

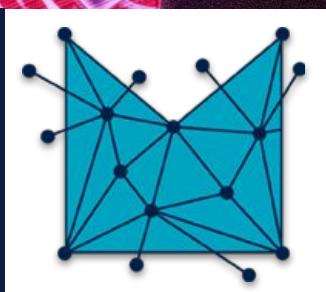
wellcome
connecting
science

ACORN.

Web Tools for Genomic Epidemiology & Surveillance

Julio Diaz Caballero

CGPS, University of Oxford



Epidemiology in Space and Time

Local or Short-Term

- Is it a hospital outbreak caused by a single strain or independent infections caused by unrelated strains?
- What is the source of an outbreak?
- Are the strains resistant to antibiotics?

Global or Long-Term

- What is the worldwide origin of the prevalent lineage of a pathogen?
- Was the disease acquired by a patient during travel?
- When did resistance to an antibiotic emerge in a pathogen species?

Epidemiology in Space and Time – WHERE?

Local or Short-Term

- Is it a hospital outbreak caused by a single strain or independent infections caused by unrelated strains?
- What is the **source** of an outbreak?
- Are the strains resistant to antibiotics?

Global or Long-Term

- What is the worldwide **origin** of the prevalent lineage of a pathogen?
- Was the disease acquired by a patient during **travel**?
- When did resistance to an antibiotic emerge in a pathogen species?

Epidemiology in Space and Time – WHEN?

Local or Short-Term

- Is it a hospital outbreak caused by a single strain or independent infections caused by unrelated strains over **time**?
- What is the source of an outbreak?
- Are the strains resistant to antibiotics?

Global or Long-Term

- What is the worldwide origin of the prevalent lineage of a pathogen?
- Was the disease acquired by a patient during travel?
- **When** did resistance to an antibiotic emerge in a pathogen species?

Epidemiology in Space and Time – WHAT?

Local or Short-Term

- Is it a hospital outbreak caused by a **single strain or independent infections** caused by unrelated strains?
- What is the source of an outbreak?
- Are the strains resistant to antibiotics?

Global or Long-Term

- What is the worldwide origin of the prevalent **lineage of a pathogen?**
- Was the disease acquired by a patient during travel?
- When did resistance to an antibiotic emerge in a pathogen species?

Epidemiology in Space and Time – HOW?

Local or Short-Term

- Is it a hospital outbreak caused by a single strain or independent infections caused by unrelated strains?
- What is the source of an outbreak?
- Are the strains **resistant to antibiotics**?

Global or Long-Term

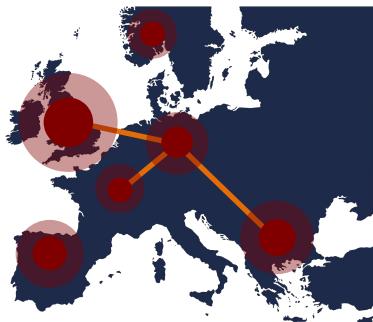
- What is the worldwide origin of the prevalent lineage of a pathogen?
- Was the disease acquired by a patient during travel?
- When did **resistance to an antibiotic** emerge in a pathogen species?

Epidemiology in Space and Time

Global



International



National

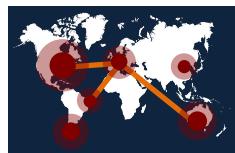


Local



Epidemiology in Space and Time

Global



International



National



Local



Long Term



Short Term

Information for Pathogen Surveillance

WHEN?
date and time

WHAT?
typing/phylogeny

WHERE?
location

HOW?
resistance

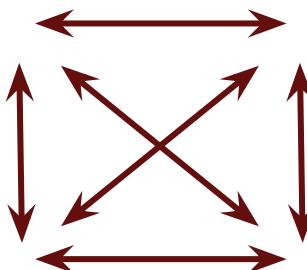
How do we bring data together?

WHEN?
date and time

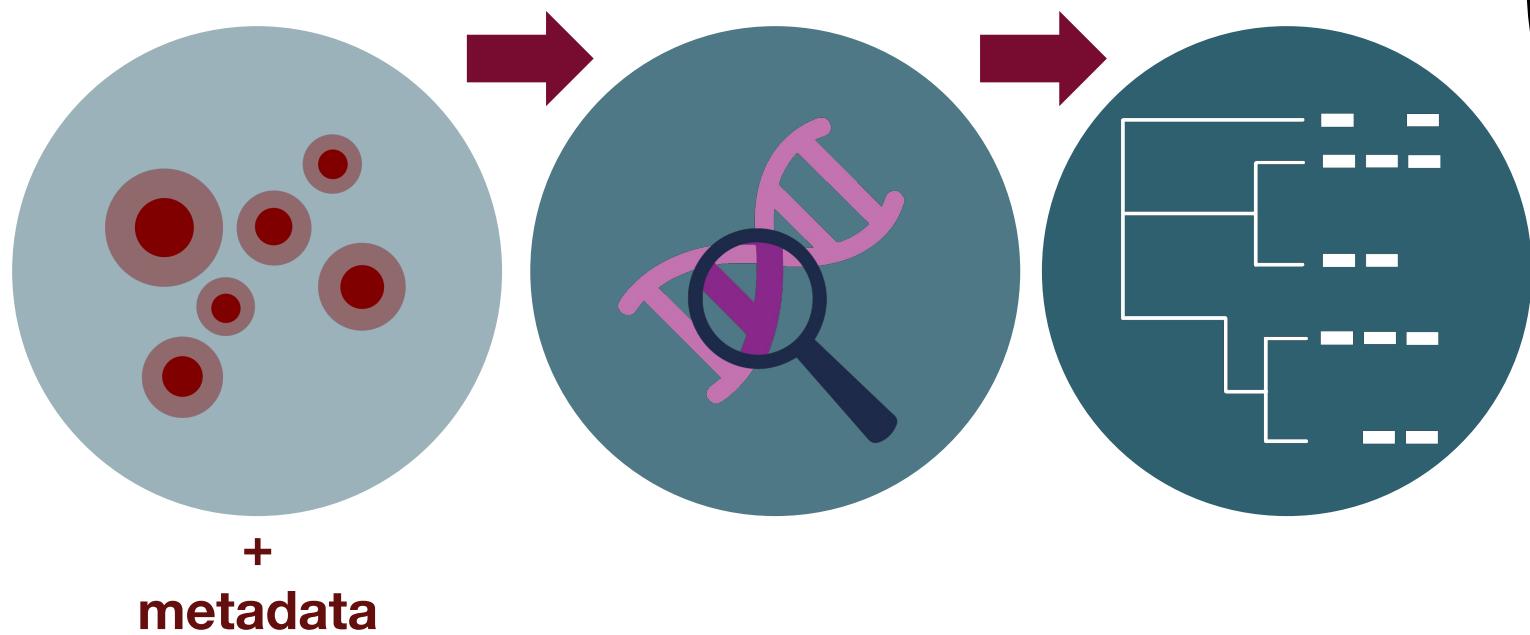
WHAT?
typing/phylogeny

WHERE?
location

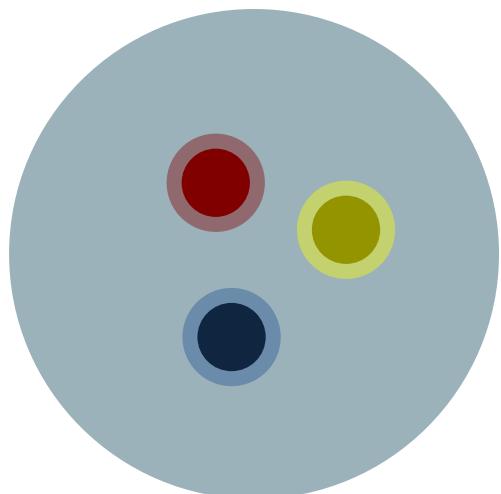
HOW?
resistance



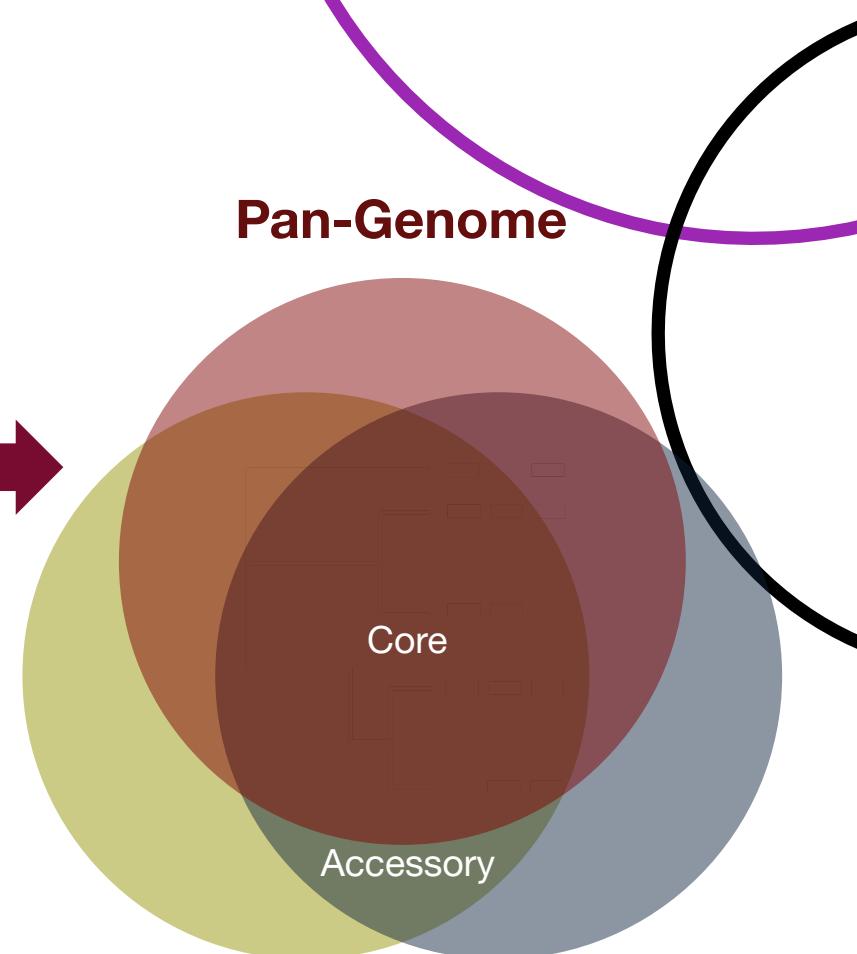
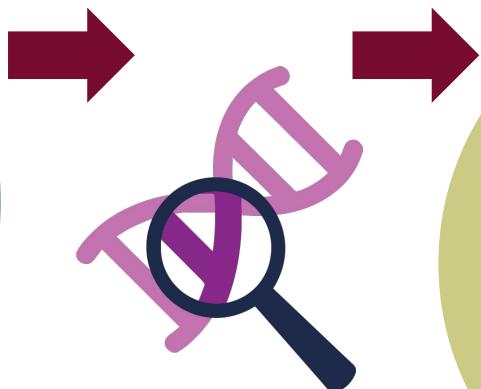
Genomic Epidemiology Method



Genomic Epidemiology Method



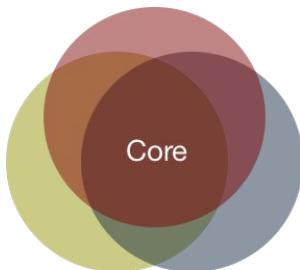
+
metadata



Genomic Epidemiology

Method

Clustering (tree)
based on core SNPs
(single nucleotide polymorphisms)



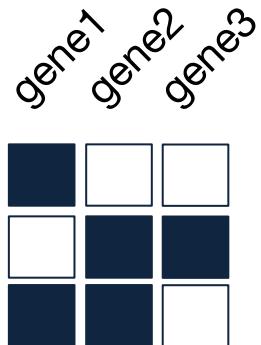
1 AATCGCTTACGACCAG...

2 AATGGCTTATGACAAG...

3 AATCGCTTATGACAAG...

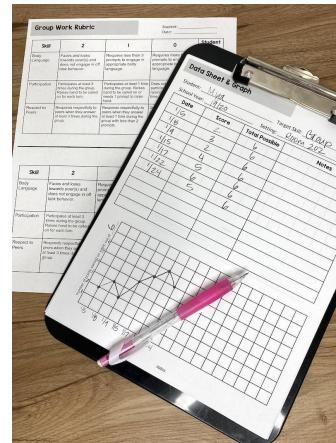
* * *
Newick Format ((2,3)1);

Distribution of accessory genes (e.g., antibiotic resistance genes, virulence genes)



Genomic Epidemiology Challenges

DATA COLLECTION



Google Drive

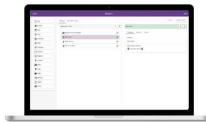




epicollect

<https://five.epicollect.net/>

- Free
- Web + Mobile based
- No specific data type
- No specific pathogen



Create



Collect

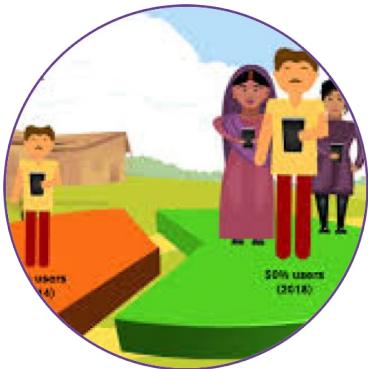


View



15,000+ projects
30,000+ users

Project Samples



RURAL SURVEY 2018

Understanding the sentiments and thoughts of rural population in India

30591 entries

SOUTHERN REGION POLES

Low voltage poles collection in Zimbabwe

76103 entries





◀ Back



✖ Exit

- Text
- Numeric
- Phone
- Date
- Time
- Dropdown
- Radio
- Checkbox
- Search
- Text Box
- Readme
- Location
- Photo
- Audio
- Video
- Barcode
- Branch

Sources ✓ Add child form +

undo

SAVE PROJECT

Form > Sources

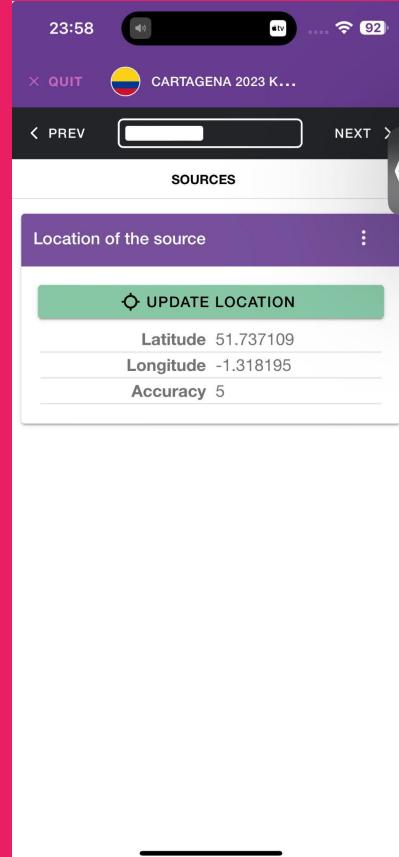
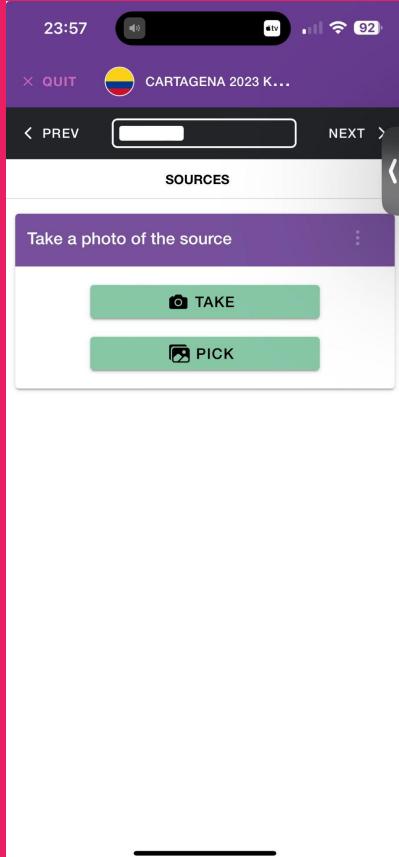
✖ ✎ ⏮

- Collector's Name and Last Name ✓
- Location Type ✓
- Please describe location ✓
- Type of Source ✓
- Please describe source ✓
- Take a photo of the source ✓
- Location of the source ✓
- Ward Number ✓
- Has a sample been collected? ✓
- Why wasn't it possible to collect a sample? ✓
- Please provide reason ✓
- Collection Date ✓
- Scan the barcode on the sample tube ✓

No input is currently selected

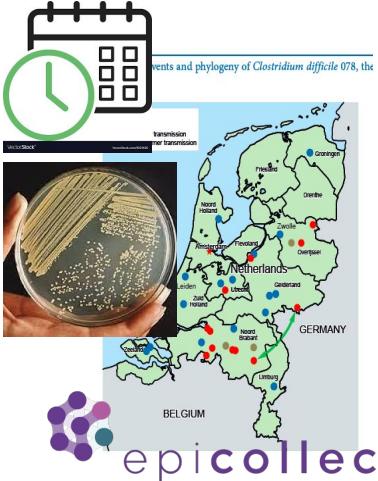


wellcome
connecting
science



Genomic Epidemiology Challenges

DATA INTEGRATION



idsheet

v Insert Format Data Tools Add-ons Help All changes saved in Drive

	A	B	C	D	E	F
1	First name	Last name	State	City	Street Address	Join date
2	Ann	Cold	WI	Springfield	7 J Dr.	2/2/1990
3	Bob	New	NC	Sit	1411 Lincoln St.	7/10/2015
4	Ard	Toothache	DE	Blob	787 An Tw	12/1/1980
5	JJ	Key	AK	Tilly	1 Kool-aid St.	11/25/2017
6	Po		7 HI	Boston	55 Globe	2/3/1990
7	III	Jeans	ND	Tub	1000 Broadway	8/8/2005
8	Gob 3	4	KY	Tab	2219 Lake Ave.	1/6/1963
9						
10						
11						
12						

+ Secret Public

TABLE 1A

Clostridium difficile type 078 isolates used in this study, the Netherlands, 2002–11 (n=65)

R_L#P	Year	City	RT	Isolate	Source	Related Isolates	Association	ERS130041
8080_2#24	2006	Leiden	078	6072310	Clinic	Non-outbreak	Healthcare	ERS130041
8080_2#25	2006	Nijmegen	078	6086336	Clinic	Non-outbreak	Healthcare	ERS130041
8080_2#26	2007	Leiden	078	7001233	Clinic	Non-outbreak	Healthcare	ERS130041
8080_2#27	2007	Groningen	078	7004578	Clinic	Non-outbreak	Unknown	ERS130041
8080_2#28	2007	Utrecht	078	7005405	Clinic	Non-outbreak	Unknown	ERS130041
8080_2#29	2007	Zwolle	078	7021455	Clinic	Non-outbreak	Healthcare	ERS130041
8080_2#30	2007	Zwolle	078	7044912	Clinic	Non-outbreak	Community	ERS130041
8080_2#31	2007	Zwolle	078	7066827	Clinic	Non-outbreak	Community	ERS130041
8080_2#32	2007	Zwolle	078	7071308	Clinic	Non-outbreak	Healthcare	ERS130041
8080_2#33	2007	Zwolle	078	7086074	Clinic	Non-outbreak	Healthcare	ERS130041
8080_2#34	2007	Leiden	078	7091952	Clinic	Non-outbreak	Healthcare	ERS130041
8080_2#35	2008	Leiden	078	8011061	Clinic	Non-outbreak	Healthcare	ERS130041



TABLE 3

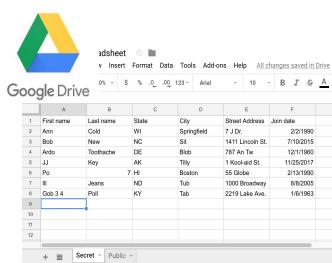
Results of Antimicrobial susceptibility testing

Isolate	Source	Streptomycin	Tn625	Tetracycline	Tn6190	Isolate	Source	Streptomycin	Tn625	Tetracycline	Tn6190
6072310	Clinic	Absent	Absent	Present		129280	Clinic	Absent	Absent	Present	
6086336	Clinic	Absent	Absent	Present		H205	Farmer	Present	Present	Absent	
7001233	Clinic	NT	Absent	Present		B37-3	Pig	Present	Present	Absent	
7004578	Clinic	Absent	Present	Present		S37-37	Clinic	NT	Absent	NT	Present
7005405	Clinic	Present	Present	Absent		47337	Clinic	Absent	Present	Present	Absent
7021455	Clinic	Absent	Present	Present		H102	Farmer	Absent	Absent	Absent	
7044912	Clinic	Present	Absent	Absent		B31-3	Pig	Absent	Absent	Absent	
7066827	Clinic	Absent	Present	Present		B17-3	Pig	Absent	Absent	Absent	
7071308	Clinic	Absent	Present	Absent		H121	Farmer	Absent	Absent	Present	
7086074	Clinic	Absent	Absent	Absent		B27-7	Pig	Absent	Absent	Absent	
7099592	Clinic	Absent	Present	Present		H230	Farmer	Absent	Absent	Present	
8011061	Clinic	Absent	Present	Absent		H189	Farmer	Absent	Absent	Absent	
B011820	Clinic	Absent	Absent	Absent		B23-6	Pig	Absent	Present	Present	
B05728	Clinic	Absent	Present	Present		H205	Farmer	Present	Absent	Absent	
B055344	Clinic	Present	Absent	Absent		B15-1	Pig	Absent	Absent	Absent	



Genomic Epidemiology Challenges

DATA INTEGRATION



Data-flo

TABLE 1
Clostridium difficile type 078 isolates used in this study, the Netherlands, 2002–11 (n=65)

Isolate	Year	City	RT	Isolate	Source	Related Isolates	Association
8080_2#24	2006	Leiden	078	6072310	Clinic	Non-outbreak	Healthcare
8080_2#35	2006	Nijmegen	078	6086336	Clinic	Non-outbreak	Healthcare
8080_2#26	2007	Leiden	078	7001233	Clinic	Non-outbreak	Healthcare
8080_2#7	2007	Groningen	078	7004578	Clinic	Non-outbreak	Unknown
8080_2#28	2007	Utrecht	078	7095405	Clinic	Non-outbreak	Unknown
8080_2#29	2007	Zwolle	078	7921455	Clinic	Non-outbreak	Healthcare
8080_2#30	2007	Zwolle	078	7044912	Clinic	Non-outbreak	Community
8080_2#31	2007	Zwolle	078	7066827	Clinic	Non-outbreak	Community
8080_2#32	2007	Zwolle	078	7071508	Clinic	Non-outbreak	Healthcare
8080_2#33	2007	Zwolle	078	7086074	Clinic	Non-outbreak	Healthcare
8080_2#34	2007	Leiden	078	7091952	Clinic	Non-outbreak	Healthcare
8080_2#35	2008	Leiden	078	8011061	Clinic	Non-outbreak	Healthcare



TABLE 2
Results of Antimicrobial susceptibility testing

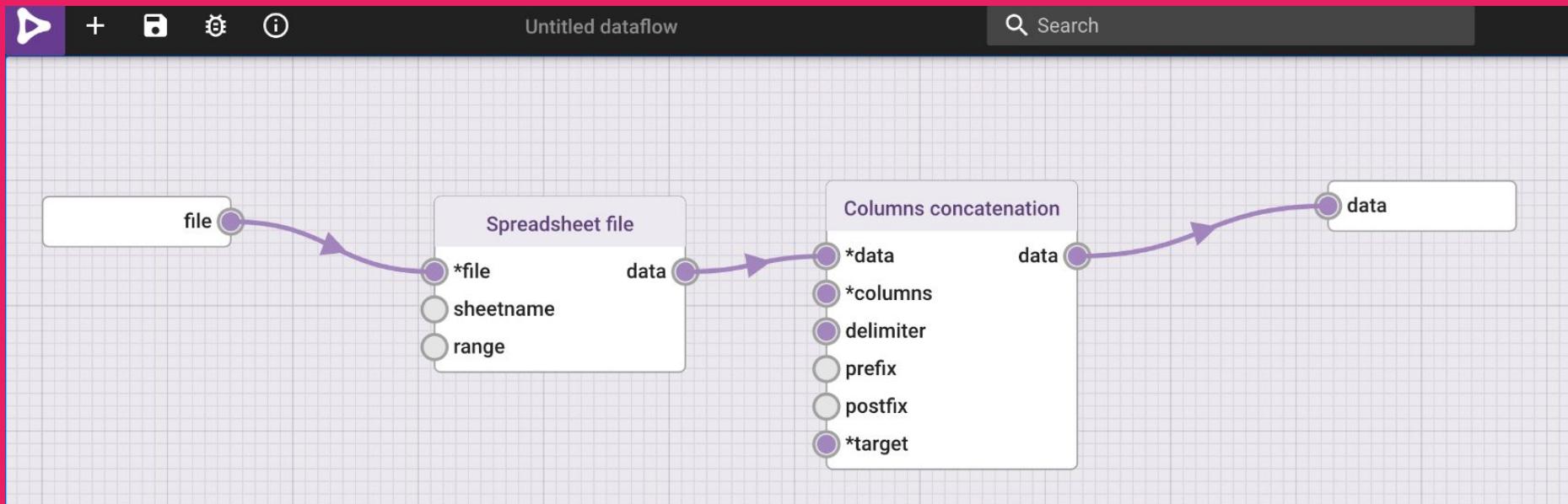
Isolate	Source	Streptomycin	Tn6235	Tetracycline	Tn6190
6072310	Clinic	Absent	Present	Present	
6086336	Clinic	Absent	Present	Present	
7001233	Clinic	NT	Absent	Absent	
7004578	Clinic	Absent	Present	Present	
7005405	Clinic	Present	Absent	Absent	
7021455	Clinic	Absent	Present	Present	
7044912	Clinic	Present	Absent	Absent	
7066827	Clinic	Absent	Present	Present	
7071508	Clinic	Absent	Absent	Absent	
7086074	Clinic	Absent	Absent	Absent	
7091952	Clinic	Absent	Present	Present	
8011061	Clinic	Absent	Absent	Absent	
8051280	Clinic	Absent	Absent	Absent	
8051728	Clinic	Absent	Present	Present	
8065344	Clinic	Present	Absent	Absent	
129280	Clinic	Absent	Present	Present	
H205	Farmer	Present	Present	Absent	
B173	Pig	Present	Present	Absent	
53737	Clinic	NT	Absent	NT	Present
47337	Clinic	Absent	Present	Absent	Present
H102	Farmer	Absent	Absent	Absent	
B113	Pig	Absent	Absent	Absent	
B173	Pig	Absent	Absent	Absent	
H121	Farmer	Absent	Absent	Absent	Present
B277	Pig	Absent	Absent	Absent	
H30	Farmer	Absent	Absent	Absent	Present
H189	Farmer	Absent	Absent	Absent	
B256	Pig	Absent	Absent	Absent	Present
H205	Farmer	Present	Present	Absent	
B151	Pig	Absent	Absent	Absent	

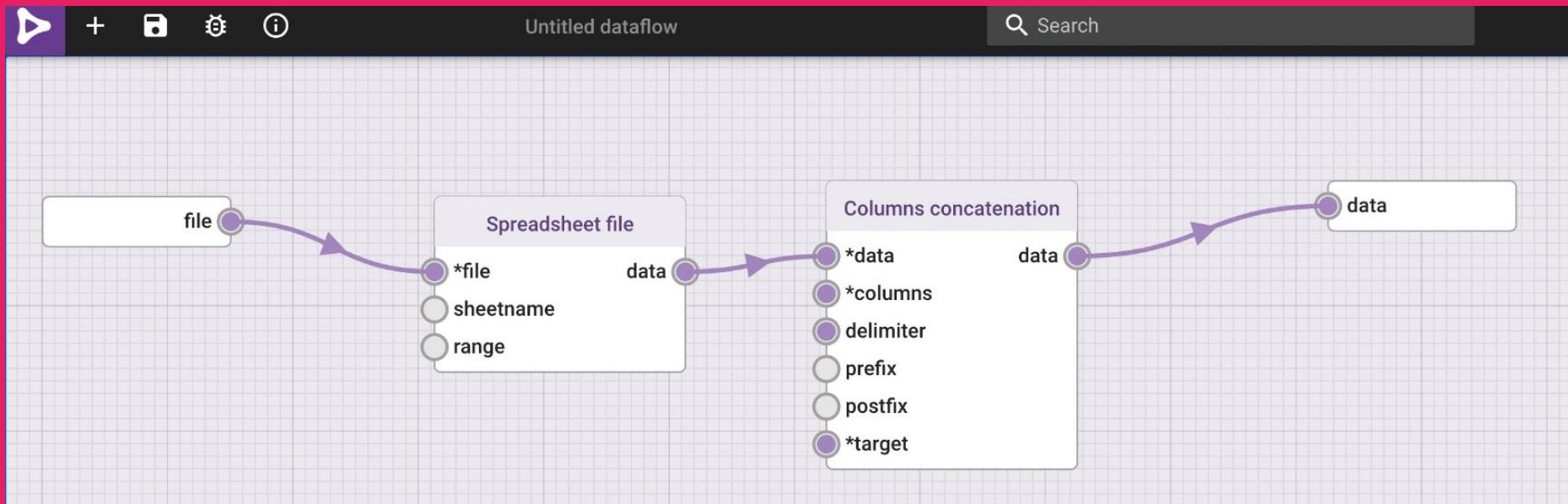




<http://data-flo.io>

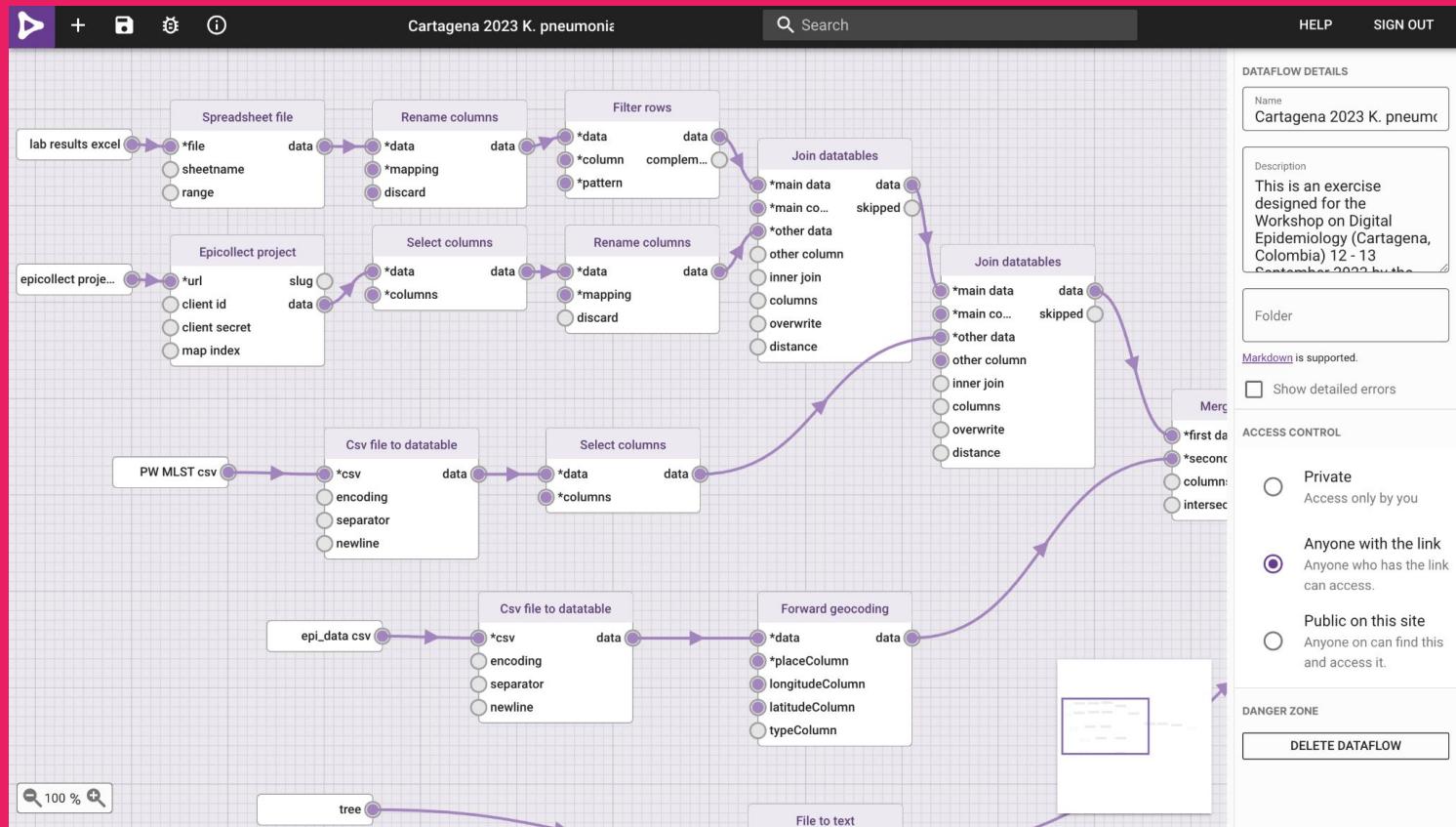
- Free
- Web based
- Many input types (and growing)
- Many transformations (and growing)
- Many output types (and growing)







welcome
connecting
science

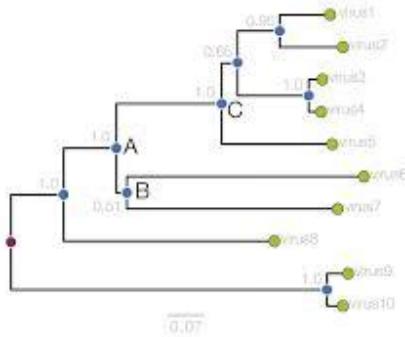


Genomic Epidemiology Challenges

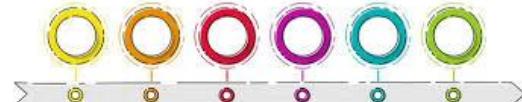
DATA EXPLORATION



map



phylogeny



timeline

ID	Accession	Organism	Host	Lab	Isolate No	ST	ST-S	ST-SST	ST-SST-S
PRY00001	MW3403_100	Escherichia	Bird		128	503	502	502	502
PRY00002	Hanford5	Escherichia	Bird		119	458	457	457	457
PRY00003	Hanford6	Escherichia	Bird		209	1000	999	999	999
PRY00004	Pitmeat	Escherichia	Bird		1958	6154	21142	21142	21142
PRY00005	Pitmeat	Escherichia	Bird		1957	6098	21140	21140	21140
PRY00006	Pitmeat	Escherichia	Bird		1959	6092	21141	21141	21141
PRY00007	Gaspero	Escherichia	Bird		1829	5002	6161	6161	6161
PRY00008	Gaspero	Escherichia	Bird		21019	186101	64550	64550	64550
PRY00009	Gaspero	Escherichia	Bird		18083	5003	6160	6160	6160
PRY00010	Gaspero	Escherichia	Bird		15628	134412	61600	61600	61600
PRY00011	Edgewood100	Escherichia	Bird		79	388	1487	1487	1487
PRY00012	Edgewood101	Escherichia	Bird		443	2093	12206	12206	12206
PRY00013	Hanford102	Escherichia	Bird		262	339	558	558	558
PRY00014	Hanford103	Escherichia	Bird		387	399	795	795	795
PRY00015	Hanford104	Escherichia	Bird		301	1048	795	795	795
PRY00016	Sherry30N	Escherichia	Bird		602	1057	1171	1171	1171
PRY00017	Sherry30S	Escherichia	Bird		123	753	454	454	454
PRY00018	Isaac212	Escherichia	Bird		202	4177	566	566	566
PRY00019	Edgewood21	Escherichia	Bird		213	594	987	987	987
PRY00020	Edgewood22	Escherichia	Bird		133	369	252	252	252
PRY00021	Theodore420	Escherichia	Bird		161	824	618	618	618
PRY00022	Theodore421	Escherichia	Bird		160	1242	617	617	617
PRY00023	Sabrina1247	Escherichia	Bird		174	388	582	582	582
PRY00024	Sabrina1248	Escherichia	Bird		368	1037	2601	2601	2601
PRY00025	Sabrina1249	Escherichia	Bird		173	373	584	584	584
PRY00026	Malene420	Escherichia	Bird		200	1051	761	761	761
PRY00027	Malene421	Escherichia	Bird		92	361	282	282	282
PRY00028	Malene422	Escherichia	Bird		153	608	547	547	547
PRY00029	Karen2111	Escherichia	Bird		189	357	887	887	887
PRY00030	Karen2112	Escherichia	Bird		180	2011	2640	2640	2640

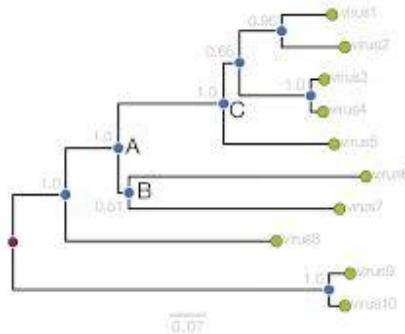
metadata

Genomic Epidemiology Challenges

DATA EXPLORATION



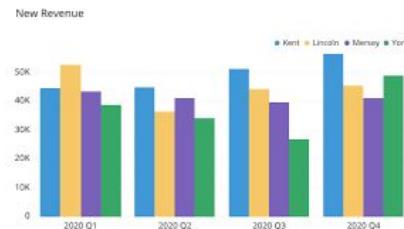
map



phylogeny



timeline

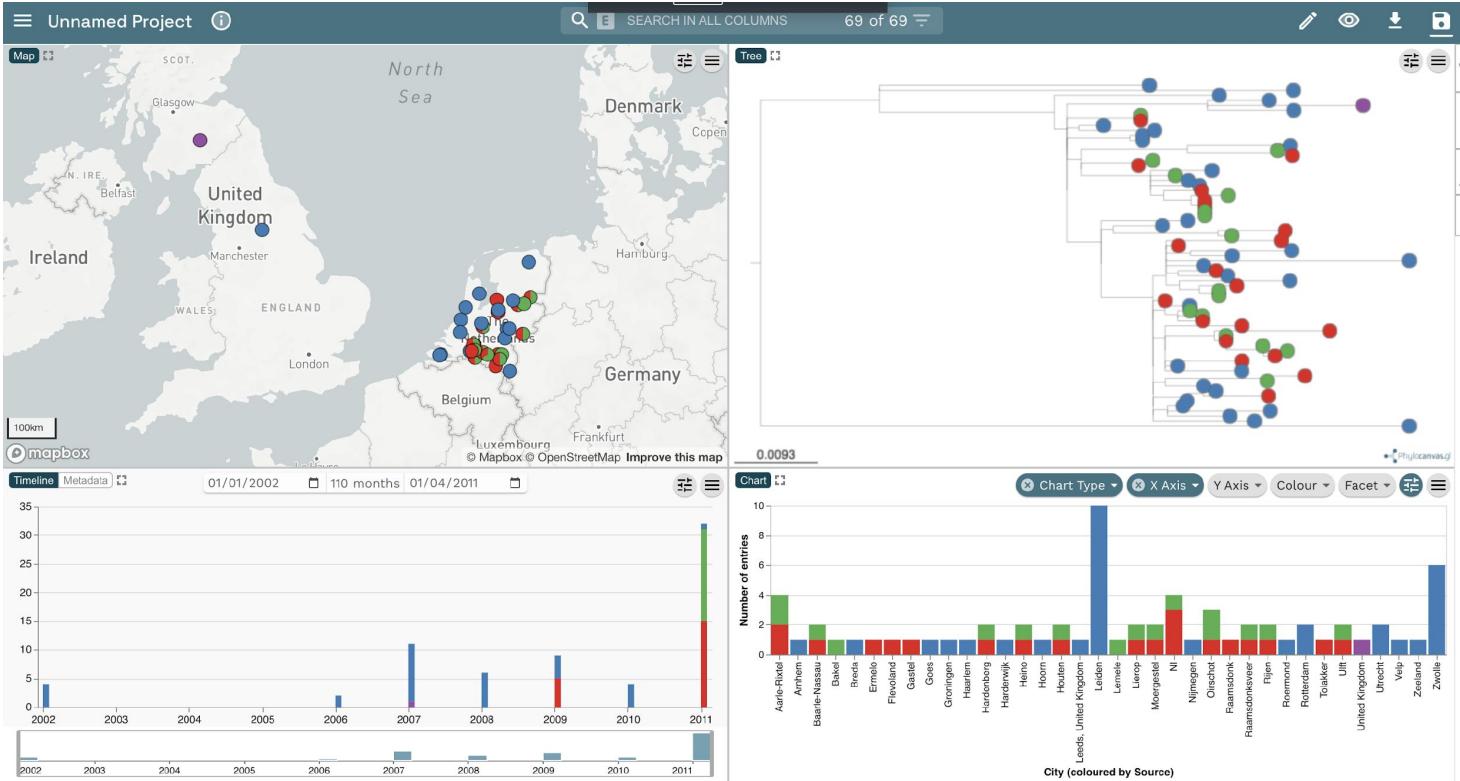


analysis

Genomic Epidemiology Challenges

DATA EXPLORATION

Microreact



Genomic Epidemiology Challenges

BIOINFORMATICS ACCESSIBILITY

RAW DATA
fastq's

```
UHRHLLGTTTCTGTTTCTTCTTGAACTTCCTTGAAAGCCTGATGCTTCAACGGAAAGATAACCTTGGCAAAGGA  
GCACCTCTACCGAGGTCTGTAATGATCTGCTTAACCATCCACGGATCTGCCCTCACGT  
TCGGAAGAGTCAAAGCTCAAGCTGGCAGCTCCCTTCAGACGAGTGTGAGCGGTG  
AAGATGGAACCGCACCCCTCTTGGCACGAATAACTCTACCATGTTGTGTTAATGT  
TCTCTGCTGAAGAGGACTTGAATAATTTTATTGGTTTTTTGGGAGTATGAGGG  
TTCTTGTTGTCGGGTTAACCTAGTGTGCTGGTCACGTGCCTTATTGGCAAGCTGTCG  
AGGTATCATAAAGGTGAGTTGAAGAGTACCTTATGGAAGACTTCGTTAGGAAGGTGTC  
GTATGATTAGTGGCGTAGGGTAATGATTAAATTCTCTTCG.....  
>NODE_1_length_1401924_cov_73.350607  
AAAGTCTCTCACGGAAACCCCTTCTGGTGGTCACGGAAGATAACCTTGGCAAAGGA  
GCACCTCTACCGAGGTCTGTAATGATCTGCTTAACCATCCACGGATCTGCCCTCACGT  
TCGGAAGAGTCAAAGCTCAAGCTGGCAGCTCCCTTCAGACGAGTGTGAGCGGTG  
AAGATGGAACCGCACCCCTCTTGGCACGAATAACTCTACCATGTTGTGTTAATGT  
TCTCTGCTGAAGAGGACTTGAATAATTTTATTGGTTTTTTGGGAGTATGAGGG  
TTCTTGTTGTCGGGTTAACCTAGTGTGCTGGTCACGTGCCTTATTGGCAAGCTGTCG  
AGGTATCATAAAGGTGAGTTGAAGAGTACCTTATGGAAGACTTCGTTAGGAAGGTGTC  
GTATGATTAGTGGCGTAGGGTAATGATTAAATTCTCTTCG.....  
>NODE_2_length_1392447_cov_73.757244  
TGTACCTACTAGCTGAATAACAAGTCTTGGGAGAAACTTGGTTTCAGAGACAAA  
GTTAAAGTACTTGACATTGGGAGCTAAGCTTCTGCTTCCTCTGTAAGAACGCTTCAG  
ACTTACCATTTGAAACAAGTGAGTTGCATTAGTATCAAAGGTTGGTATGATATAGGT
```

QC & assembly



ASSEMBLY
fasta

Genomic Epidemiology Challenges

BIOINFORMATICS ACCESSIBILITY

RAW DATA
fastq's

```
GAAGAAAAGAAGACACAGAGAGGGTTGGAAATTGTGAGAGAGAGAAAGAGACATCAAGAGGTTT  
+  
F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F  
Q@00178:71:H07705XX:1:1597:38291:3424 1:N:8:ACAGCAC-CTTGCTGT  
ACATAGAGCTTATGATTCCTCTGTGTTGAAAGGTCAGAAGGGGCCCTCTGGGACAAAGGAGCCTGACTCAAGG  
+  
F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F  
Q@00178:71:H07705XX:1:1597:38291:3424 2:N:8:ACAGCAC-CTTGCTGT  
CTGGATGAGAAACCTGAGAGATACCAAGAGAGATGATGCTTCTATAAGAGCTTACAATGACTGGAGAGCATCTGGCTGCAAG  
+  
F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F  
Q@00178:71:H07705XX:1:2413:2289:3579 1:N:8:ACAGCAC-CTTGCTGT  
GCTTGATGATTTGGCCCTCTCTGTGTCAGAAAGGGGCCCTCTGGGACAAAGGACAGCCTGACTCAAGCTGCCCTC  
+  
F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F  
Q@00178:71:H07705XX:1:2413:2289:3579 2:N:8:ACAGCAC-CTTGCTGT  
GAGAAAGAAAAGAACATCACAGAGGAGTTCTCAAGTGGATCTTACACAGAGAACCTACTGTGAGGAAGGAGCTGAG  
+  
F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F  
Q@00178:71:H07705XX:1:2354:5620:8876 1:N:8:ACAGCAC-CTTGCTGT  
ATGTTGTCGCTCTTCTGGTGTCAAGAGGAAAGGGGCCCTCTGGGAAACAAAGGAGAGCTGAGTCAGGTCAG  
+  
F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F  
Q@00178:71:H07705XX:1:2354:5620:8876 2:N:8:ACAGCAC-CTTGCTGT  
AGAAGGAAAGAAAAGAGAAAAGAACATCACAGAGTTCTCATGATGAGGCTCTGGATGAAAGCAAGAACCTACTGTGAG  
+  
F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F  
Q@00178:71:H07705XX:1:1568:4741:9815 1:N:8:ACAGCAC-CTTGCTGT  
CGAACATTACATGATGACATTTCATGOCAGAGAGCTGATTTTGCAGATGTTGATGTTAGATGAGCTTAAAGGTTGGTATAGGT  
+  
F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F::::::::::::::::::F
```

QC & assembly

ASSEMBLY
fasta

```
>NODE_1_length_1401924_cov_73.350607  
AAAGTCTCTCAGCAAACCCCTTGTGGGTACCGGAAGATAACCTGGCCAAGGA  
GCACCTCTACAGGGTGTGAATGATCTGCTTAACATATCCACGGATCTGCTTCACGT  
TCGGAAGAGTCAAAGCTTAAGCTTGGCAGCTCCCTTCAGACGAGTGTGAGCGGT  
AAGATGGAACCCAGCACCCCTTCTTGAGCACGAATAACTCTACCCATGTTGTAAATGT  
TCTTGCTGAAGAGGACTTGTGAAATTTTTATTGGTTTTTTGGGAGATGAGGG  
TCTTGTTGTCGGGTTAACCTAGTGTGTCAGTGCCTTATTGGGCAAGCTGTG  
AGGTATCATAAAGGTGGTAGTTGAAAGGTACCTTATGGAAGACTTCGTTAGGAAGGTGTCT  
GTATGATTAGTGTGGCGTAGGGTGAATGATTATTCTTCG.....  
>NODE_2_length_1392447_cov_73.757244  
TGACCTACTAGCTGAATAACAAGTTATCTGGGTTAGAGACAAA  
GTTAACCTACTGACATTGGGAGCTAAGCTGCTCTCTGAAAGACTTCAG  
ACTTACCATTTGAAACAAGTGAGTTGCATTAGTATCAAAGGTTGGTATGATAGGT
```

DOWNSTREAM
ANALYSIS

- Typing
- Contextualization
- AMR marker identification
- Storage



Pathogenwatch

<http://pathogen.watch>

- Fast predictions of resistant genotypes and relatedness from whole genome assemblies
- Contextualization with public data
- Real-time analytics and genomic epidemiology
- User-friendly web application

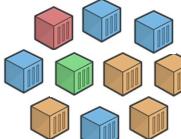
Sequences and Metadata
(Fastq assembled)



Speciation



Analytics
(containers)

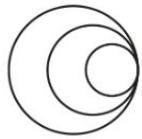


Species-specific analytics

Available Community-Developed Analytics



Species ID (Mash+RefSeq)	MLST	Additional Genotyping	Serotyping	cgMLST Network View	Core SNP trees Collection View	Curated AMR prediction	Plasmid Inc types (plasmidFinder)	Curated virulence prediction
<i>A. baumannii</i> <i>P. aeruginosa</i> <i>K. pneumoniae</i> complex <i>E. coli</i> <i>Shigella</i> spp. <i>E. cloacae</i> complex <i>E. faecium</i> <i>S. aureus</i> <i>C. coli/jejuni</i> <i>S. enterica</i> <i>S. Typhi</i> <i>N. gonorrhoeae</i> <i>S. pneumoniae</i> <i>M. tuberculosis</i> <i>V. cholerae</i> <i>L. monocytogenes</i> <i>S. equi</i> <i>R. salmoninarum</i> <i>C. auris</i> Zika <i>SARS-CoV-2</i>	Kleborate		Kaptive Kaptive		Kleborate		Kleborate	



wellcome
connecting
science

ACORN.

thank you

