hits: 83

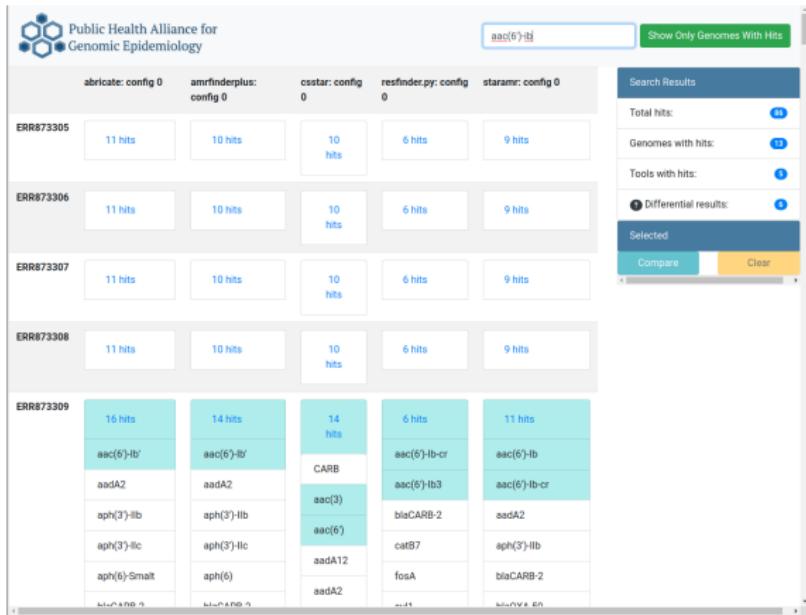
Genomes with hits: 13

Tools with hits: 5

Differential results: 4

Selected

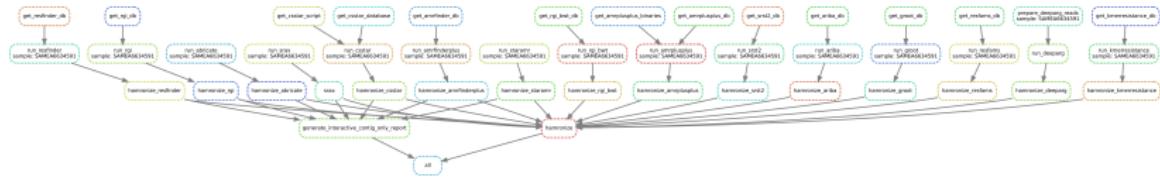
Compare Clear



Front-end Wizardry: Alex Manuele

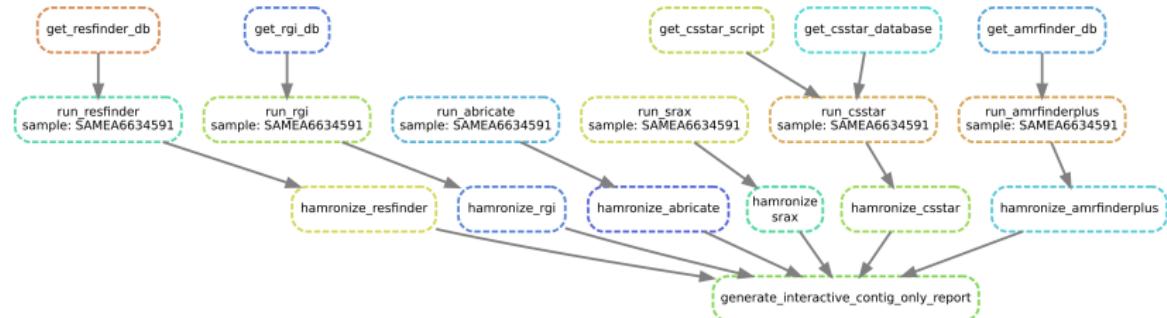
How do you use these?

hAMRonization workflow



Predict AMR genes for 14 tools for an arbitrary number of sequences and hAMRonize into a single report. (Ines Mendes, Simon Tausch, Adam Witney).

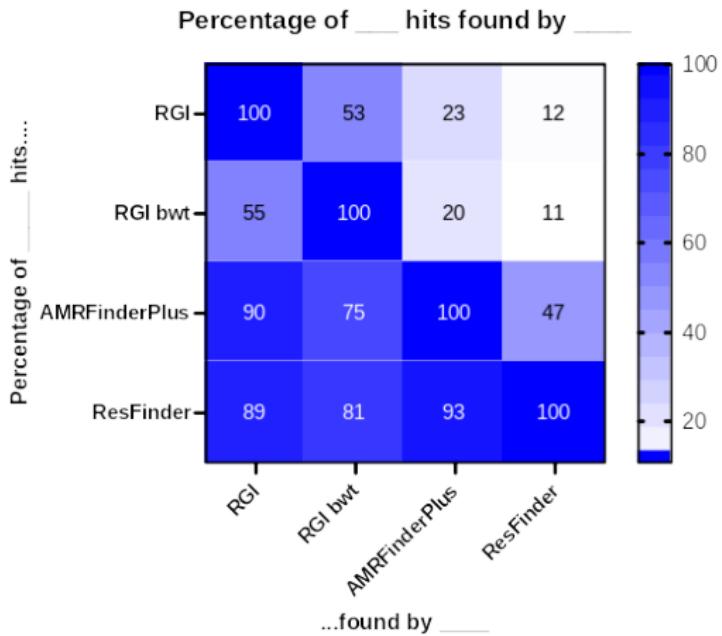
hAMRonization workflow



Predict AMR genes for 14 tools for an arbitrary number of sequences and hAMRonize into a single report. (Ines Mendes, Simon Tausch, Adam Witney).

Using hAMRonization

89 Klebsiella Genomes



RGI: 140, RGI-bwt: 223, AFP: 70, ResFinder: 46
(Beth Culp, Amos Raphenya)

Current beta-deployment



PAHO AMR Reference Labs: Chile, Colombia, Argentina, Mexico. Canada,
Africa CDC: Nigeria, Portuguese Public Health. (Led by Josefina Campos)

Summary

- hAMRonization solves the tool proliferation issues in AMR gene detection

Summary

- hAMRonization solves the tool proliferation issues in AMR gene detection
- Public Health labs have been very excited about it

Summary

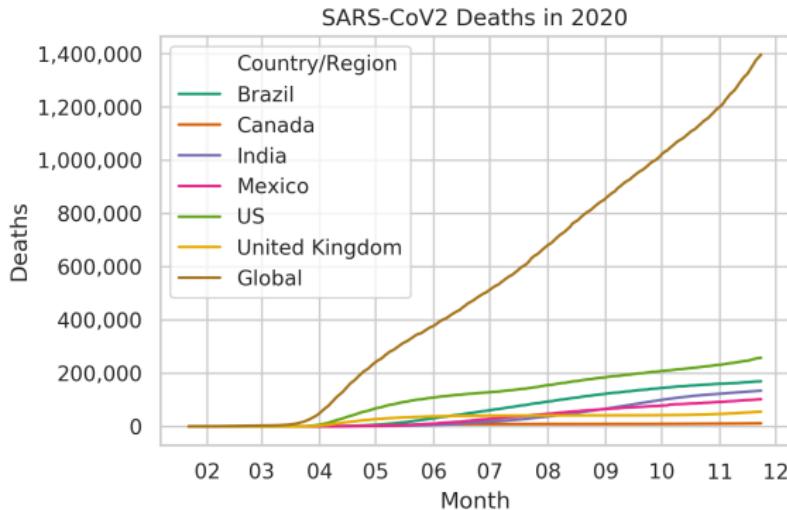
- hAMRonization solves the tool proliferation issues in AMR gene detection
- Public Health labs have been very excited about it
- Currently doesn't support variant AMR: currently being implemented

Summary

- hAMRonization solves the tool proliferation issues in AMR gene detection
- Public Health labs have been very excited about it
- Currently doesn't support variant AMR: currently being implemented
- Doesn't solve cross-AMR database issues: future work

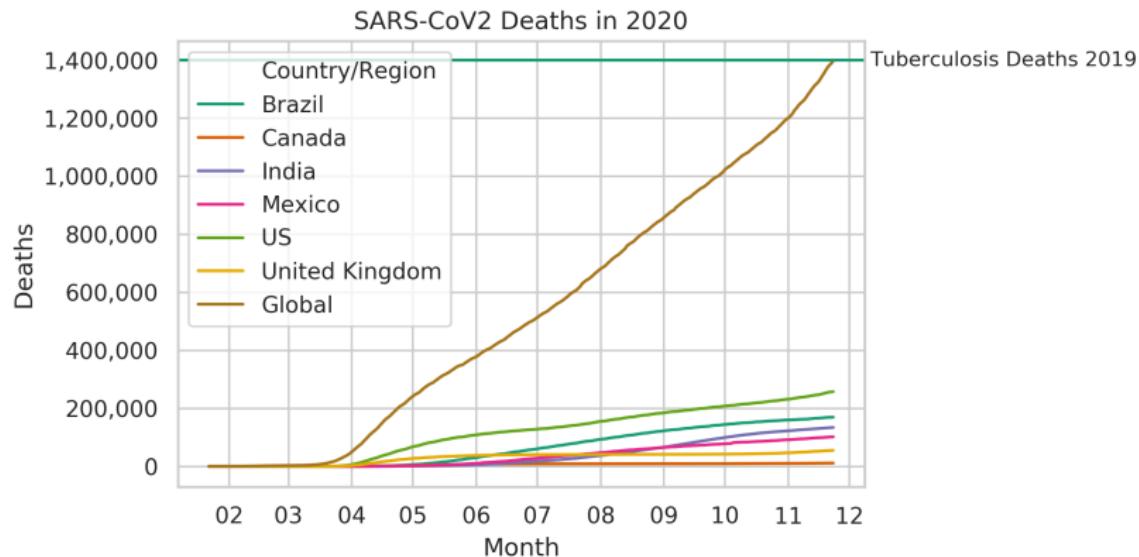
Genomic Epidemiology of SARS-CoV2

Global Crisis



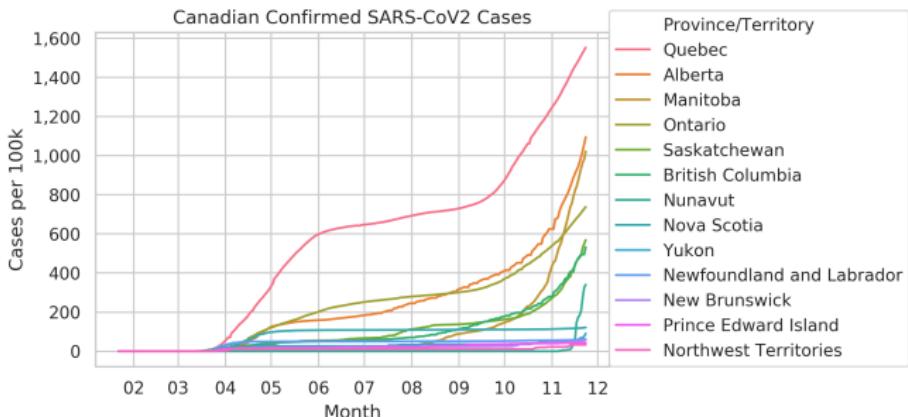
Data from github.com/CSSEGISandData/COVID-19

Global Crisis



Data from github.com/CSSEGISandData/COVID-19

National Crisis



Data from github.com/CSSEGISandData/COVID-19 and StatCan

Why genomic epidemiology?

- Real-time genomic surveillance key to managing Ebola, Lassa virus, and Yellow fever outbreaks

Why genomic epidemiology?

- Real-time genomic surveillance key to managing Ebola, Lassa virus, and Yellow fever outbreaks
- Insights into circulation lineages, evolution, and outbreak dynamics

Why genomic epidemiology?

- Real-time genomic surveillance key to managing Ebola, Lassa virus, and Yellow fever outbreaks
- Insights into circulation lineages, evolution, and outbreak dynamics
- Fix failures of contact tracing, diagnostics

Why genomic epidemiology?

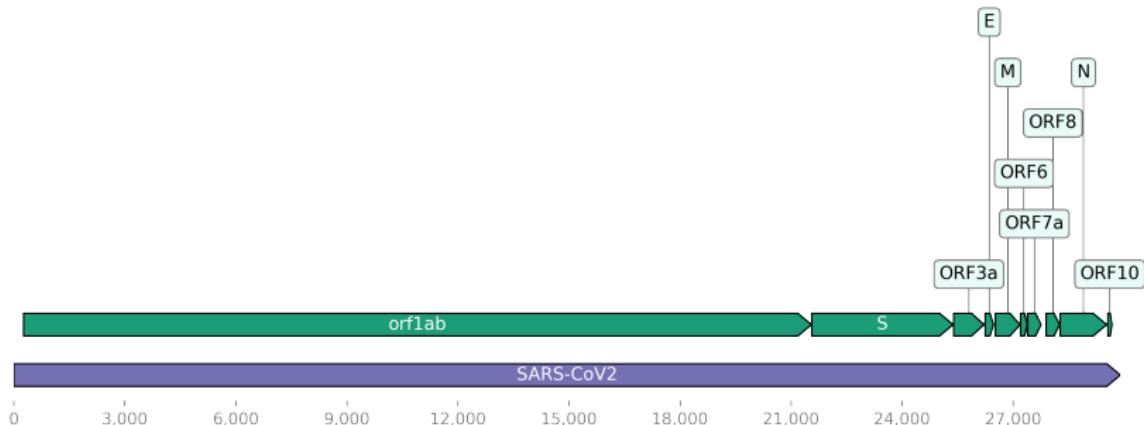
- Real-time genomic surveillance key to managing Ebola, Lassa virus, and Yellow fever outbreaks
- Insights into circulation lineages, evolution, and outbreak dynamics
- Fix failures of contact tracing, diagnostics
- Optimise policy, treatment, and vaccine development

SARS-CoV2 Genome



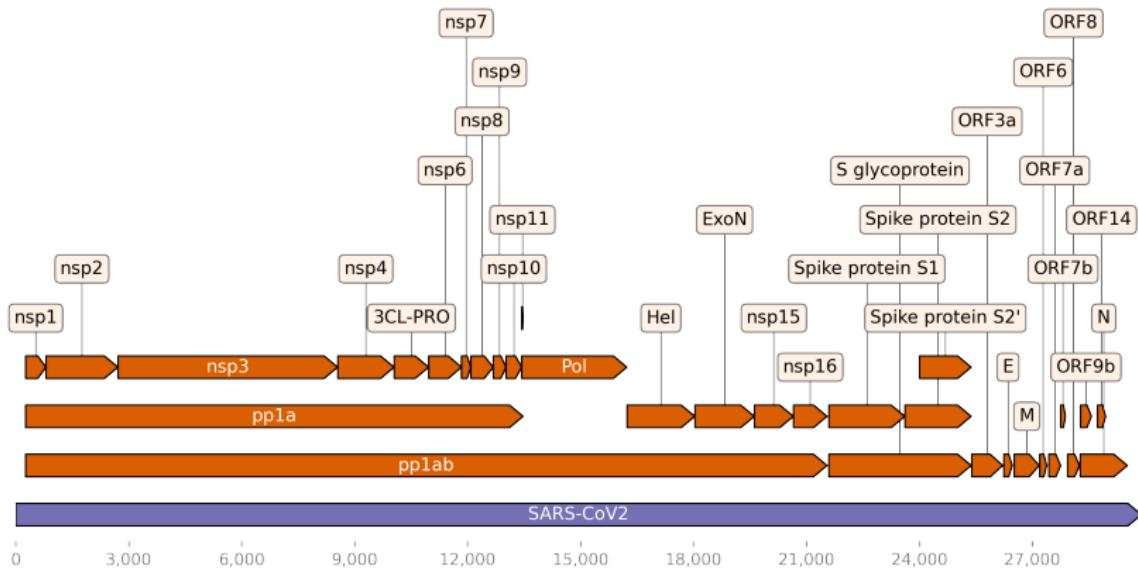
Replotted from UCSC Genome Browser

SARS-CoV2 Genome



Replotted from UCSC Genome Browser

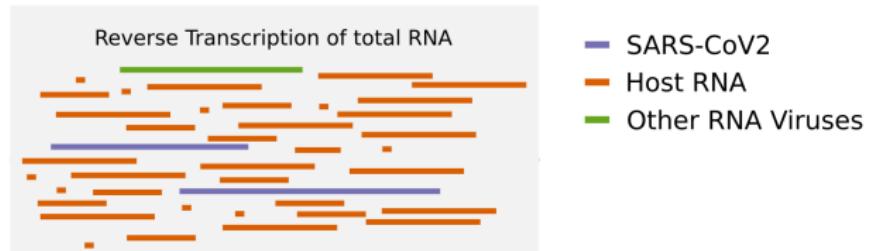
SARS-CoV2 Genome



Replotted from UCSC Genome Browser

Sequencing SARS-CoV2

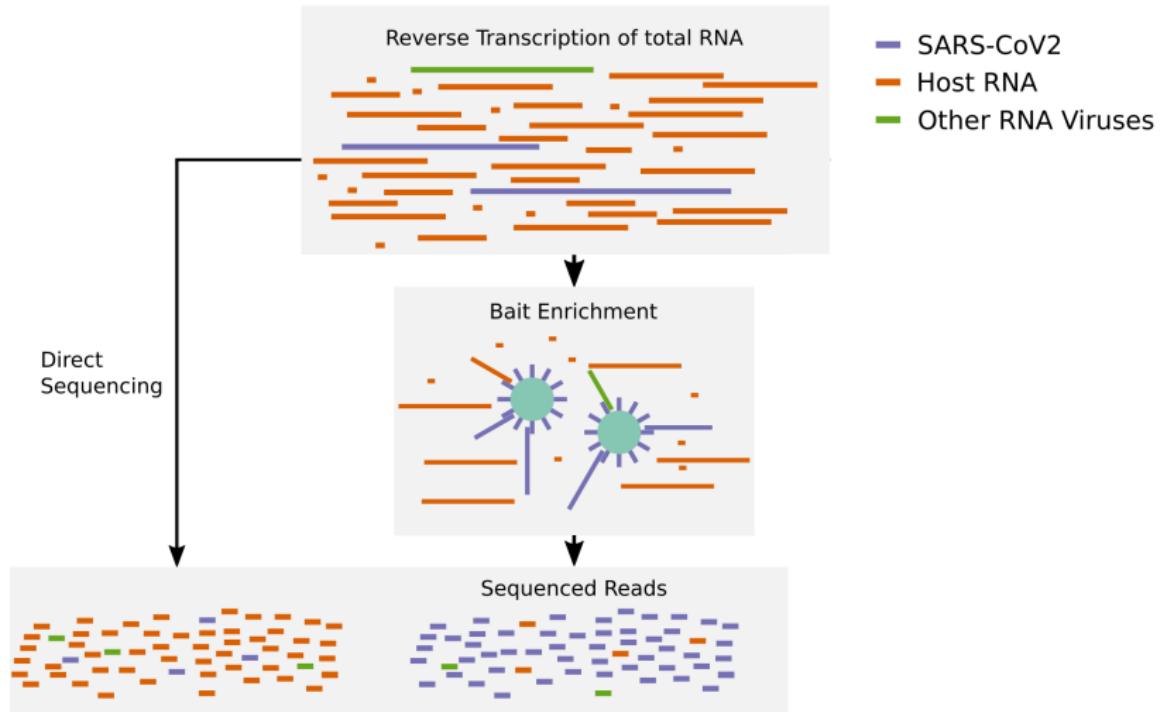
SARS-CoV2 Sequencing Methods



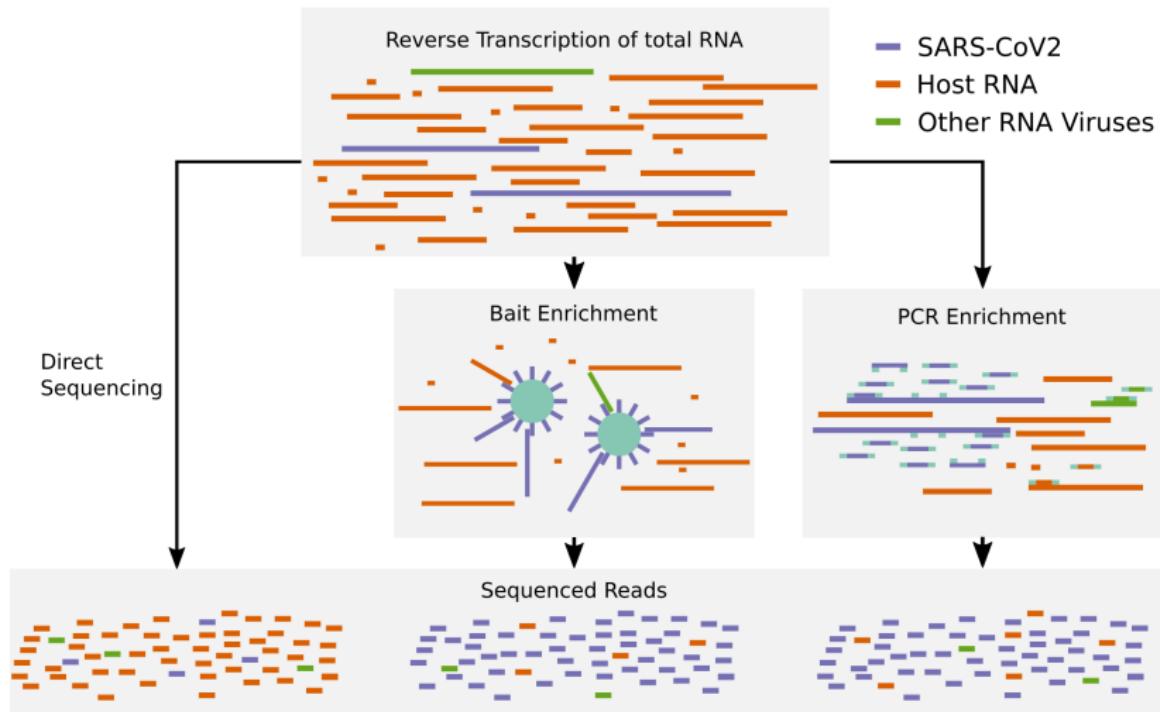
SARS-CoV2 Sequencing Methods



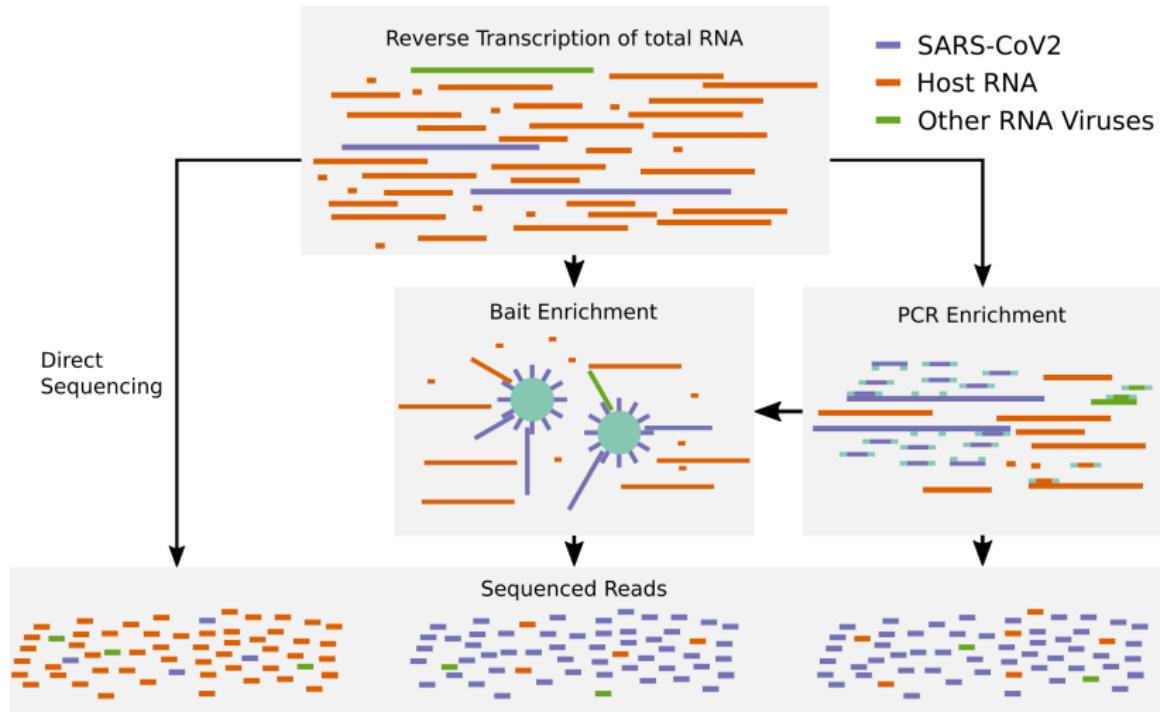
SARS-CoV2 Sequencing Methods



SARS-CoV2 Sequencing Methods



SARS-CoV2 Sequencing Methods

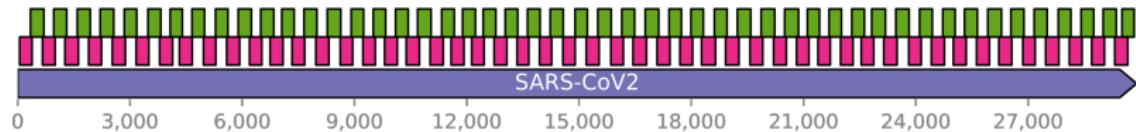


PCR Amplicon Schemes



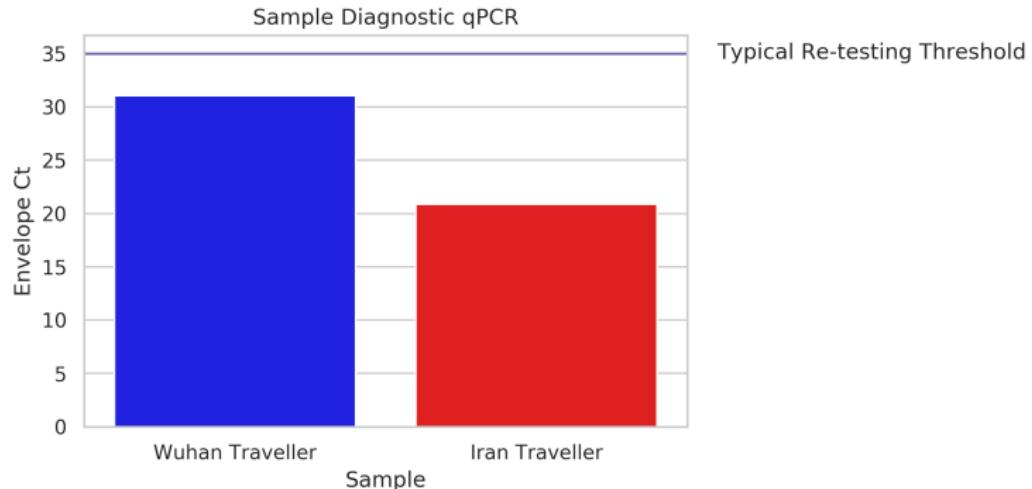
ARTICv3: 96 x 400bp amplicons

PCR Amplicon Schemes



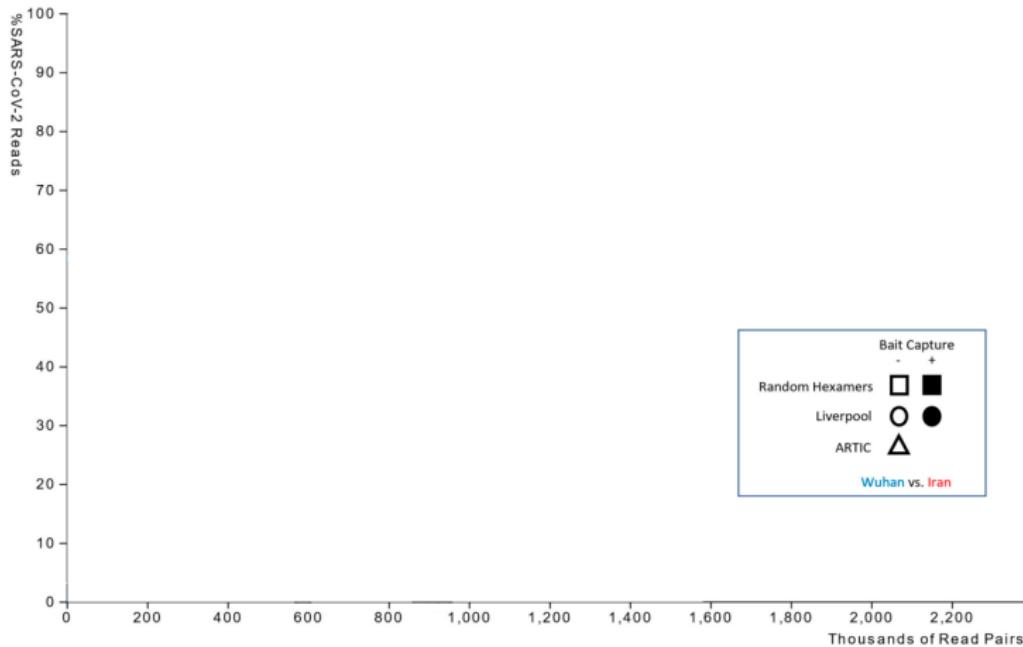
ARTICv3: 96 x 400bp amplicons

Which Approach?



- Direct cDNA sequencing
- Bait Enrichment
- ARTICv3 (96 amplicons)
- Liverpool (30 amplicons)
- Liverpool (30 amplicons) + Bait Enrichment

Which Approach?

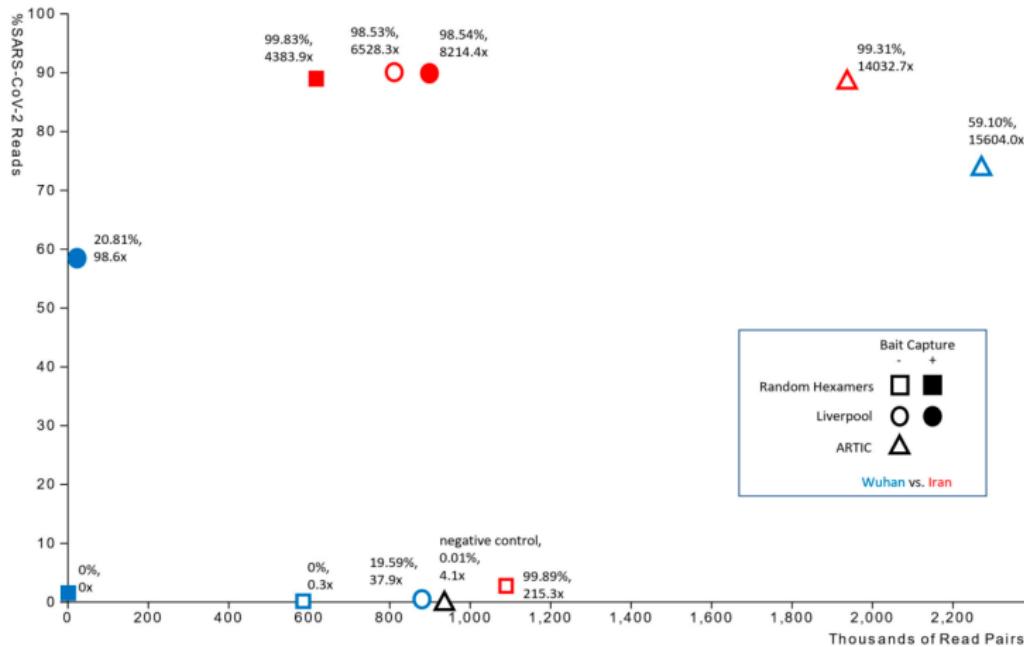


Wuhan Traveller: Low concentration sample

Iran Traveller: High concentration sample

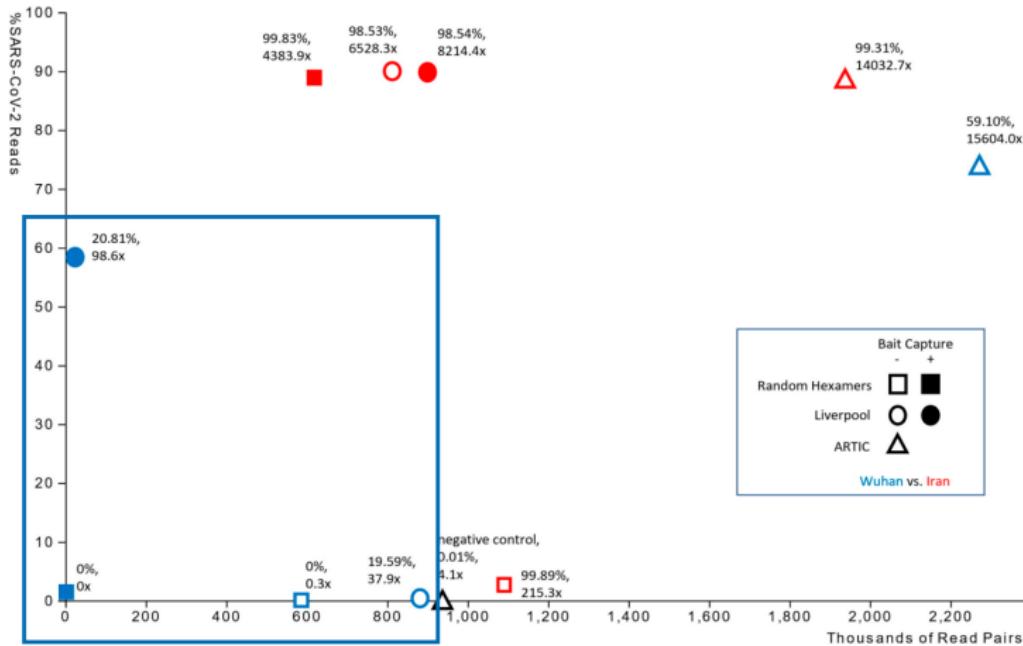
(Nasir et al., 2020)

Which Approach?



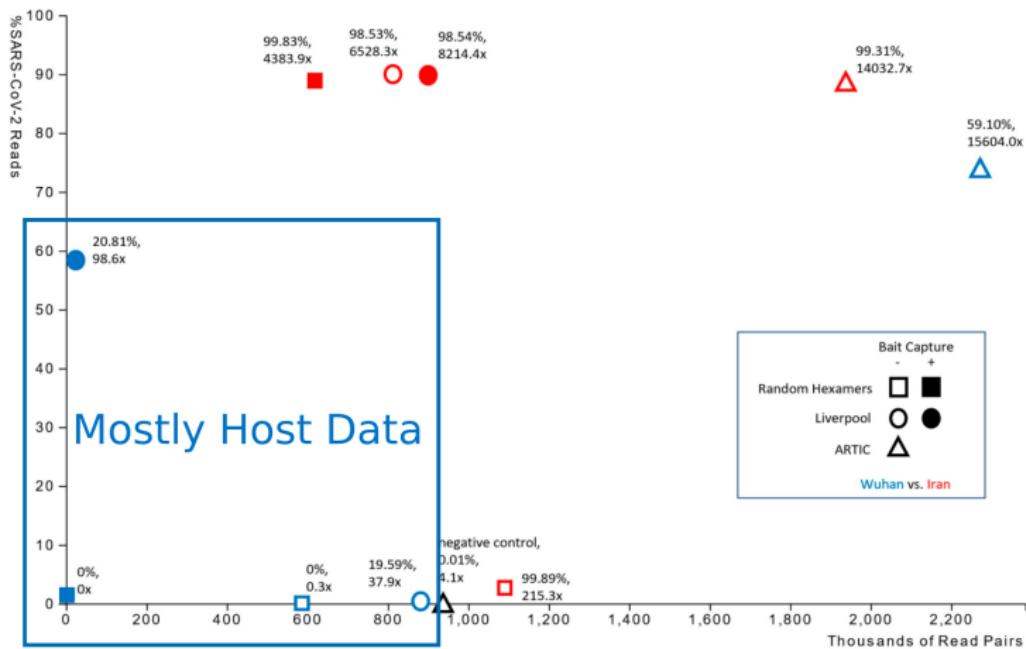
Wuhan Traveller: Low concentration sample
Iran Traveller: High concentration sample
(Nasir et al., 2020)

Which Approach?



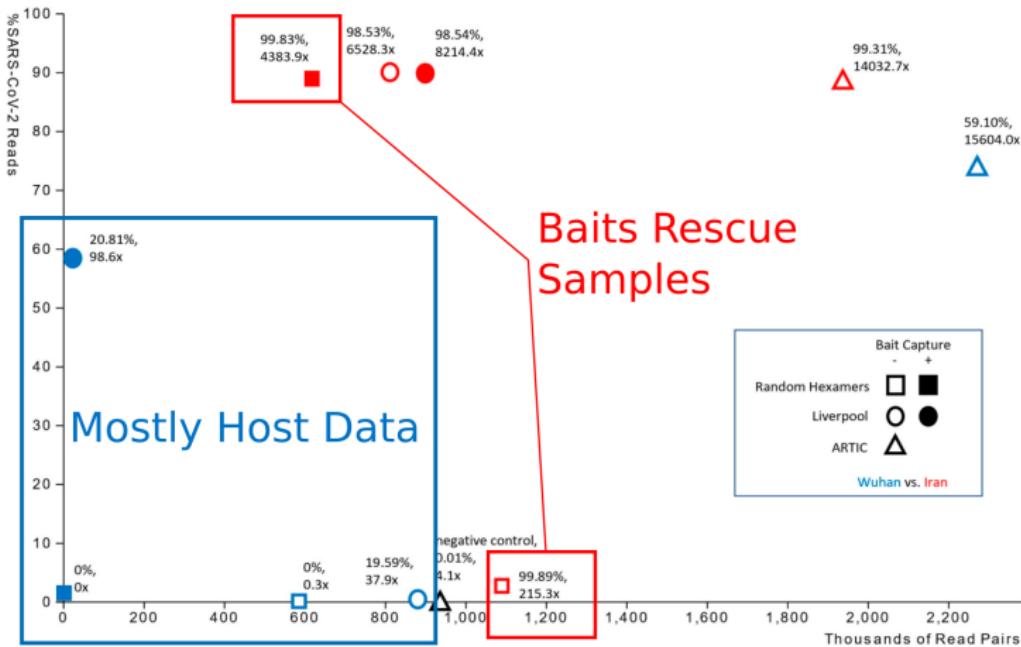
Wuhan Traveller: Low concentration sample
Iran Traveller: High concentration sample
(Nasir et al., 2020)

Which Approach?



Wuhan Traveller: Low concentration sample
Iran Traveller: High concentration sample
(Nasir et al., 2020)

Which Approach?

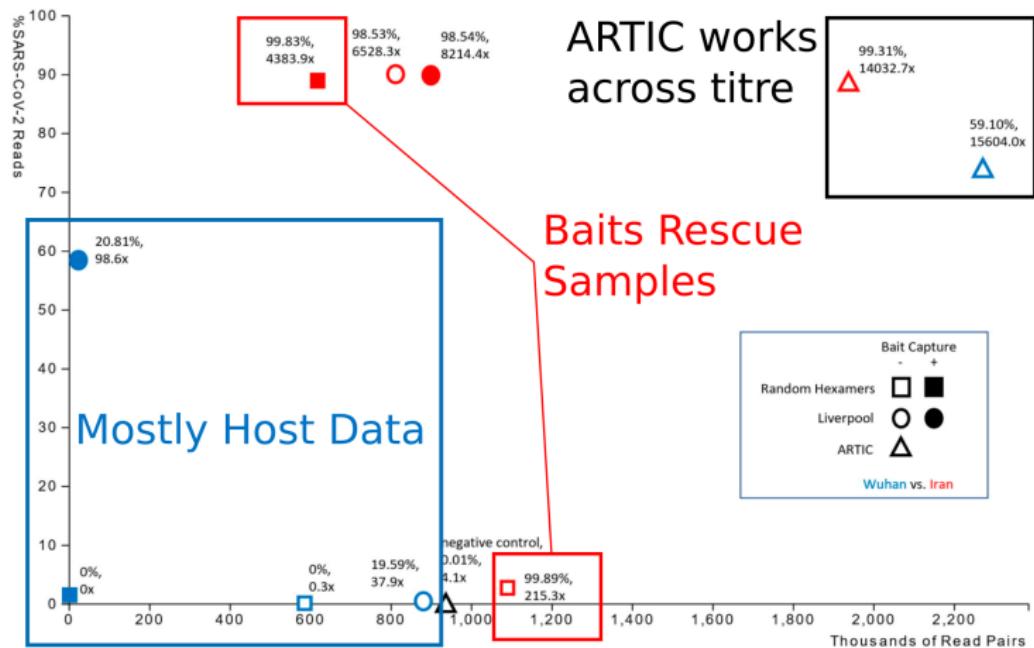


Wuhan Traveller: Low concentration sample

Iran Traveller: High concentration sample

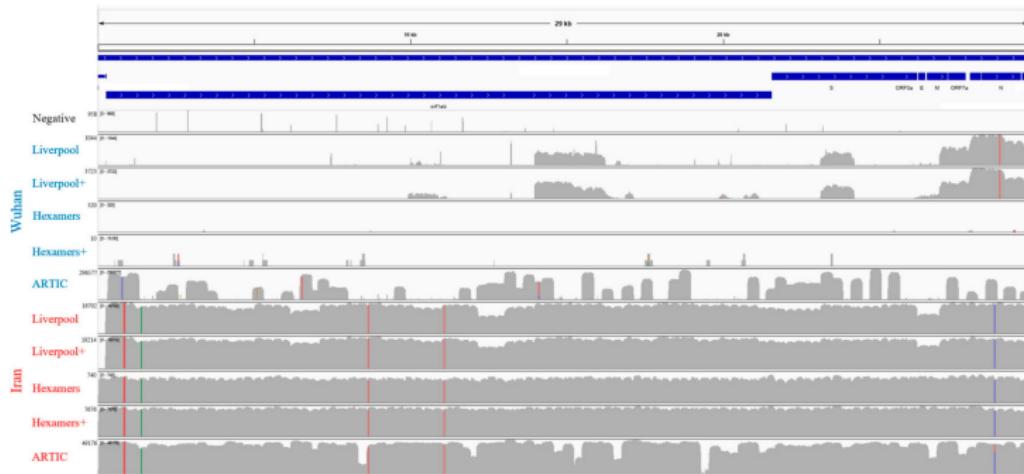
(Nasir et al., 2020)

Which Approach?

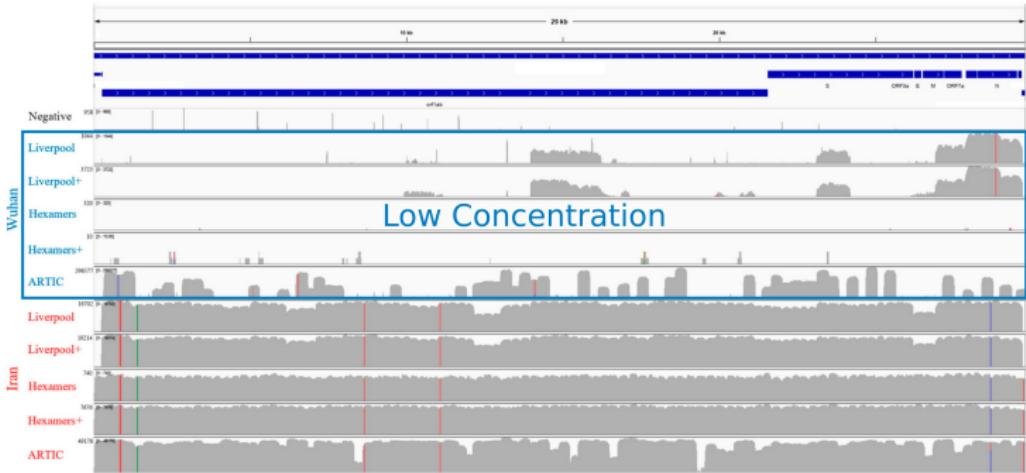


Wuhan Traveller: Low concentration sample
Iran Traveller: High concentration sample
(Nasir et al., 2020)

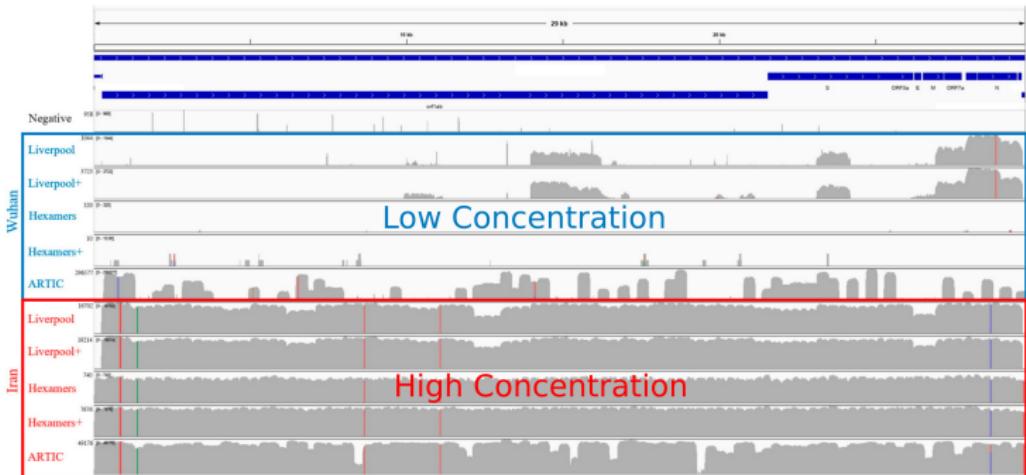
Concentration Matters



Concentration Matters

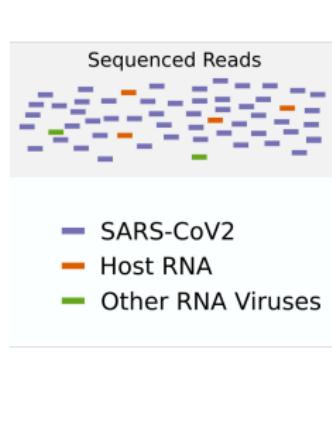


Concentration Matters

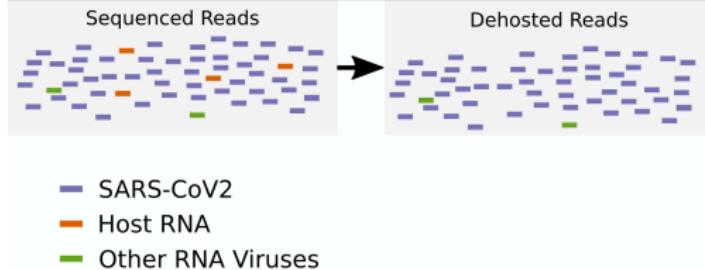


**What do I do with all these
reads?**

Assembling SARS-CoV2



Assembling SARS-CoV2

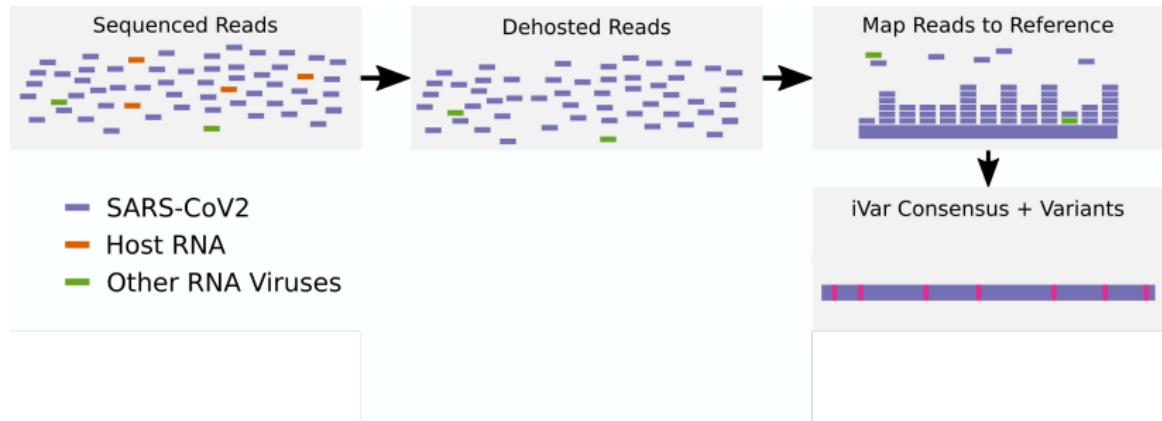


Assembling SARS-CoV2

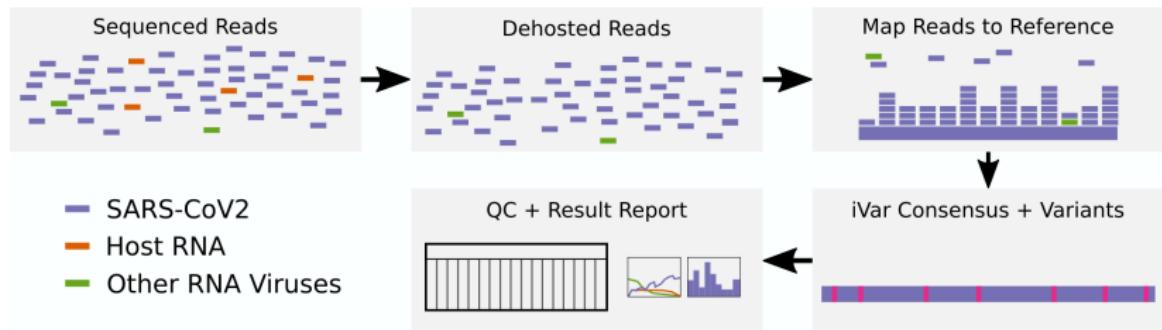


- SARS-CoV2
- Host RNA
- Other RNA Viruses

Assembling SARS-CoV2



Assembling SARS-CoV2



**OK, that is 1 genome, what
about 1000s?**

SIGNAL you need to scale-up

fmaguire Merge pull request #102 from jaleezyy/dev

3 commits · 3 days ago · 359 commits

.github/workflows	Update github action to SIGNAL name	7 months ago
conda_envs	Integrate latest version of ncov-tools	13 days ago
ncov-tools @ 649d96e	Integrate latest version of ncov-tools	13 days ago
resources	update workflow figure to composite host removal	4 months ago
scripts	Integrate latest version of ncov-tools	13 days ago
.gitattributes	c19_postprocess.py: add version string	7 months ago
.gitmodules	restored submodule download	6 months ago
PIPELINE.md	Add png version of workflow for embedding	5 months ago
README.md	Update README closes #99	13 days ago
Snakefile	Integrate latest version of ncov-tools	13 days ago
config.yaml	Update to ivar 1.3	13 days ago
example_config.yaml	Make bresseq optional and fix postprocessing to handle optionality miss...	13 days ago
example_sample_table.csv	Update to ivar 1.3	13 days ago
generate_sample_table.sh	fix typo	9 days ago

Readme

Releases

v1.1.0 (Latest)
3 days ago
+ 3 releases

Packages

No packages published.
Publish your first package

Contributors

Languages

Language	Percentage
Python	72.7%
Shell	18.1%
Tex	4.7%
Perl	1.8%
Dockerfile	0.4%

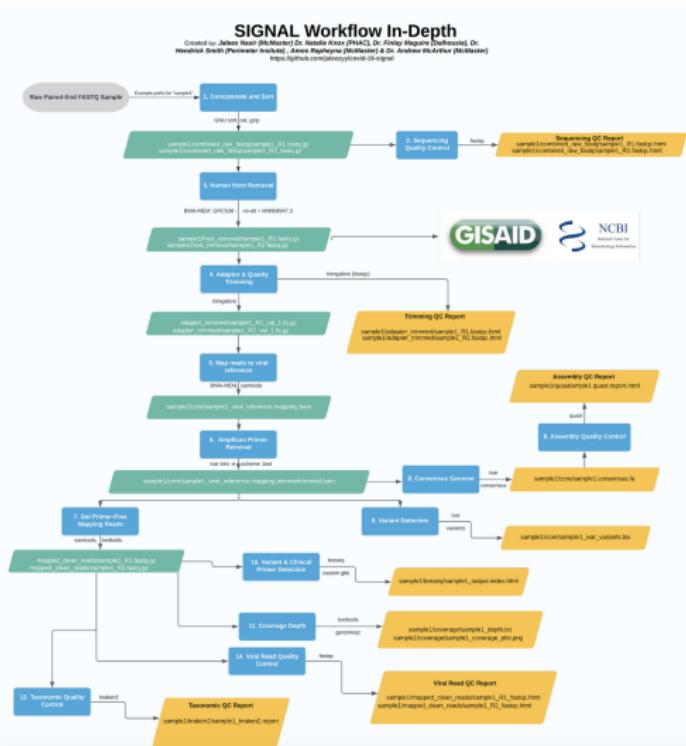
Prototyping: Jalees Nasir/Amos Raphenya

Development: Finlay Maguire/Kendrick Smith

Contributors: Natalie Knox/Jared Simpson

Main Users: Emily Panousis/Natalie Knox

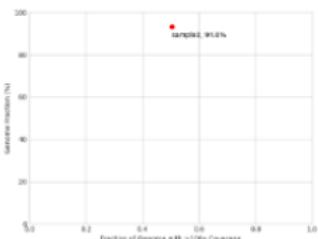
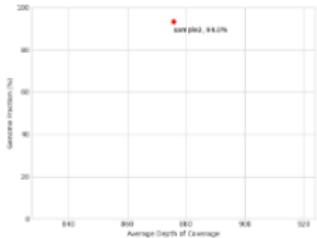
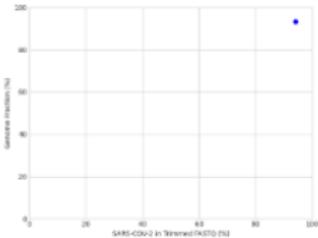
SIGNAL



Flowchart created by Emily Panousis

SIGNAL

SARS-CoV-2 Human Genome Assembly Line (SIGNAL), version b158164 (<https://github.com/talherry/covid-19-signal>)



Sample	Data Volume	Quality Control Flags				Kraken2	QUAST	Variant (Var)	Variant (Briefed)
		Genome	Reads	Depth Variants ORF	Parse: Pass	Genome Coverage (% per 100 bp)	Average Depth Coverage		
sample2	103209 103209	PASS	PASS	FAIL, FAIL	PASS	97%	7133.06	075.7	GRIMMETT
sample2	103209 103209	PASS	PASS	FAIL, FAIL, FAIL	PASS	97%	7133.06	075.7	GRIMMETT

SIGNAL

Sample	Data Volume	Quality Control Flags										Kraken2	QUAST				Variants (iVar)	Variants (BreSeq)	
		Raw	Post	Genome	No.	Depth	Variants	ORF	FastQC	FastQC	90%	99%	Reads	Genome	Length	N's per	Average		
		Data	Trim	(read	Indels	>2000	>90%	Frameshifts	quality	cov	cov	SARS-CoV-2	(bp)	Fraction	100 kbp	Depth of	Coverage		
sample1	103299	103299	PASS	PASS	FAIL	FAIL	FAIL	PASS	PASS	FAIL	WARN	94.05	29903.0	93.188	7133.06	875.7	G11083T		
sample2	103299	103299	PASS	PASS	FAIL	FAIL	FAIL	PASS	PASS	FAIL	WARN	94.05	29903.0	93.188	7133.06	875.7	G11083T		

SIGNAL

SARS-CoV-2 Illumina GeNome Assembly Line (SIGNAL), version b158164 (<https://github.com/jaleezyy/covid-19-signal>)

Sample: sample1

Data Volume [[sample1_trim_galore.log](#)]

Raw Data (read pairs)	103299
Raw Data (base pairs)	28481100
Post Primer Removal (read pairs)	103299
Post Primer Removal (base pairs)	28353773
Post Trim (read pairs)	103299

Quality Control Flags

Genome Fraction greater than 90%	PASS
No indels detected (maximum length 85bp)	PASS
Depth of coverage >= 2000x	FAIL
All variants with at least 90% frequency among reads	FAIL
Frameshifts in SARS-CoV-2 open reading frames	FAIL
Reads per base sequence quality	PASS
Sequencing adapter removed	PASS
At least 90% of positions have coverage >= 100x	FAIL
At least 90% of positions have coverage >= 1000x	WARN

FASTQC Flags [[sample1_R1_val_1_fastqc.html](#) | [sample1_R2_val_2_fastqc.html](#)]

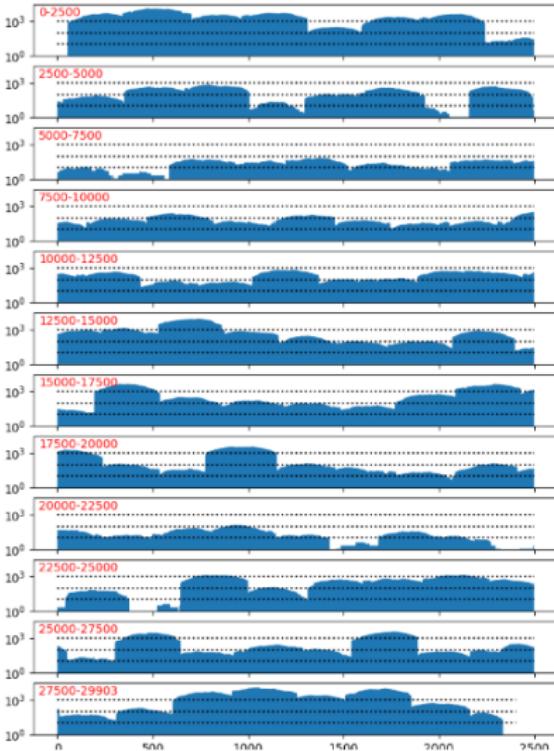
Per tile sequence quality	FAIL
Per base sequence content	FAIL
Per sequence GC content	FAIL
Sequence Length Distribution	WARN

Kraken2 [[sample1_kraken2.report](#)]

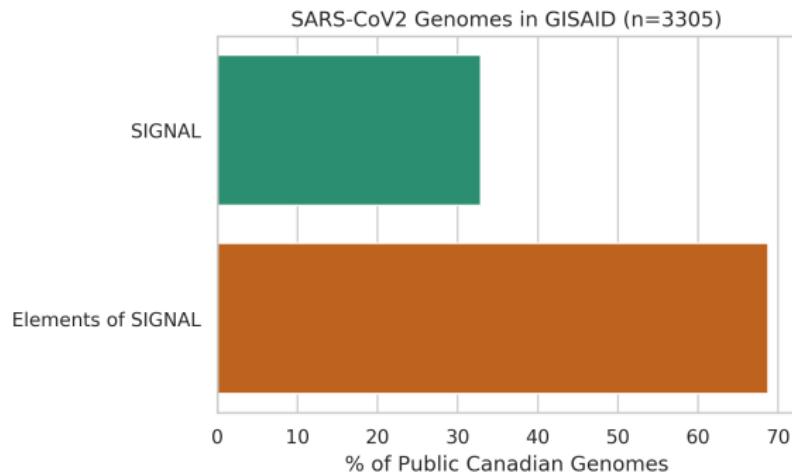
Reads SARS-CoV-2 (%)	94.05
Genome Length (bp)	29903.0
Genome Fraction (%)	93.188
N's per 10k kbp	7133.06
Genomic Features	16 + 8 part
Mismatches	1.0
Mismatches per 100 kbp	3.59
Indels	0.0
Indels per 100 kbp	0.0
Average Depth of Coverage	875.7
Fraction with 0 coverage	0.014
Fraction with 1x-10x coverage	0.063
Fraction with 11x-100x coverage	0.422
Fraction with 101x-1000x coverage	0.307
Fraction with 1001x-2000x coverage	0.051
Fraction with 2001x-10000x coverage	0.134
Fraction with > 10000x coverage	0.008
5' Ns	54
3' Ns	67

Variants in Consensus Genome (iVar)

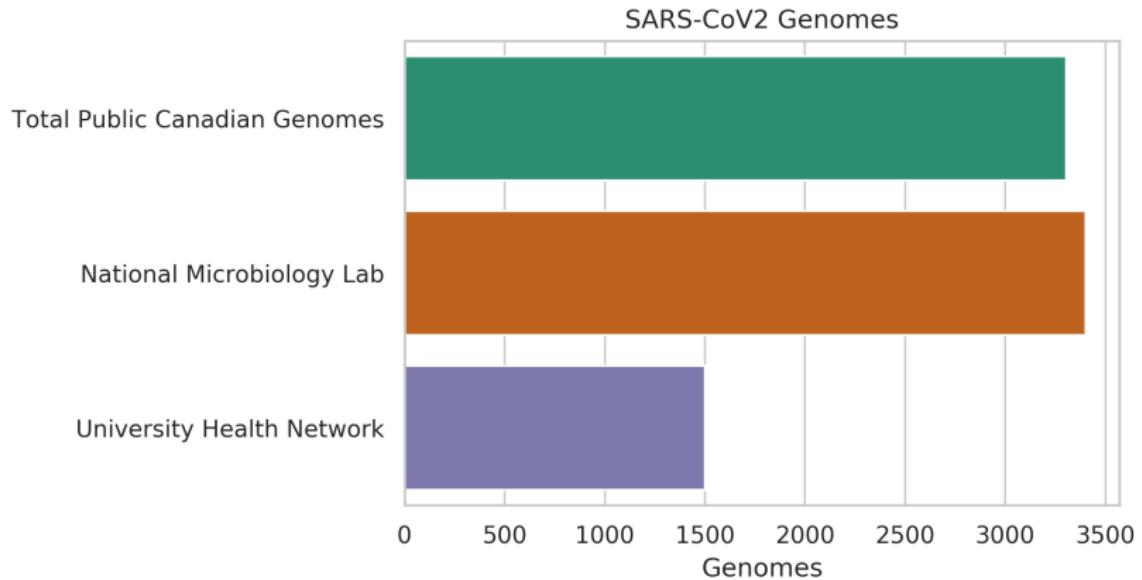
G11083T



Strong SIGNAL



Strong SIGNAL



Data: ✓

Metadata: ?

Metadata is messy

unknown	152280
Male	28899
Female	26423
Unknown	766
male	76
Male	55
female	17
Unkown	5
Woman	5
Male	3
unknown	2
FEmale	2
Famle	1
Famale	1
MAle	1
Not reported to protect privacy	1
Female	1
unknow	1

Actual values in GISAID

Metadata is messy



JINGXIN® BioelectronSeq 4000 Sequencer System | Next- Generation Sequencing System

From

JINGXIN® BioelectronSeq 4000 System (abbreviation: BES 4000) is a next-generation sequencing system developed and produced by the cooperation between CapitalBio Corporation and Thermo Fisher Scientific Inc, in China.

IN STOCK

Actual values in GISAID

PHA4GE Contextual SARS-CoV2 Metadata Scheme



Figure by Duncan MacCannell (Griffiths et al., 2020). Used by CaCOGeN, US-SPHERES, AusTrakka, South Africa's Baobab LIMS

PHA4GE Contextual SARS-CoV2 Metadata Scheme

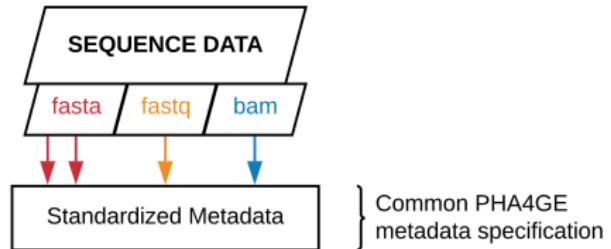


Figure by Duncan MacCannell (Griffiths et al., 2020). Used by CaCOGeN, US-SPHERES, AusTrakka, South Africa's Baobab LIMS

PHA4GE Contextual SARS-CoV2 Metadata Scheme

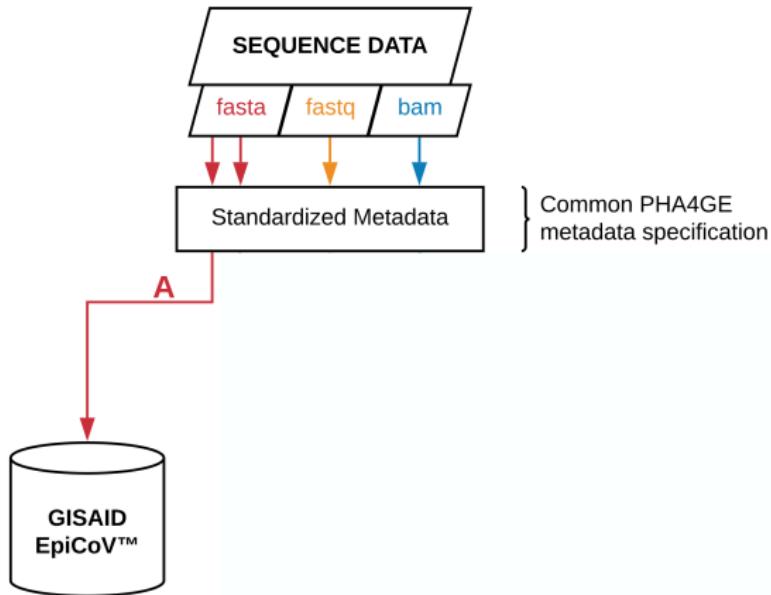


Figure by Duncan MacCannell (Griffiths et al., 2020). Used by CaCOGeN, US-SPHERES, AusTrakka, South Africa's Baobab LIMS

PHA4GE Contextual SARS-CoV2 Metadata Scheme

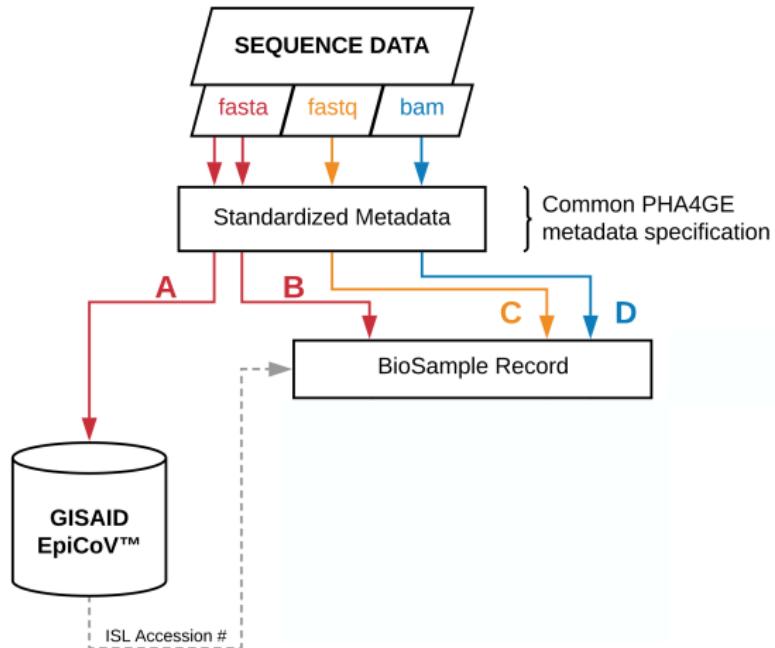


Figure by Duncan MacCannell (Griffiths et al., 2020). Used by CaCOGeN, US-SPHERES, AusTrakka, South Africa's Baobab LIMS

PHA4GE Contextual SARS-CoV2 Metadata Scheme

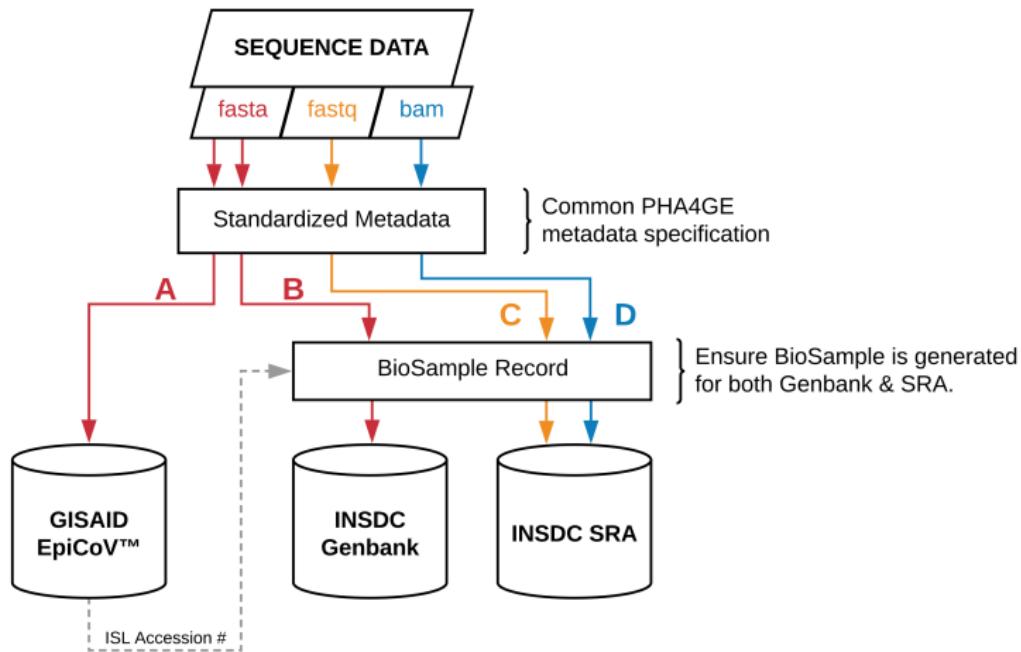


Figure by Duncan MacCannell (Griffiths et al., 2020). Used by CaCOGeN, US-SPHERES, AusTrakka, South Africa's Baobab LIMS

Data: ✓

Metadata: ✓

Analysis: ?

Genomic Epidemiology

GISAID/Nextstrain Genomes + Metadata

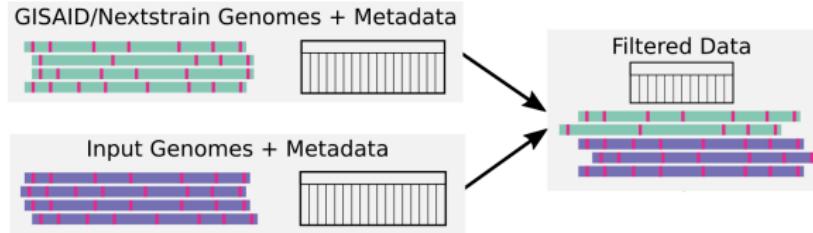


Input Genomes + Metadata



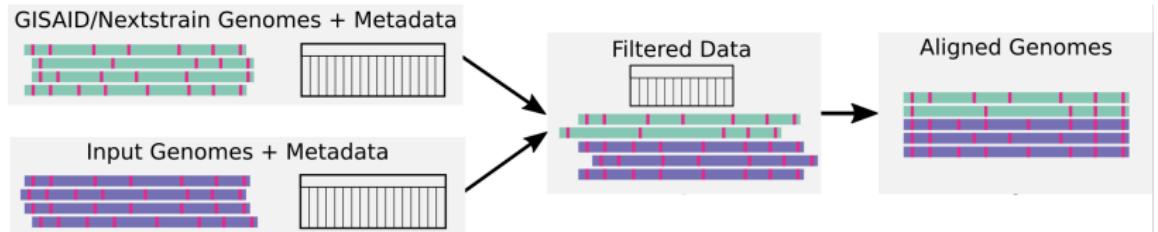
Nextstrain's Augur/Auspice

Genomic Epidemiology



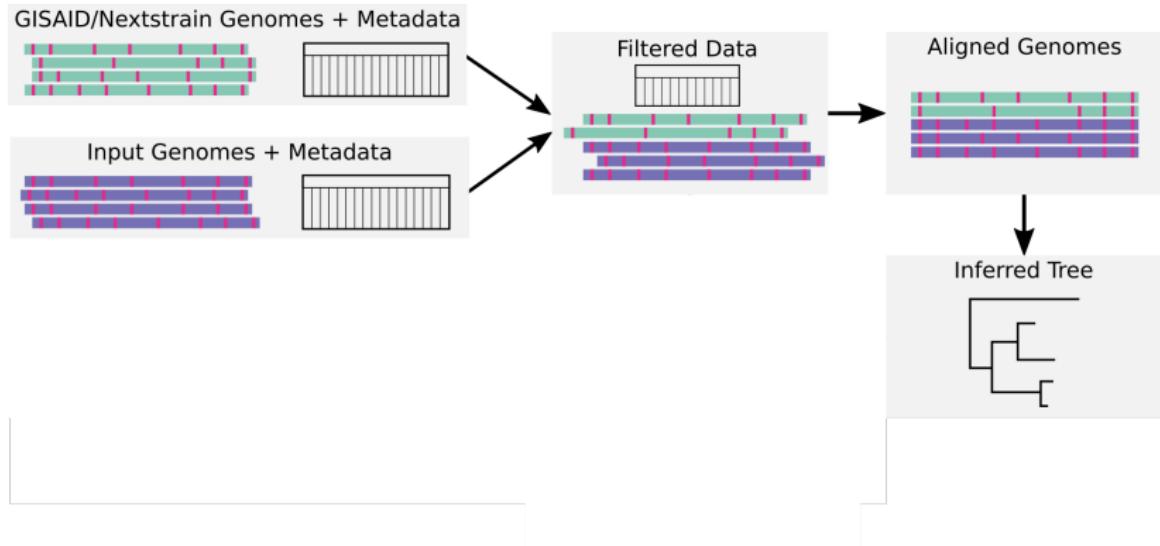
Nextstrain's Augur/Auspice

Genomic Epidemiology



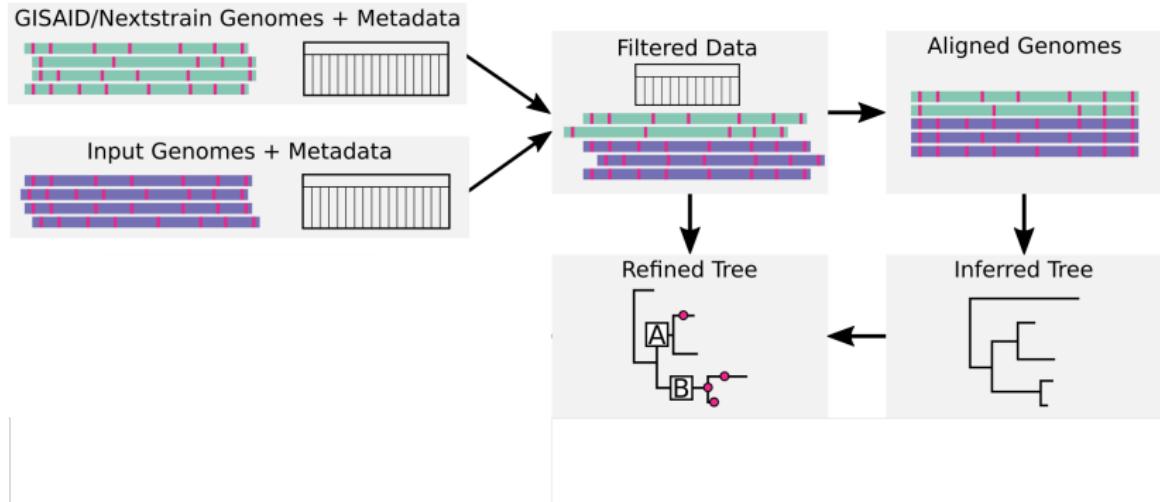
Nextstrain's Augur/Auspice

Genomic Epidemiology



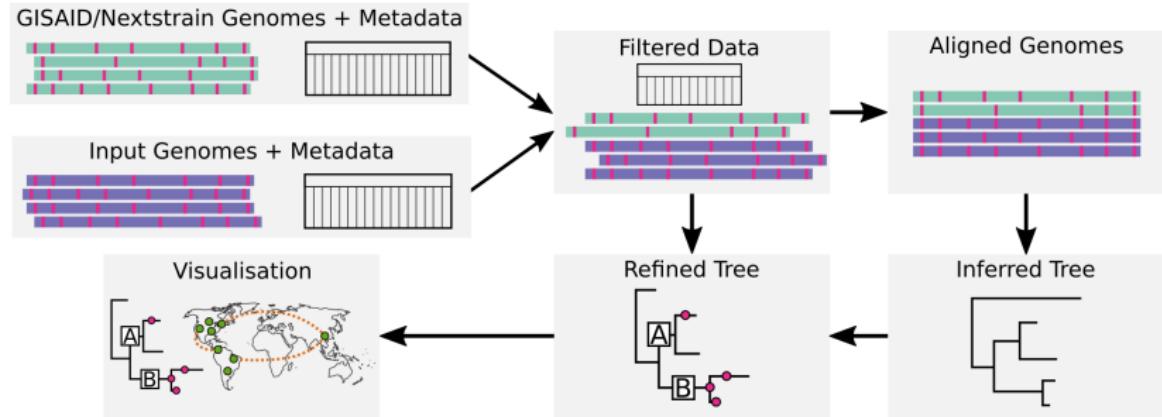
Nextstrain's Augur/Auspice

Genomic Epidemiology



Nextstrain's Augur/Auspice

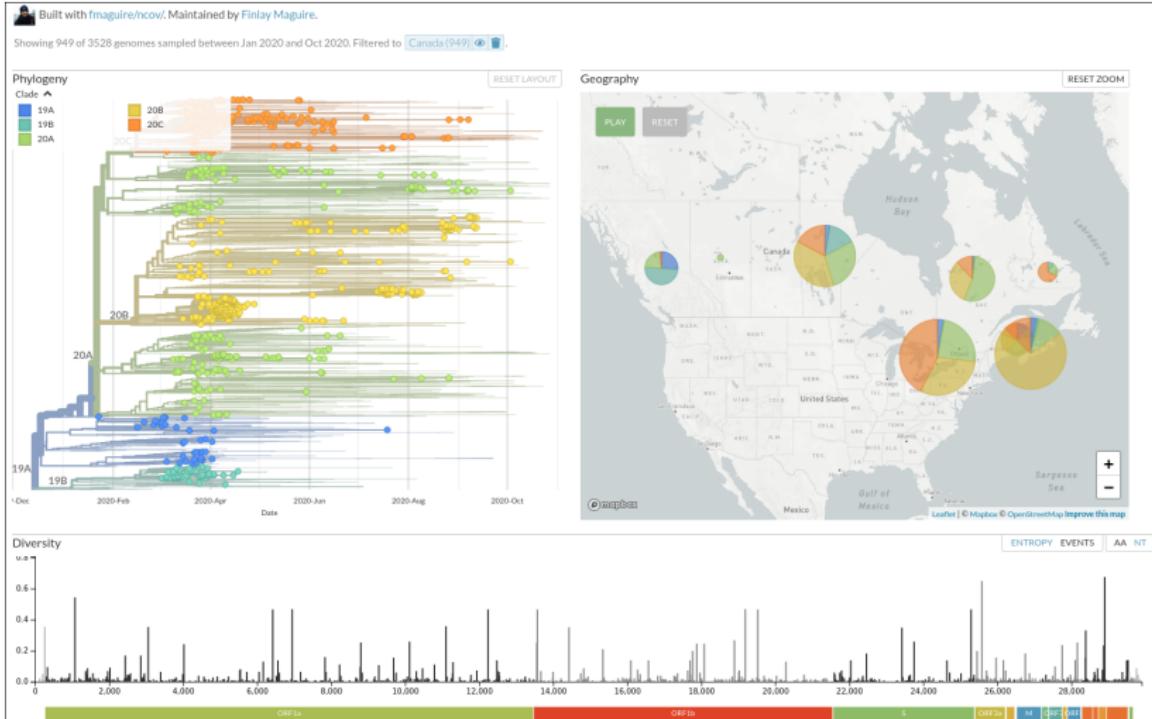
Genomic Epidemiology



Nextstrain's Augur/Auspice

On-going Public Health Analyses

Monthly Surveillance Builds



Regular builds for NSHA and UHN (moving to NML hosting)

Identifying Diagnostic Primer Failures



nodrogluap Merge pull request #2 from fmaguire/master ...

8873de6 on Oct 23 37 commits



First 28 reinfection cases

last month



LICENSE

Initial commit

6 months ago



README.md

Fix nextstrain exclude link in README

last month



iedb_csv2fasta

Initial commit of csv -> fasta conversion script for epiropes

2 months ago



pokay

Fixed 5' alignment extension bug, 3' end pos reporting

2 months ago

README.md

Pokay

Report salient DNA (qPCR primer/probe) or amino acid (epitope) mismatches info against a set of pathogen isolate genomes.

Pronunciation

/pou'keɪ/ as in the [delicious Hawaiian raw fish dish](#), or Pokay as in "my PCR design is okay" or "my epitope is okay".

Paul Gordon

Report salient qPCR primer/probe or immune epitope mismatches info against a set of pathogen isolate genomes

Readme

GPL-3.0 License

Releases

No releases published

Packages

No packages published

Contributors

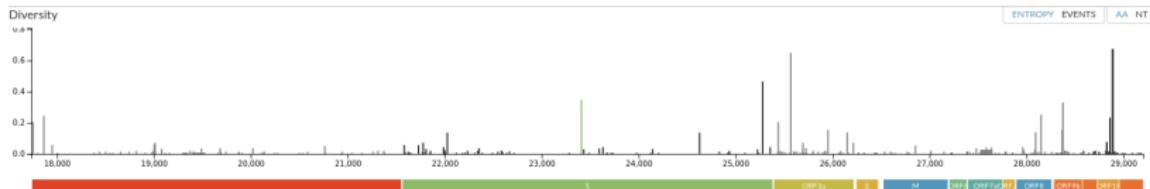
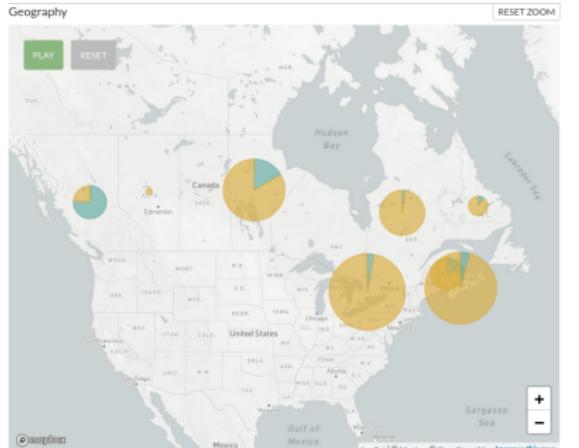
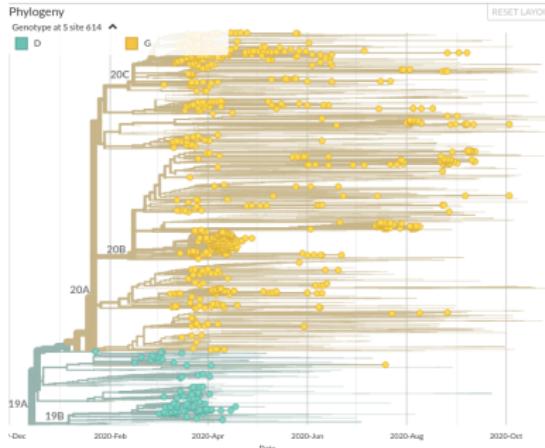
nodrogluap Paul Gordon

fmaguire Finlay Maguire

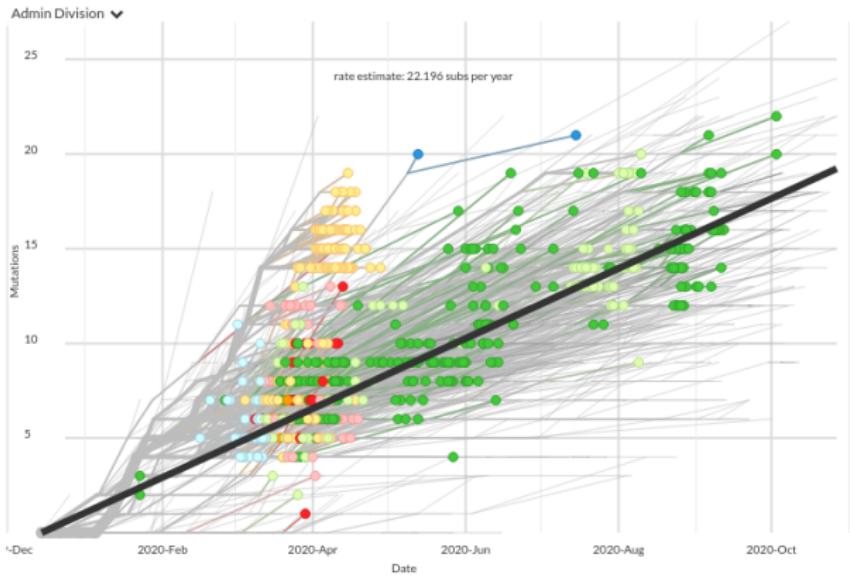
Investigating Specific Variants

Built with [fmaguire/ncoiv](#); Maintained by Finlay Maguire.

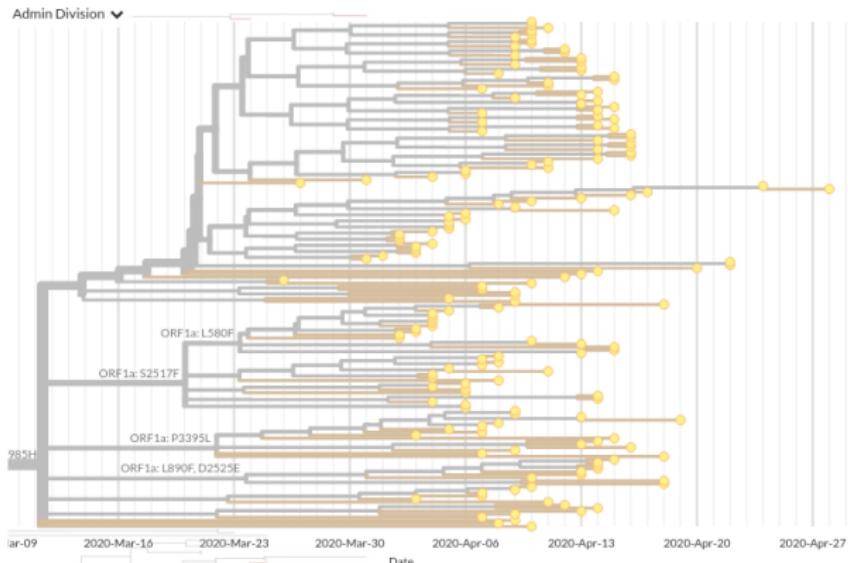
Showing 949 of 3528 genomes sampled between Jan 2020 and Oct 2020. Filtered to [Canada \(949\)](#).



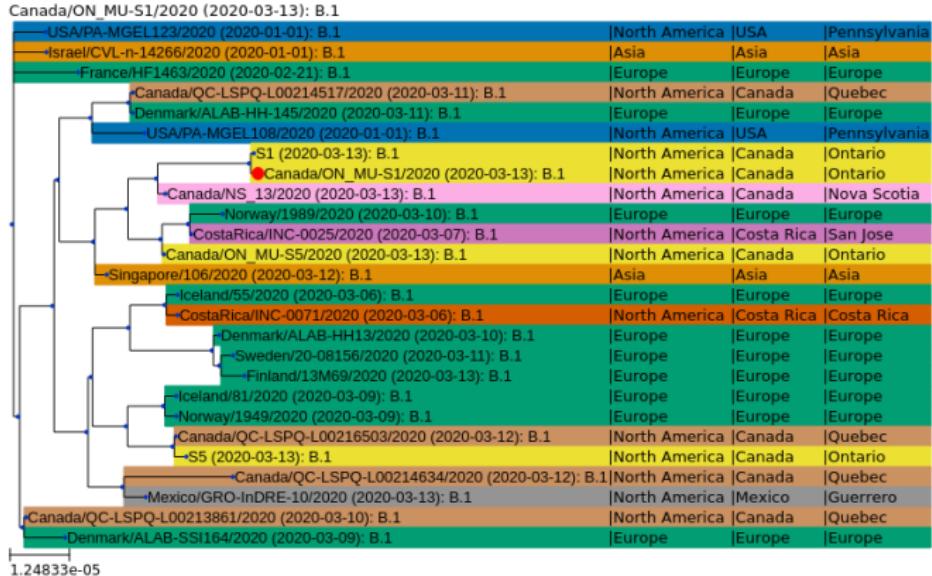
Outbreak Analysis



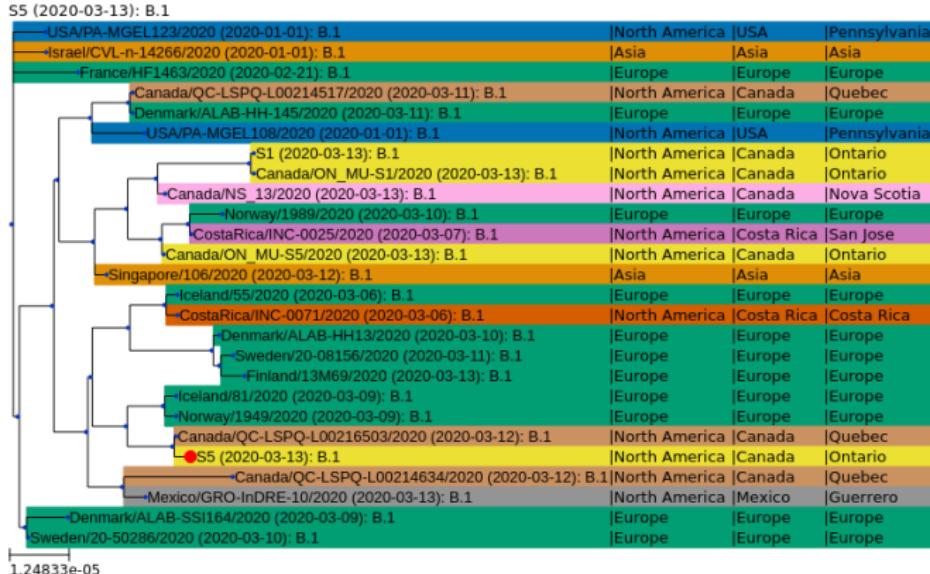
Outbreak Analysis



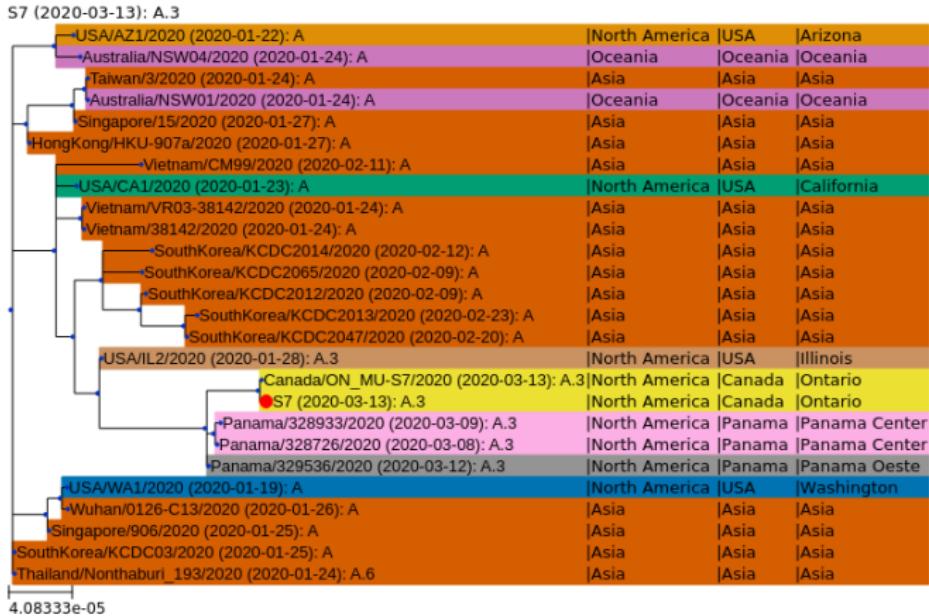
Initial Ontario Cases



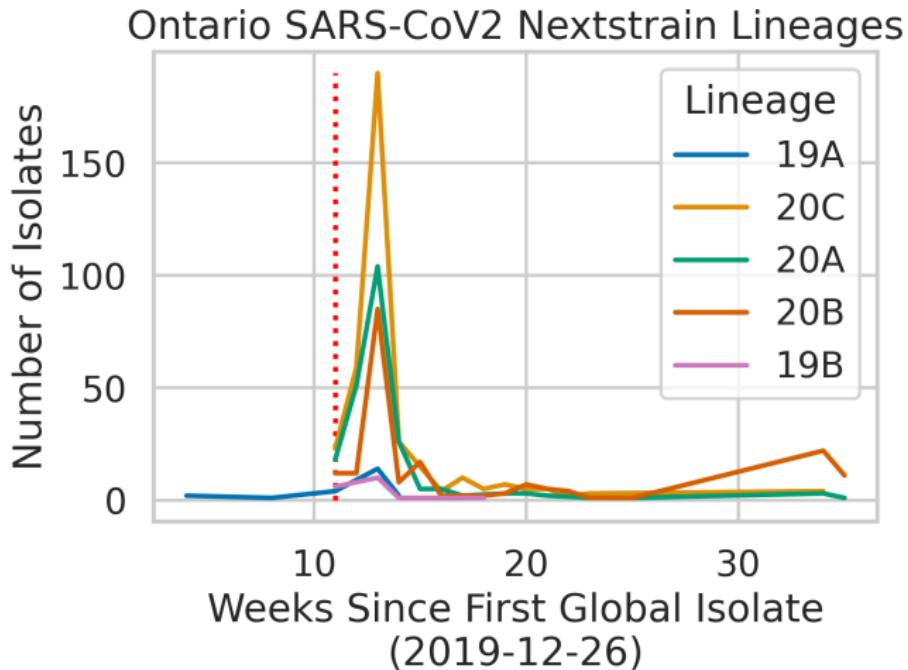
Initial Ontario Cases



Initial Ontario Cases



Investigating Policy Impact



Summary

- Genomic Epidemiology is an important public health tool

Summary

- Genomic Epidemiology is an important public health tool
- ARTICv3 Amplicon tiling is an effective sequencing method

Summary

- Genomic Epidemiology is an important public health tool
- ARTICv3 Amplicon tiling is an effective sequencing method
- Workflow engines are key to scaling SARS-CoV-2 genomics

Summary

- Genomic Epidemiology is an important public health tool
- ARTICv3 Amplicon tiling is an effective sequencing method
- Workflow engines are key to scaling SARS-CoV-2 genomics
- Nextstrain's Augur+Auspice is a powerful tool to interactive explore GenEpi results

Summary

- Genomic Epidemiology is an important public health tool
- ARTICv3 Amplicon tiling is an effective sequencing method
- Workflow engines are key to scaling SARS-CoV-2 genomics
- Nextstrain's Augur+Auspice is a powerful tool to interactive explore GenEpi results
- Actively using these methods for retrospective and prospective public health analyses

Acknowledgements

Acknowledgements 1

- Dalhousie: Alex Manuele, Rob Beiko
- NSHA: Jason LeBlanc, Daniel Gaston
- US CDC: Duncan MacCannell, Lee Katz
- US FDA: Ruth Timme
- McMaster University: Jalees Nasir, Beth Culp, Amogelang Raphenya, Emily Panousis and Andrew McArthur
- Sunnybrook: Samira Mubareka
- BCCDC: Emma Griffiths
- Ontario Institute of Cancer Research: Jared Simpson, Richard de Borja
- University of Calgary: Paul Gordon
- Public Health England: Adam Witney
- COG UK: Samuel Nicholls

Acknowledgements 2

- Africa CDC: Paul E Oluniyi and Idowu B Olawoye
- South African Bioinformatics Institute: Peter van Heusden
- AusTrakka: Anders Gonçalves da Silva
- Quadram Institute: Andrew Page
- Public Health Agency of Canada: Natalie Knox
- Argentinian National Laboratories and Institutes of Health: Josefina Campos
- Institute of Molecular Medicine (Portugal): Ines Mendes
- **Funding:** Bill and Melinda Gates Foundation, Pan-American Health Organisation, Genome Canada (CanCOGeN-V), Donald Hill Family Fellowship

References

- Cassini, A., Högberg, L. D., Plachouras, D., Quattrocchi, A., Hoxha, A., Simonsen, G. S., Colomb-Cotinat, M., Kretzschmar, M. E., Devleesschauwer, B., Cecchini, M., et al. (2019). Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the eu and the european economic area in 2015: a population-level modelling analysis. *The Lancet Infectious Diseases*, 19(1):56–66.
- Griffiths, E. J., Timme, R. E., Page, A. J., Alikhan, N.-F., Fornika, D., Maguire, F., Mendes, C. I., Tausch, S. H., Black, A., Connor, T. R., et al. (2020). The pha4ge sars-cov-2 contextual data specification for open genomic epidemiology.

Nasir, J. A., Kozak, R. A., Aftanas, P., Raphenya, A. R., Smith, K. M., Maguire, F., Maan, H., Alruwaili, M., Banerjee, A., Mbareche, H., et al. (2020). A comparison of whole genome sequencing of sars-cov-2 using amplicon-based sequencing, random hexamers, and bait capture. *Viruses*, 12(8):895.

Organization, W. H. et al. (2014). *Antimicrobial resistance: global report on surveillance*. World Health Organization.

Silver, L. L. (2011). Challenges of antibacterial discovery. *Clinical microbiology reviews*, 24(1):71–109.

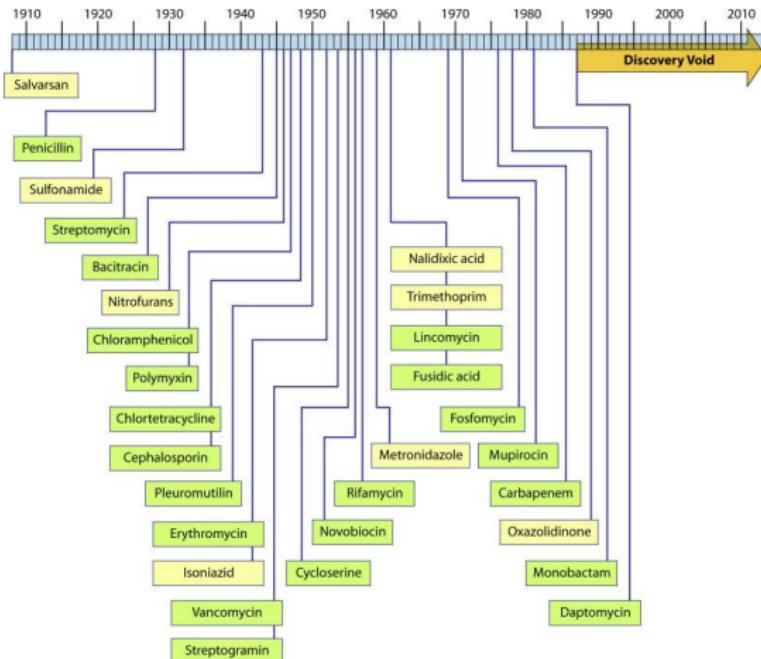
Backup

Caveat on data

	Antibacterial resistance	Studies included in SR (n)	Country income status ^a	
			Low income (n)	Lower-middle income (n)
<i>Escherichia coli</i>	3 rd generation cephalosporin-resistant	25	0	0
	Fluoroquinolone-resistant	12	0	0
<i>Klebsiella pneumoniae</i>	3 rd generation cephalosporin-resistant	24	0	0
	Carbapenem-resistant	13	0	0
<i>Staphylococcus aureus</i>	Methicillin-resistant	147	0	2

Studies included in systematic review on healthcare implications of AMR
Organization et al. (2014)

New Antibiotics



Silver (2011). Teixobactin in 2015, still in trials.