Klebsiella 2024

Genomic Pathogen Surveillance - Tools & Interpretation

Cape Town, South Africa

9-11 September 2024

Course Material

Centre for Genomic
Pathogen Surveillance

Index

Centre for **Genomic**
Pathogen Surveillance

Programme

Day 1   Monday 9th September

08:30 – 09:00	Registration
09:00 – 10.00	**Welcome & introduction**
10:00 – 11:00	**Speed dating introductions**
11:00 – 11:15	**Break**
11:15 – 12:00	**Recorded talk: Practical aspects of Pathogen Genomics** Nabil-Fareed Alikhan
12:00 – 12:30	**Introduction to Pathogenwatch** Sophia David
12.30-13.00	**Pathogenwatch exercise (1)** Sophia David and Julio Diaz Caballero
13:00 - 14:00	**Lunch**
14:00 - 15:15	**Pathogenwatch exercise (1)** Sophia David and Julio Diaz Caballero
15:15 -15:30	**Break**
15:30 - 16:30	**Talk: Genomic surveillance of carbapenem-resistant *Klebsiella pneumoniae* in Europe** Sophia David

Centre for Genomic
Pathogen Surveillance

Day 2   Tuesday 10th September

09:00 – 10:00	**Pathogenwatch wrap-up session** Julio Diaz Caballero and Sophia David
10:00 - 11:00	**Introduction to data visualisation in Microreact** Nicole Dagata
11:00 – 11:15	**Break**
11:15 – 13:00	**Microreact exercise (2)** Sophia David, Julio Diaz Caballero and Nicole Dagata
13:00 - 14:00	**Lunch**
14:00 - 14:30	**Microreact wrap-up session** Julio Diaz Caballero and Sophia David
14.30 - 15.15	**Recorded talk: Using genomics to understand *Klebsiella pneumoniae* transmission between clinical and non-clinical settings** Natacha Couto
15:15 -15:30	**Break**
15:30 - 16:30	**Data sharing workflow mapping** Nicole Dagata

Centre for Genomic
Pathogen Surveillance

Day 3   Wednesday 11th September

09:00 – 10:00	**Recorded talk: Real-world applications of genomic pathogen epidemiology** David Aanensen
10.00 - 11.00	**Workshop - Investigating a hospital outbreak of MDR *K. pneumoniae* (3)** Julio Diaz-Caballero and Sophia David
11:00 – 11:15	**Break**
11:15 – 13:00	**Workshop - Investigating a hospital outbreak of MDR *K. pneumoniae* (3)** Julio Diaz-Caballero and Sophia David
13:00 - 14:00	**Lunch**
14.00 - 14.30	**Workshop - Group presentations** Julio Diaz-Caballero and Sophia David
14:30 - 15:15	**Talk: Genomic surveillance of AMR using amr.watch** Julio Diaz-Caballero
15:15 – 15:30	**Break**
15:30 - 16:30	**Data sharing workflow mapping** Nicole Dagata

Centre for **Genomic Pathogen Surveillance**

CGPS Team

Trainer contacts

Name	Contact information
David Aanensen Director of CGPS	david.aanensen@cgps.group
Sophia David Senior Research Fellow	sophia.david@cgps.group
Julio Diaz Caballero Lead Bioinformatician	julio.diaz@cgps.group
Nicole Dagata Surveillance Program Manager	nicole.dagata@cgps.group

Training handbook developers

Current team members:

Nabil-Fareed Alikhan
Krisandra Allen
Natacha Couto
Nicole Dagata
Sophia David
Julio Diaz Caballero
Emma Kumaran
Georgina Lewis-Woodhouse

Previous team members:

Monica Abrudan
Silvia Argimon

Software developers

Khalil Abu-Dahab
Corin Yeats
James Baker
Pranit Shinde

Centre for Genomic
Pathogen Surveillance

Workshop overview

Genomic epidemiology aims to understand the evolution and spread of pathogens with the goal of implementing evidence-based interventions to protect public health. The increased genetic resolution afforded by genomic data has proven useful at different geographic scales, including for local outbreak investigations and for examining patterns on a broader global scale. It has been revolutionary in understanding the population structure, diversity and evolutionary dynamics of pathogens, especially those that exhibit little genetic variation (e.g. *Salmonella* Typhi ([Wong et al. 2015](#))). In the era of increasing antimicrobial resistance, genomic epidemiology is also a vital tool for identifying the routes and/or reservoirs via which resistant pathogens are spreading, investigating important strains (e.g. high-risk clones) and tracking the underlying resistance mechanisms and plasmid vectors.

During this workshop, we will explore pathogen genome analytics and visualisation using free web-based applications developed by the Centre for Pathogen Surveillance (CGPS) including 1) Pathogenwatch ([pathogen.watch](#)) and 2) Microreact ([https://microreact.org](#)). We will use data from the important bacterial pathogen, *Klebsiella pneumoniae*, and explore analytics appropriate for this species although many of the tools and concepts are broadly applicable across different bacterial species.

K. pneumoniae is a leading cause of healthcare-acquired infections globally and causes a range of disease types including pneumonia, skin and wound infections, urinary tract infections and sepsis. There has been an increasing prevalence of infections with multidrug-resistant (MDR) *K. pneumoniae* over the last two decades. A major concern now is the rise of infections that are resistant to carbapenems, a vital class of antimicrobials for treating severe infections in hospitalised patients ([WHO, 2024](#)). There have also been increasing reports of *K. pneumoniae* infections that are both hypervirulent and MDR, presenting an even heightened threat ([ECDC, 2024](#)).

This three-day workshop will include talks describing pathogen genomic workflows, genomic epidemiological studies of *K. pneumoniae* and different public health applications, combined with hands-on exercises. For the latter, we will first explore the use of Pathogenwatch for genomic typing (e.g. MLST, cgMLST) of *K. pneumoniae* genomes, clustering and phylogenetic analyses, identification of resistance and virulence markers, plasmid replicon typing and determination of K- and O-types. We will then use Microreact to link epidemiological and genomic data in interactive data visualisations with a focus on interpretation and data sharing for public health value. In a final exercise, we simulate a hospital outbreak scenario and show how the Pathogenwatch and Microreact tools can be used effectively in a genomic epidemiological investigation.

Please feel free to ask questions to the facilitators at any time during the workshop. Our training materials are also in continuous development and we will be grateful for any feedback during the course or once you are back home.
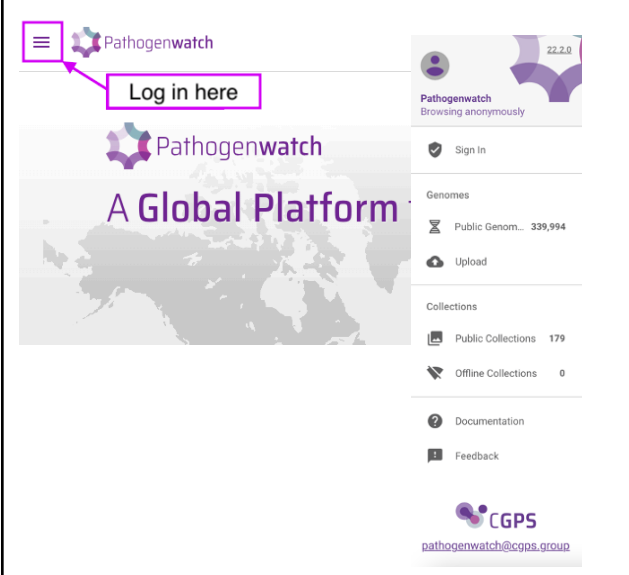
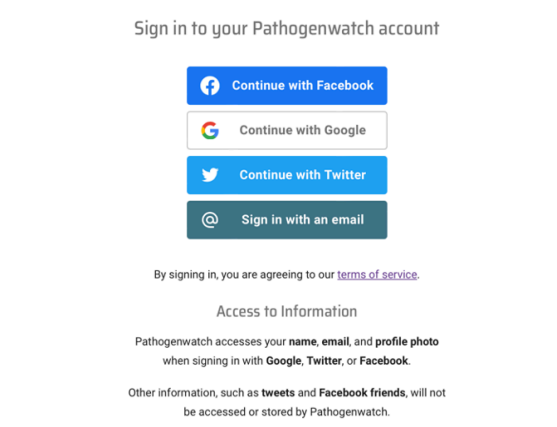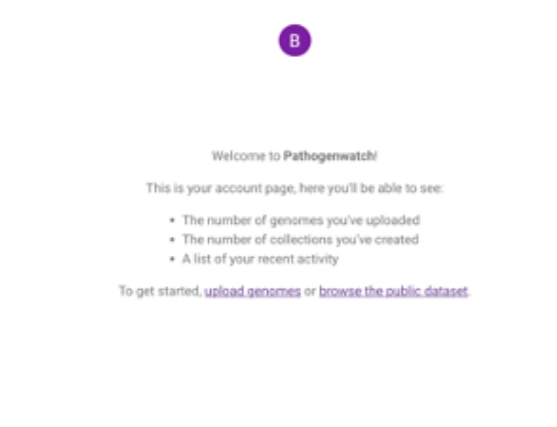# Exercise 1: Pathogen genome analysis with Pathogenwatch

During the exercise we will introduce **Pathogenwatch**, a web-based tool for genomic surveillance that enables users to upload and analyse their own genome data. Pathogenwatch provides species and taxonomy prediction for over 60,000 variants of bacteria, viruses, and fungi. The tool then performs a series of species-specific analytics including MLST, identification of resistance and virulence loci, replicon typing and genome clustering. MLST is available for over 100 species using schemes from PubMLST, Pasteur, and Enterobase. Users can create phylogenetic trees of bespoke genome collections incorporating their own uploaded data and/or publicly-available genomes for supported pathogens which can be analysed with metadata in interactive visualisations. Data can also be exported for downstream purposes such as for use in customisable dashboards in Microreact.

The data for this exercise is available here:
https://drive.google.com/drive/folders/1yZsX5HJFq1_wsY6qfY9rc2N4k4NXQLP2
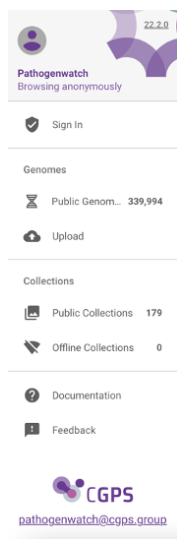
Creating an account

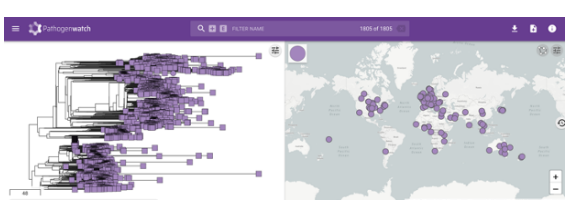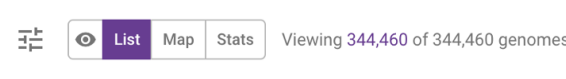| Logging into Pathogenwatch	
Open Pathogenwatch on **Google Chrome** or **Mozilla Firefox**. For this exercise, we will use the PW-demo site available at the URL: **https://demo.pathogen.watch/** This is a "light" version of Pathogenwatch used for training, but it looks and feels like the main Pathogenwatch site. Click on "Sign In" by clicking on the ☰ icon in the upper left corner of the page. Note - Never upload confidential metadata to the PW-demo site. Only https://pathogen.watch is appropriate for viewing such metadata. It is also worth looking at private metadata browsing if it is not possible to upload your metadata at all, e.g. due to personally identifying information (PII) or patient confidentiality. For more details, see: https://cgps.gitbook.io/pathogenwatch/how-to-use-pathogenwatch/private-metadata	

Sign in using your credentials
 	
Select your preferred sign in method (the options are restricted to Google/email on the PW-demo site).
 	
Input your login credentials.
 	
    ● You will create your own login credentials (or use the credentials associated with your sign-in method).
    ● Check your SPAM folder if you choose to sign in by email and have difficulties receiving the login link.
	Sign in to your Pathogenwatch account
	
	Continue with Facebook
	Continue with Google
	Continue with Twitter
	Sign in with an email
	
	By signing in, you are agreeing to our terms of service.
	
	Access to Information
	
	Pathogenwatch accesses your **name, email**, and **profile photo** when signing in with **Google, Twitter**, or **Facebook**.
	
	Other information, such as **tweets** and **Facebook friends**, will not be accessed or stored by Pathogenwatch.
	
	Public login page

View My Account page
 	
Now that you are logged in, you will see your Account page where you can manage your genomes, collections and account settings.
 	
Note that any data in your PW-demo account will be removed after this training session. Use the main site (https://pathogen.watch) for uploading your own data outside of this workshop.
	B
	
	Welcome to **Pathogenwatch!**
	
	This is your account page, here you'll be able to see:
	
	  • The number of genomes you've uploaded
	  • The number of collections you've created
	  • A list of your recent activity
	
	To get started, upload genomes or browse the public dataset.
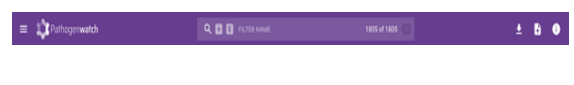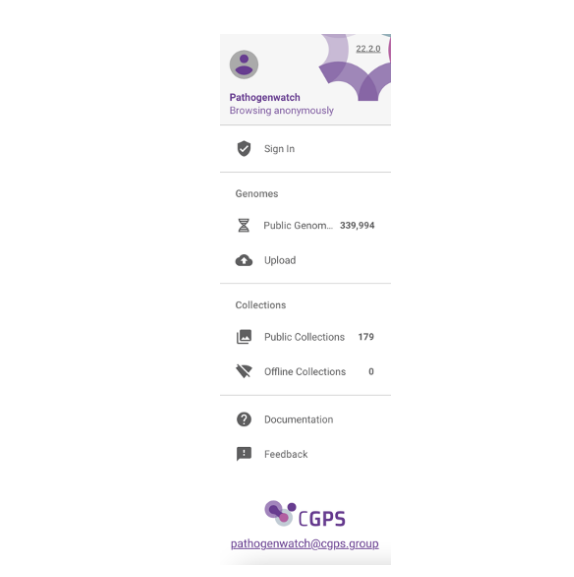
Navigating Pathogenwatch

Navigating the site	
Click the menu icon The main menu bar can be accessed by clicking on the ≡ icon on the upper left corner of the page.	≡ Pathogenwatch Pathogenwatch A **Global Platform**

Centre for Genomic
Pathogen Surveillance
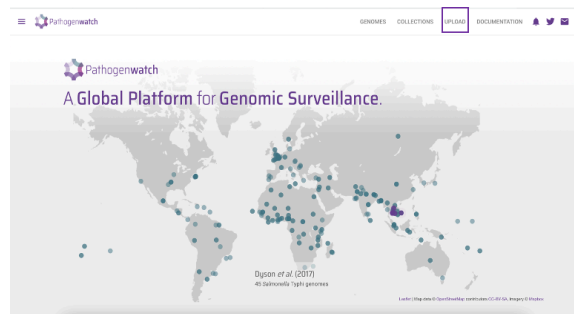
View the four main pages This menu allows you to navigate the site. There are four key pages: ● Genomes ● Collections ● Upload ● Documentation You may also use this menu to sign out of your account by clicking "Sign Out". You may also submit feedback to the developers about Pathogenwatch by selecting "Feedback". 💡 Tip - These four menu options are also easily accessible on the upper right corner of the page. ⓘ For more information about navigating Pathogenwatch visit: https://cgps.gitbook.io/pathogenwatch	 OR

Pathogenwatch terminology

Key terminology		
Name	**Description**	**Image**
View	Each individual dashboard element in a Collection View is called a **View**. The dashboard in the image has two views - a Tree View and a Map View.	
Toolbox	Each panel in the Genome Page.	
Search-and-filter bar	Typing in the top search-and-filter bar in the Collection View will filter the data across all columns in the metadata. Clicking the "+" will let you filter data based on a	

Centre for Genomic Pathogen Surveillance

	specific column. Clicking the "E" will enforce an exact match requirement.	
Top menu bar	Always visible along the top in the Collection View.	
Left menu bar	Accessed through the hamburger icon at the left of the top menu bar.	

Importing data

Upload files
Navigate to the "Upload" page First make sure that you are logged into the PW-demo site (https://demo.pathogen.watch). On the home screen, click on "Upload" in the upper right corner. You may also use the "Upload" button located in the main menu (see Navigating the Site section of this tutorial).

Select the file type to upload

Clicking the "Upload" button brings you to the screen shown here.

For this exercise, click on "Single Genome FASTAs".

ⓘ Single Genome FASTAs represent one or more FASTA files with one genome per FASTA file.
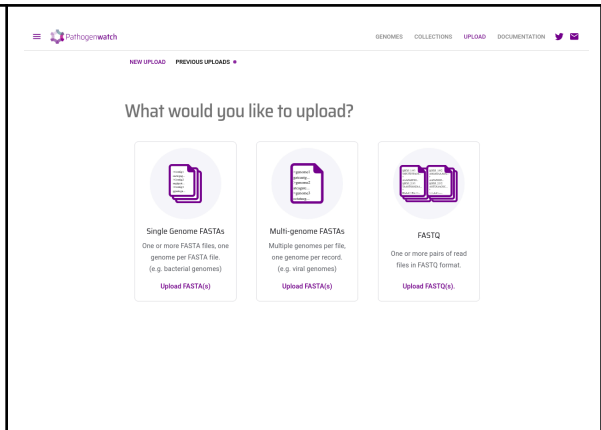
Browse your files

On the bottom right corner, use the ⊕ icon to select files to upload. Alternatively, you can drag and drop files directly into the centre circle.

In this exercise we will first create a project with the following files:
Test1.fasta
Test2.fasta
Test3.fasta
metadata.csv

These files can be downloaded from:
https://drive.google.com/drive/folders/1yZsX 5HJFq1_wsY6qfY9rc2N4k4NXQLP2

💡 Tip - If your internet connection regularly drops, using the "Compress files" option will increase the chance that each file will upload successfully.

💡 Tip - Metadata has to be in CSV format, with the extension `.csv`.
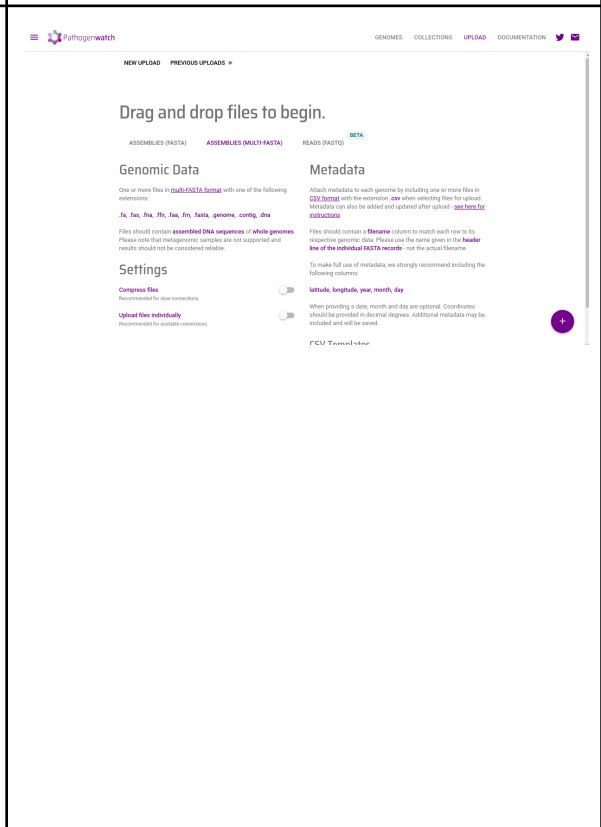
Processing screen and viewing results

The tasks being carried out and their individual progress are tracked on the left hand side. The overall progress and current stage is tracked on the top right and indicated when complete.

As results arrive from Speciator and then MLST, the species and ST are displayed for each submitted assembly in the animated circle.

Once all tasks are complete, you can click the "View Genomes" button to view the results in your "Genomes" page.

Centre for Genomic Pathogen Surveillance
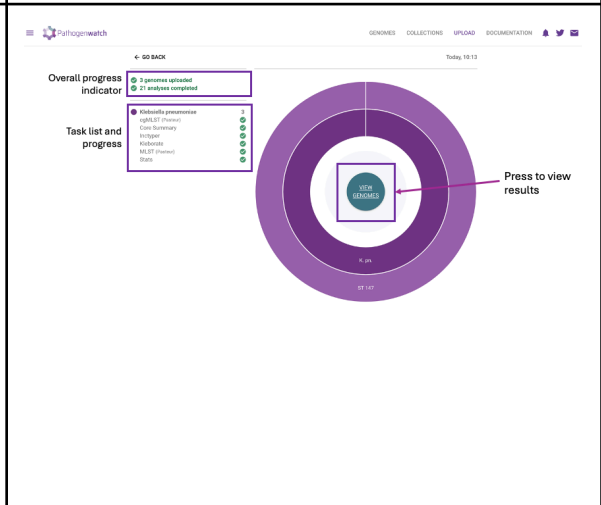
Viewing the Genomes page
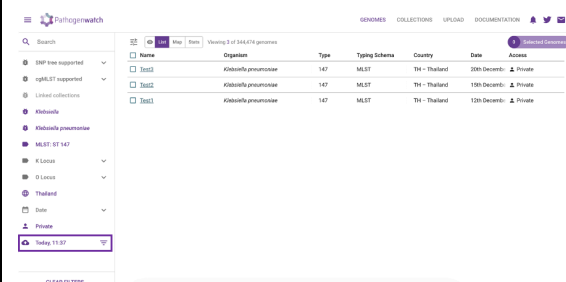
Browsing genomes

View your uploaded genomes

The uploaded assemblies are presented automatically, and tagged as a single group in the bottom left, with tag "Today" and the time of upload.

The Genomes page also includes publicly-available genomes which can be accessed using various filters on the left hand side.

Note that the PW-demo site contains only a small subset of genomes that are available on the main Pathogenwatch site.

List of uploaded genomes.
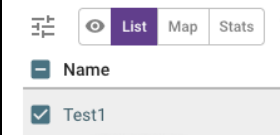
Recently uploaded assemblies are presented as a single group on the bottom left of the page.
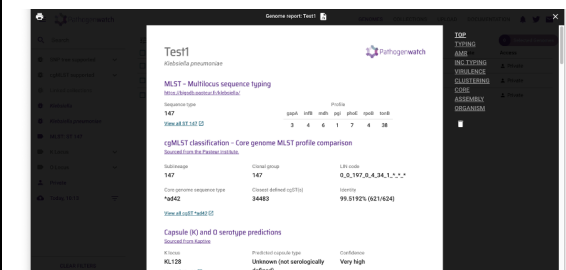
Genome Reports

Click on one of the uploaded assembly names. This will show a detailed report of the assembled genome.

The Genome Report includes results from each tool run by Pathogenwatch (e.g. MLST), along with assembly metrics (e.g. N50) and a core genome clustering analysis (if a cgMLST scheme is available). The version of each tool is also tagged and a link to the original FASTA is provided at the top.

Assembly name of a recently-uploaded file.

Genome report.

Genomic context search

For species where there is a cgMLST scheme available, Pathogenwatch provides a single-linkage clustering-based service for finding other genomes within a specified threshold.

The first time a genome is uploaded, it is necessary to first run the clustering by clicking on the "RUN CLUSTERING" button in the Genome Report.

💡 Tip - The threshold can be modified using the slider under the chart showing the number

Centre for Genomic Pathogen Surveillance

13
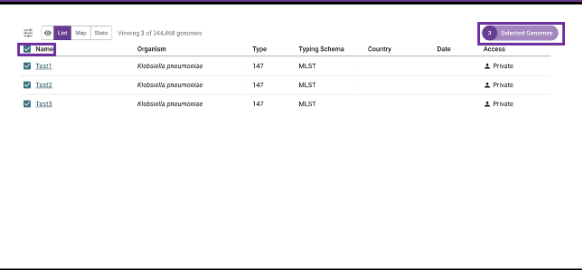
of genomes at each value.	

Creating a Collection

Filtering and selecting genomes	
Selecting your genomes to build a collection Ensure that only your three uploaded genomes are listed in the Genomes page. Then click the tick that selects all genomes. You will see that the number of "Selected Genomes" is now "3" in the top right.	
Adding more genomes to your selection Click on "CLEAR FILTERS" at the bottom left corner. You can now filter all genomes within Pathogenwatch from the left sidebar. Filter the genomes by: ● Genus - Klebsiella ● Species - Klebsiella pneumoniae ● MLST - Pasteur - ST 147 ● Date - Jan 2020 - Dec 2022 ● Click on the tick that selects all genomes and then click on "Selected Genomes". You can see that the number of genomes has now increased. Tip - You can type what you are looking for in each filtering step.	
Creating a collection Select the assemblies for which you want to create a collection by ticking the green checkboxes. Once the required assemblies are selected, click the "Selected Genomes" button and then "Create Collection".	
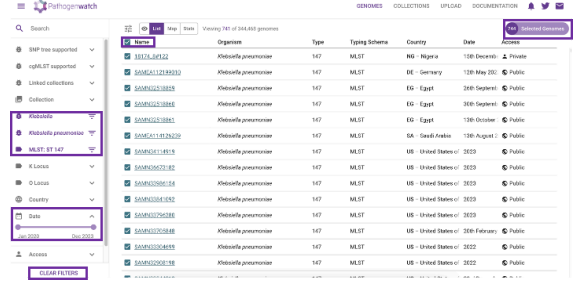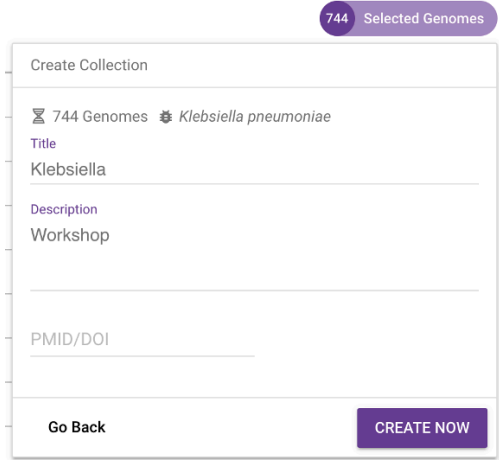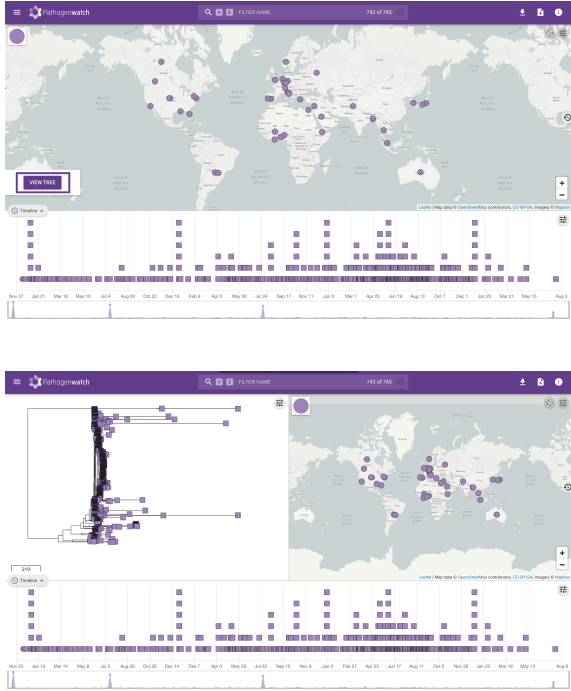
Centre for Genomic
Pathogen Surveillance

Naming a collection

At this point, you are presented with the option of filling in a description of the collection. The fields include:

Title - This will be included in the collection URL.
Description - A short description of the collection, perhaps to explain to other people what it is for.
PMID/DOI - (optional) The PubMed or DOI identifier if it is linked to a publication.

Finally, click "Create Now" and you will be taken to the "Collection View".

Creating a SNP-based tree

For a subset of species, listed under the "SNP tree supported" filter, a tree will be calculated using a neighbour-joining method. If the tree of the collection has not been previously calculated, it may take a few seconds to a few minutes to create, depending on the server load and collection size. Once it is ready, you can click on "VIEW TREE".

Centre for Genomic
Pathogen Surveillance

15

Viewing a Collection

Browse an existing collection

Explore a *K. pneumoniae* collection comprised of data from a Global Health Research Unit (GHRU) AMR project by either typing the link: **https://tinyurl.com/2ctwwxc5** Or by scanning the QR code below: Make sure that you are using one of the PW supported browsers: Mozilla Firefox or Google Chrome. Locate the five interactive components: 1. A phylogenetic tree. 2. A global map showing the locations of genomes that have geospatial metadata. 3. A metadata display area that includes the user-supplied data, calculated typing assignments such as MLST, and antimicrobial resistance predictions. 4. A query bar that allows the selection or finding of assemblies by searching the metadata fields. 5. A timeline showing the collection dates of assemblies that have temporal metadata. The header row also contains links to the home page and a "Downloads" button in the top right.	
Map View The map view plots the locations of all assemblies that have been provided with latitude and longitude coordinates. Selecting a circle will highlight the corresponding taxa in the "Tree View" and in the "Metadata Table".	

Centre for Genomic
Pathogen Surveillance

Assemblies that have been selected in the "Tree View" or "Metadata Table" will be highlighted with a purple ring.	
When a column in the "AMR Determinants" table is selected, circles on the map will be coloured in a ratio of resistance/intermediate/sensitive phenotype (red/yellow/white) or present/absent (red/white) genotype. The circle in the top left shows the overall ratio for the selected resistance phenotype or genotype.	
The group selection tool allows you to draw a line around a set of points to select. These are then highlighted in the "Tree View" and "Metadata Table".

Tip - If you click on the ⚎ icon within the Map View, you can adjust the marker sizes. You can also zoom in or out (swipe gestures work if using a touchpad), move the map around, and download an image by right-clicking on the map view. |

The group selection tool, shown in purple, in the top right corner of the map

 |
| **Tree View**

The displayed trees are fully interactive and allow drawing of subtrees, selection of assemblies by clicking on parent nodes, and highlighting of the phylogenetic distribution of AMR and other metadata.

Assemblies can be selected by either clicking on specific leaves or by clicking internal nodes, and are highlighted with a purple halo. Selected assemblies are highlighted in the "Map View" also with a purple halo and shown exclusively in the "Metadata Tables".

Tip - If you click on the ⚎ icon, you can change the tree topology, zoom in or out (swipe gestures work if using a touchpad), adjust nodes & labels, move the tree around, and download an image by right-clicking on the tree view. |

Click on the ⚎ icon to change the tree topology. |

Centre for Genomic
Pathogen Surveillance

Tip - Your uploaded genomes are shown as circles in the tree while public genomes from Pathogenwatch are shown as squares.	
Metadata Tables The "Metadata Tables" display uploaded and calculated attributes for each assembly, grouped according to the type of information. Assemblies are sorted according to the order of the leaves on the rectangular tree, and can be selected by clicking on rows, or by clicking on column headers and then typing in the "Filter Bar" at the top of the page. The <u>Metadata</u> tab shows all the data fields connected with an assembly.	
Clicking on the assembly name in any tab will bring up the Genome Report for that assembly. All tabs allow you to download the FASTA and GFF files for each assembly. Clicking on the top of the column will download all selected files in the current collection as a single zip archive.	 *Clicking on the ".fa" or the ".gff" icons will download the respective files for that respective assembly.* *An example of a genome report.*
The <u>Typing</u> tab shows MLST results and any additional species-specific results.	
The <u>Stats</u> tab shows you a summary set of QC statistics calculated by Pathogenwatch.	

Centre for Genomic
Pathogen Surveillance

The <u>Antibiotics</u> tab shows you a predicted susceptibility profile based on AMR genotypes.	
The <u>AMR Determinants</u> tab shows you the presence and absence of specific AMR genes associated with specific drug classes (available for some species).	
Timeline The "Timeline" plots the collection dates of all assemblies that have been provided with year, month and day information. You can click on an assembly to highlight it in the "Tree View" and "Map View". Tip - If you click on the icon, you can play the timeline, change nodes size, choose how to group assemblies, choose speed and reset zoom.	
Filter Bar The "Filter Bar" allows you to select assemblies according to their attributes in the "Metadata Tables", including by typing assignments or antimicrobial resistance. You can filter on multiple columns by clicking on the '+' symbol. Click a field you are interested in, and you will be taken to a selection of possible terms, with the number of associated assemblies next to each term. Clicking a term will add it to a filter. You can then choose to add another term that the assemblies must match *as well* by clicking 'AND'. The new term must come from a different field. You can also add an 'OR' statement. This allows you to create another "sub"-filter that assemblies can *also* match. You may have multiple 'OR' and 'AND' terms to build up complex queries. First filter by COUNTRY and select India. Then select NDM-1 from AMR Determinants. Now,	

you can look at all the *K. pneumoniae* isolates with NDM-1 from India. 💡 Tip - Remember clicking on the "E' in the filter bar will enforce an exact match.	

Downloading data

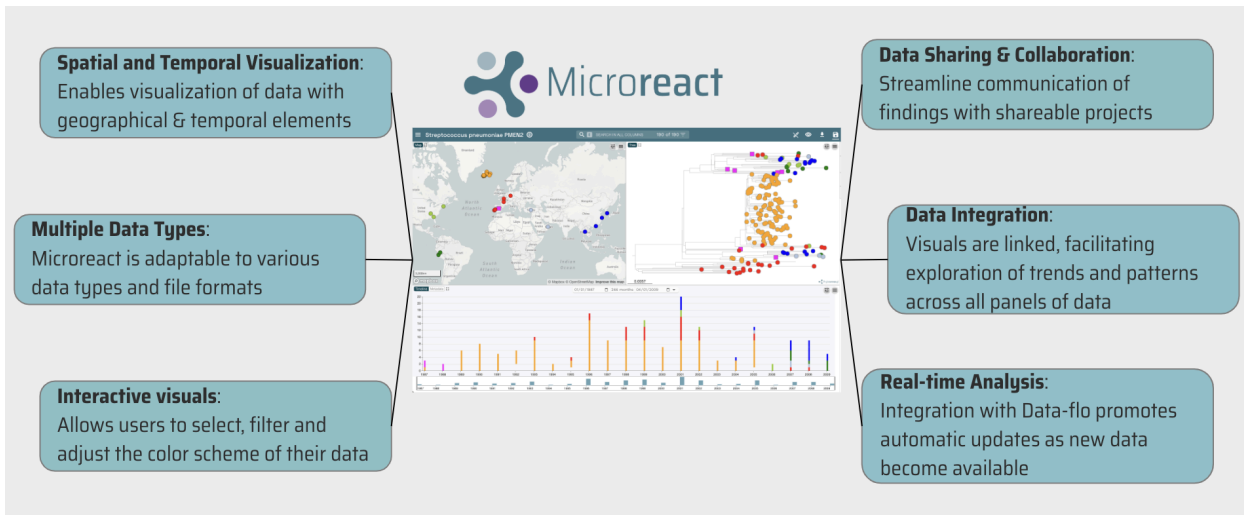|	
Downloading data You can download data in the Collection View by clicking on the ⬇ icon in the upper right corner and then clicking on the data you want to download. Let's download the Tree (.nwk), the Metadata table and the Typing table. The input fasta files can be downloaded from the Metadata window, by clicking on the ".fa" icon next to the genomes' names.	

Exercise 2: Visualisation and interpretation of pathogen genomes using Microreact

Microreact (https://microreact.org/) is a public, open-source, web-based application that promotes data review and analysis through the rapid generation and linkage of interactive data visualisations.

Microreact users build projects either by uploading data files or by providing links to data files via URLs. Users can then build panels to showcase the data in a customisable way. A panel within Microreact is a pre-configured visual (Tree, Map, Table, Chart, Network, Matrix) each with its own set of customisable features (colour, layout, labels, etc).
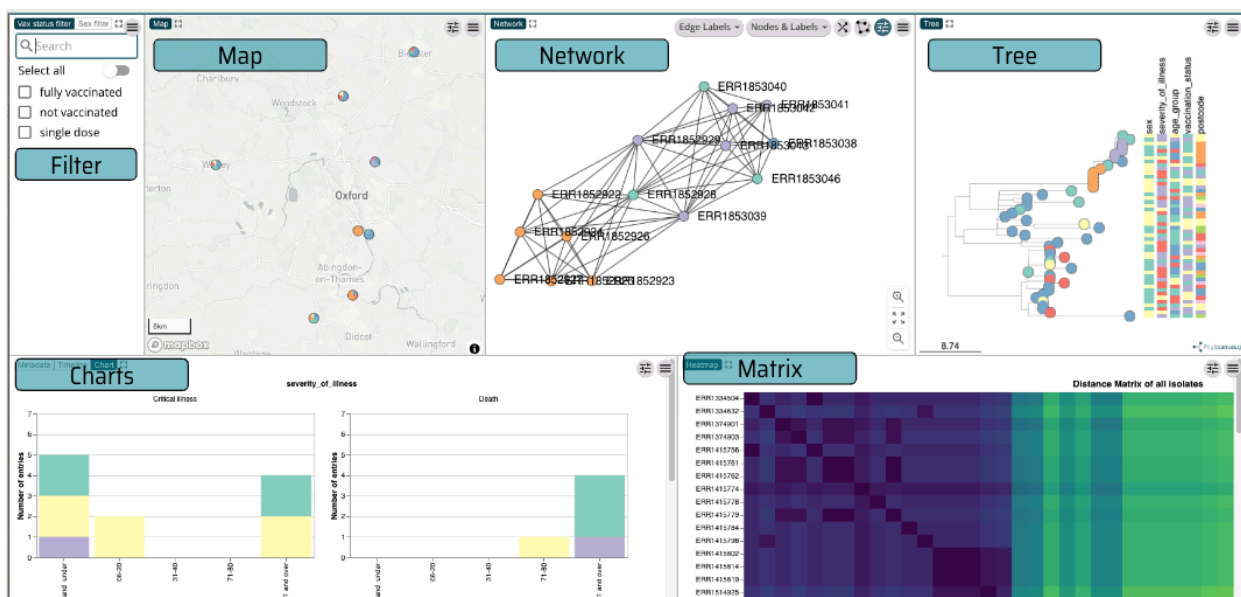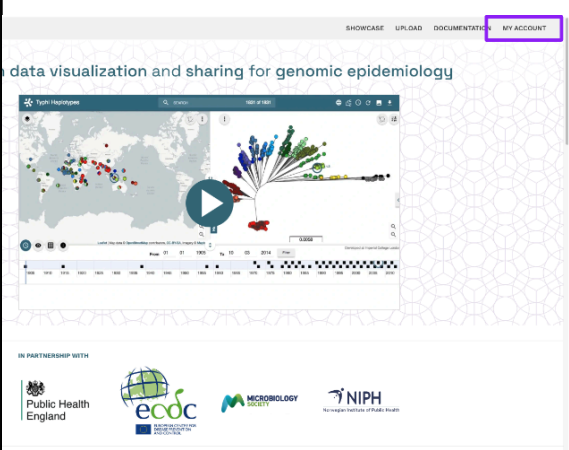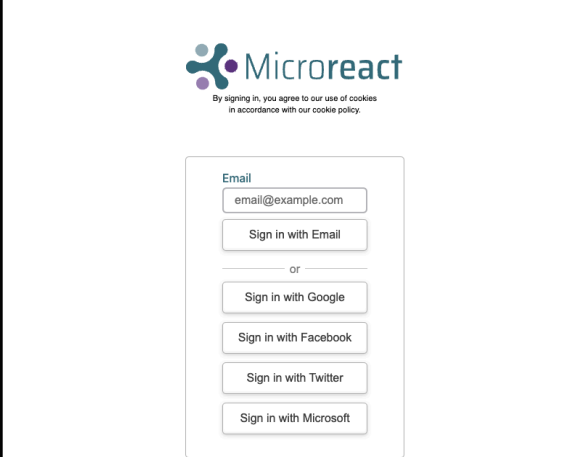
Panels can be filtered and snapshots of data can be saved and shared. These snapshots enable users to set up a particular filter or explore a selection of samples or group of visuals to refer back to with the goal of facilitating understanding of complex public health data.

Microreact can be deployed securely on a local server, adhering to local data governance. Users can make their Microreact projects findable on local networks, or share them privately with other users. Microreact projects for public consumption can be shared on the public Microreact site.

The data for this exercise can be found here:
https://drive.google.com/drive/folders/1nqQ10s-GX7Nsu9-HN7w82Wph_wDBtxSG

Creating an account

Logging into your Microreact account	
Open Microreact ● Use https://microreact.org/ to access the public site. Click on "My Account" in the upper right corner of the page. Tip - Make sure you don't put protected health information into the public site. Always use a local installation for sensitive data.	
Sign in using your credentials Select your preferred sign in method. Input your login credentials. ● If using the public site, you will create your own login credentials (or use the credentials associated with your sign-in method i.e. Google or Twitter) ● If using a locally-installed version, use the login credentials assigned by your IT department	 *Public login page*

View My Account page

Now that you are logged in, you will see your Account page where you can manage your projects and account settings.

Navigating Microreact

Navigating the site

Click the menu icon

The menu bar can be accessed by clicking on the ☰ icon on the upper left corner of the page.

Clicking the icon brings up the main menu.
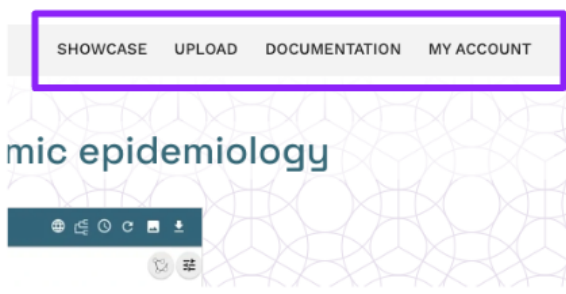
View the four main pages

This menu allows you to navigate the site. The site has four main pages:

- Showcase
- Upload
- Documentation
- My Account

You may also use this menu to sign out of your account by clicking "Sign Out". You may also submit feedback to the developers about Microreact by selecting "Send Feedback".

💡 Tip - These four menu options are also
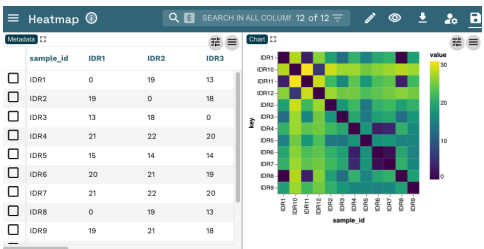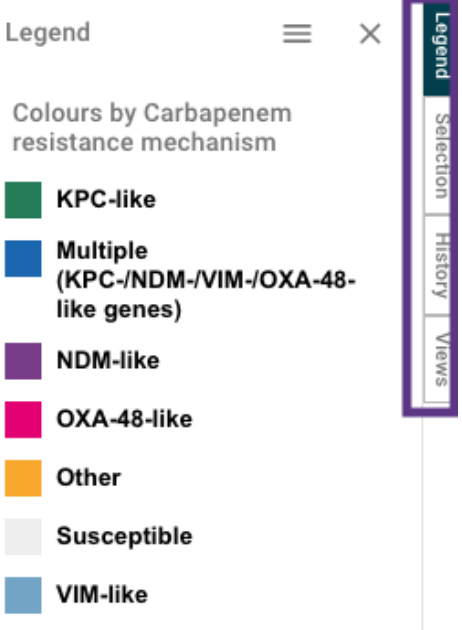
OR

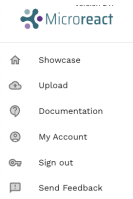Centre for Genomic Pathogen Surveillance

easily accessible on the upper right corner of the page.

(i) For more information about navigating Microreact visit: https://docs.microreact.org/instructions/navigating-the-site

Microreact terminology

Microreact icon summary	
Icon	**Description**
☰	Display additional menu options. It can be found either within individual panels, or at the top left of your microreact account.
✏	Opens a menu to let you edit or delete existing panels, as well as add a new panel.
👁	In this menu you can control selections for labels, colours, and shapes.
⬇	Download individual project files, such as the metadata file, .newick files, etc.
👤⚙	Update who has access to the project as well as find sharing links and create an easier alias link.
💾	Allows you to save a new project, update a current project, or download a file containing the complete project.
⚙	This is the configure icon. It expands or collapses additional buttons or menus.
⛶ ⛶	The maximise icon expands one panel to cover the whole dashboard; the minimise icon returns the panel to standard size.
🔗	This is the lasso tool (or polygon lasso). It allows you to draw a filter around points of interest on a network or tree to create a custom filter.

Centre for **Genomic Pathogen Surveillance**

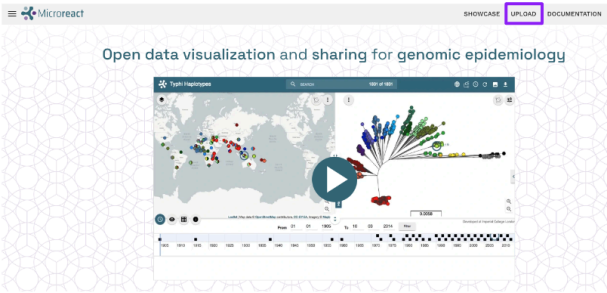|	
	Shuffles the nodes in a network.
	Automatically adjust the tree or network size to the current panel size.
	Controls whether the tree is stretched/compressed in all four directions, only horizontally, or only vertically. Click the icon to toggle through the three modes.
	This appears in the data table headers and allows you to filter by values or build filter conditions in any of the columns.
	In the map panel this viewport button filters the dashboard to only things currently visible in the map panel.

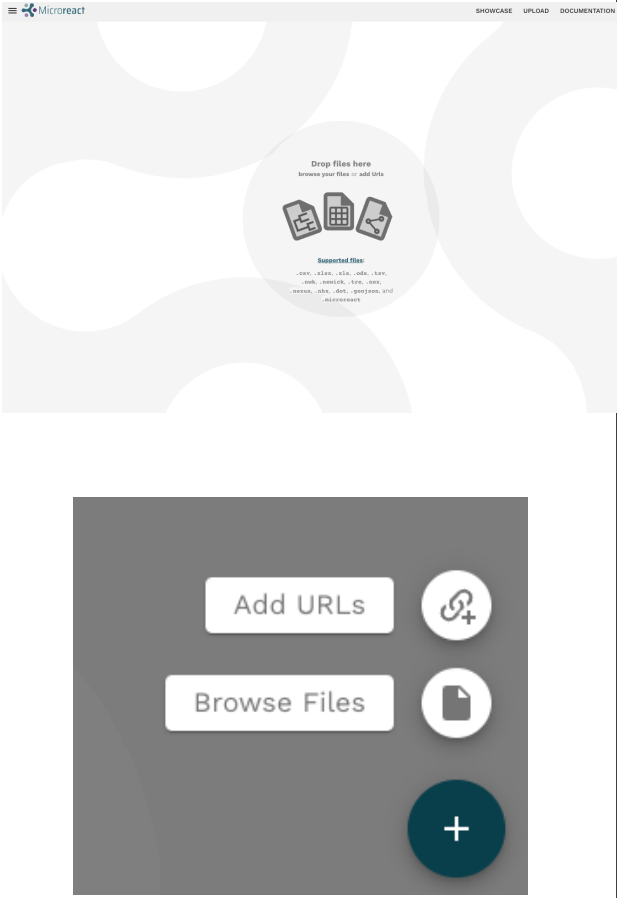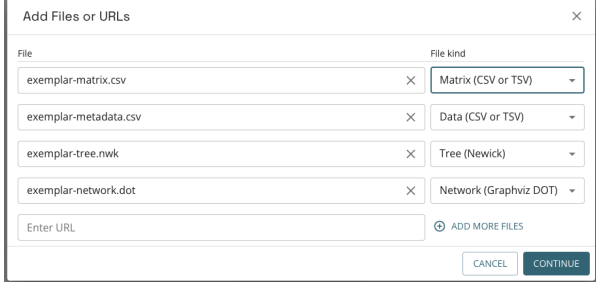| Microreact terminology		
Name	**Description**	**Image**
Panel	Each individual dashboard element in Microreact is called a panel. The dashboard in the image has two panels - metadata and chart.	
Side panels	On the far right of the screen is a side panel with options: **Legend:** shows the legend for the colour and shapes in use in the project. **Selection:** shows a donut chart of the distribution of whatever column you choose for the currently selected samples. **History:** shows a history list and allows you to revert to a prior state. **Views:** allows you to create and save multiple customised dashboard presentations in a single project.	

Centre for Genomic
Pathogen Surveillance

Title lozenge	Each panel has an editable title. If you have panels overlaid, there will be multiple sections.	Epi-links **Heatmap**
Search-and-filter bar	Typing in the top search-and-filter bar will filter the data across all columns in the metadata. Clicking the "E" will enforce an exact match requirement.	🔍 E kaufma	0 of 110
Top menu bar	Always visible along the top.	☰ Demo project ⓘ 🔍 E SEARCH IN ALL COLUMN 110 of 110 ✏ 👁 ⬇ 👥⚙
Left menu bar	Accessed through the hamburger icon at the left of the top menu bar.	Microreact — Showcase, Upload, Documentation, My Account, Sign out, Send Feedback

Building an exemplar data dashboard

Now that you're familiar with the basics of navigating Microreact, this section will walk you through an exemplar outbreak. All metadata in this outbreak is simulated and not based on any real data. Nonetheless, we hope this will help you understand how you can use Microreact to visualise different datasets together to make relevant decisions.

Importing data

Upload files
Navigate to the Upload page On the home screen, click the "Upload" button in the upper right corner. You may also use the "Upload" button located in the main menu (see Navigating the Site section of this tutorial).

Centre for Genomic Pathogen Surveillance

Browse your files

Clicking the "Upload" button brings you to the screen shown here.

Use the ⊕ icon to select the files you want to upload. Alternatively, you can drag and drop the files directly into the centre circle.

For this exercise we will create a project with the following files:

Exemplar_metadata.csv
Exemplar_tree.nwk
Exemplar_network.dot
Exemplar_matrix.csv

The files can be found here:
https://drive.google.com/drive/folders/1n qQ10s-GX7Nsu9-HN7w82Wph_wDBtxSG

💡 Tip - Microreact supports the following file types: **.csv**, **.xlsx**, **.xls**, **.ods**, **.tsv**, **.nwk**, **.newick**, **.tre**, **.nex**, **.nexus**, **.nhx**, **.dot**, **.geojson**, and **.microreact**

Verify your file type

Microreact will automatically identify each file and each file type. Double check that each file is present and the file types are correct.

Exemplar_metadata.csv > Data (CSV or TSV)
Exemplar_tree.nwk > Tree (Newick)
Exemplar_network.dot > Network (Graphviz DOT)
Exemplar_matrix.csv > Matrix (CSV or TSV)

Select the ID column for each file

You will then be asked to specify the ID column in the metadata file. This needs to be a unique value for each row in the dataset. In this example, it's the sample_id column.

If you upload multiple data files, they **must** have a common ID that allows Microreact to join the datasets.

Microreact will make a guess at each data files' ID column. Verify it is correct and then click "Continue" to move to the next file.
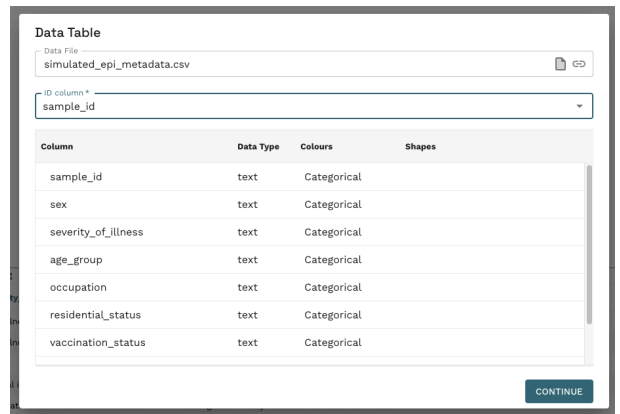
Select the labels column for the tree, matrix and network files

The tips of the tree need to have an ID that matches to an ID in your metadata in order to be able to colour, label, and add metadata blocks.
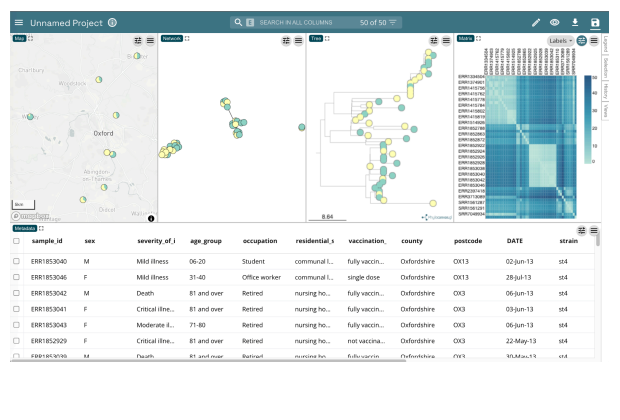
Similarly, each node in your network file needs to have an ID that matches to an ID in your metadata in order to be able to colour nodes and display different metadata as node labels.
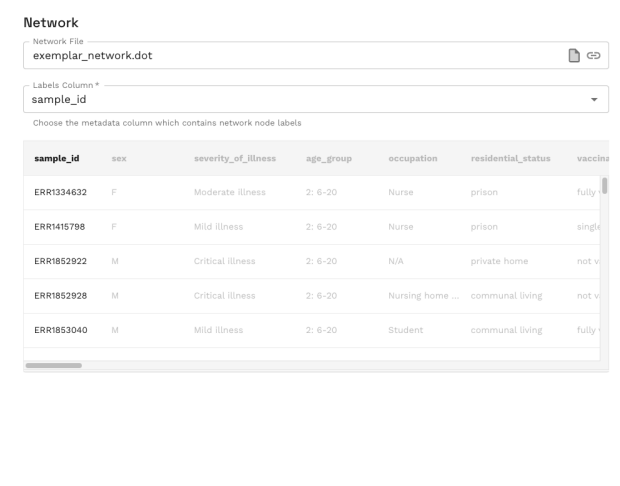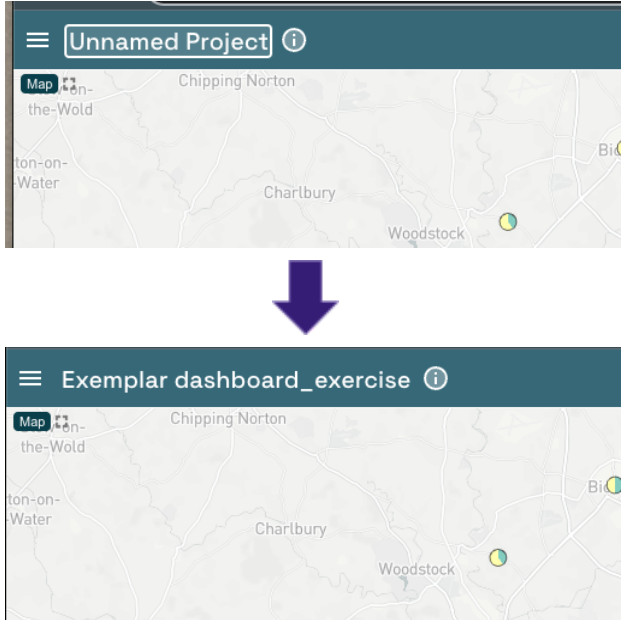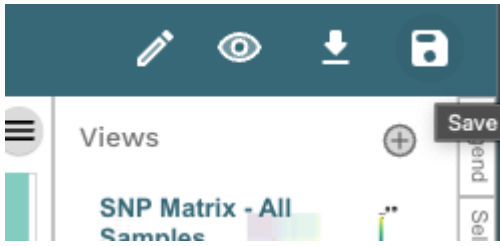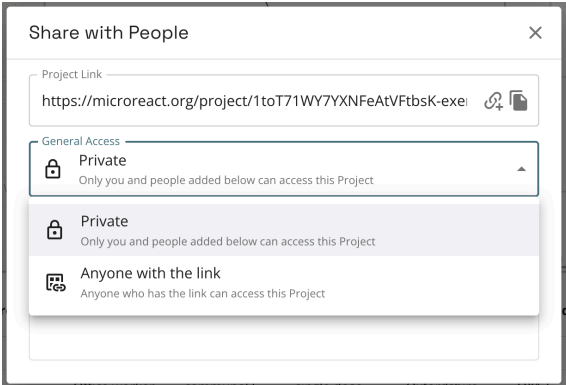
View your initial dashboard

Microreact will automatically bring you to your dashboard.

For some purposes, this output might be satisfactory for your use case. However we will be enhancing this to showcase many of Microreact's features.

Saving your project

Saving your Microreact project	
Rename your dashboard Since we will continue to edit this dashboard, let's give it a name so we can easily locate it on our Accounts page. In the upper left corner, double click the field "Unnamed Project". You will now be able to edit the title. Rename the project "Exemplar dashboard_exercise" and hit enter on your keyboard.	
Save your project Navigate to the upper right corner of the page and click the Save icon ![save icon] .	
Save as new project Click "Save as New Project" to add this dashboard to your account.	
Setting access permissions for your project You will then be shown a link to access your project at a later time or share your project with others. Select whether you would like to save the project to your individual account or whether you would like it to be accessible to others. For this exercise, let's select "Private" so it will only be available in your personal account.	

Editing the panels
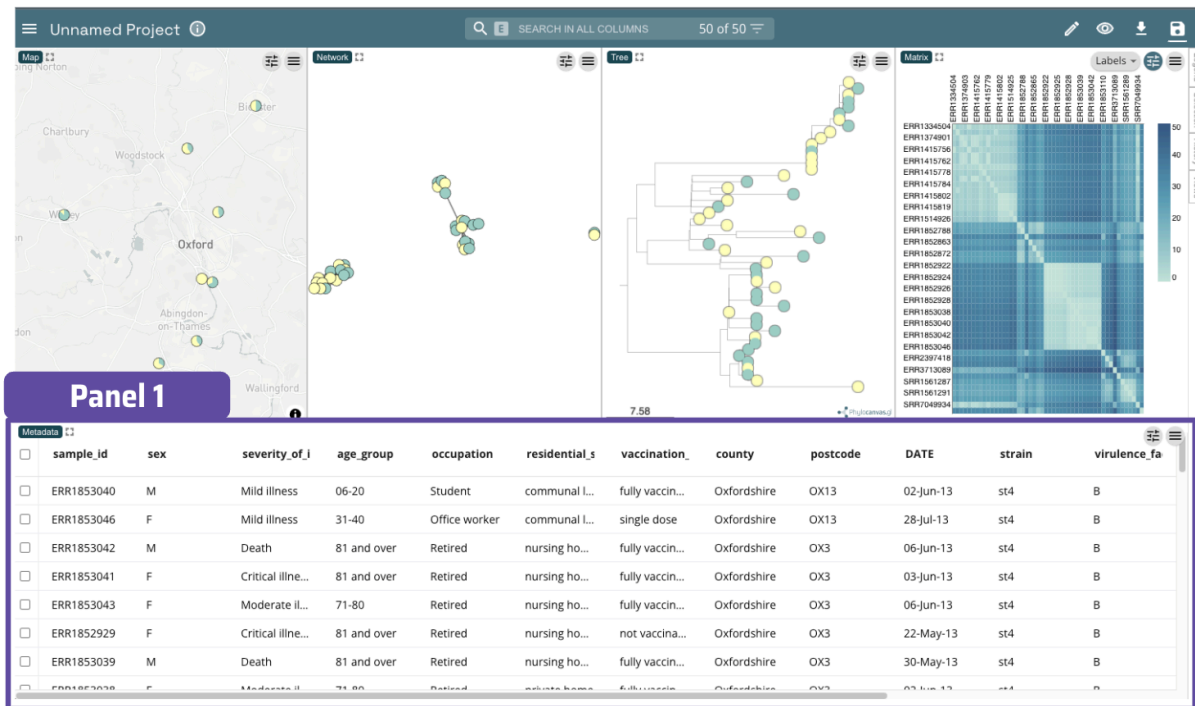
Panel 1

Panel 1: Metadata panel

View the metadata

The Metadata Panel appears at the bottom of the dashboard. Scroll through the rows and columns to view the imported data.
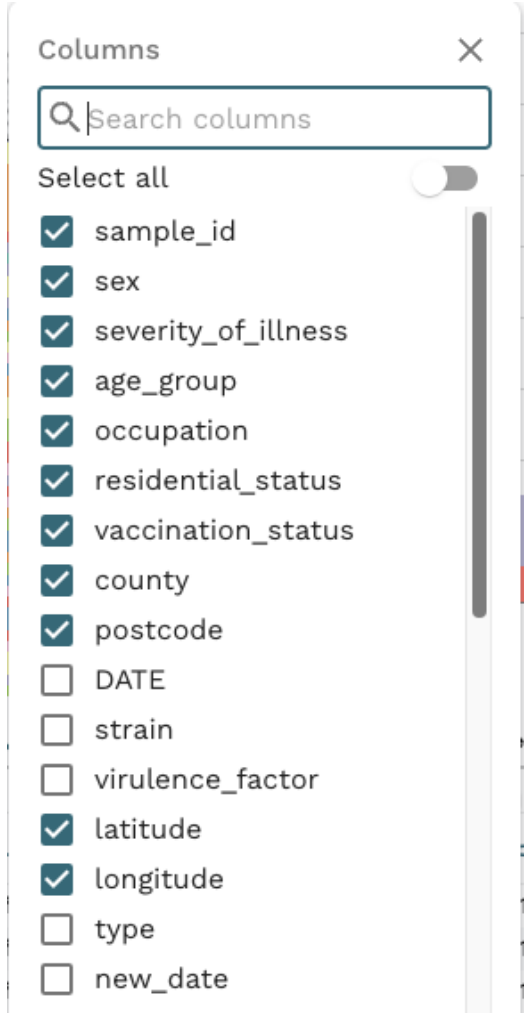
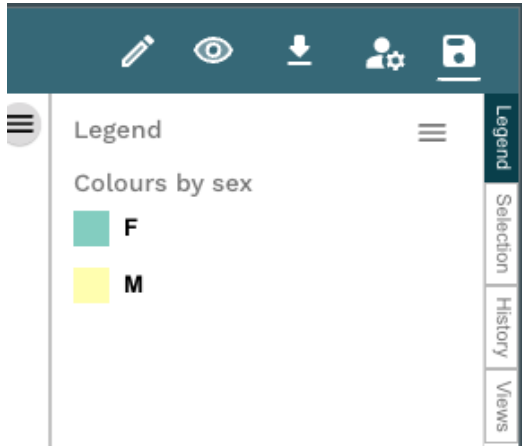You can use the Metadata Panel to search through your metadata, sort the columns, rearrange the columns or hide unneeded columns.

Adjust displayed data

Select the ⚏ icon. Two new lozenges will appear called "Columns" and "Density"

Filter columns

By default all columns in your dataset are shown in the metadata table.

Select the "Columns" lozenge and a list of all columns in the metadata will appear.

Tick/untick the green checkboxes to show/hide metadata columns on your dashboard.

Toggle the "Select all" button off. Now manually check the following columns:
- sample_id
- sex
- severity_of_illness
- age_group
- occupation
- residential_status
- vaccination_status
- country
- postcode
- latitude
- longitude

Tip - you can also filter the metadata rows. Tick any of the checkboxes to the left of your sample_id column. This action highlights the selected sample(s) in **all** dashboard visuals.

Columns	✕
Q Search columns	
Select all	⬤
☑ sample_id	
☑ sex	
☑ severity_of_illness	
☑ age_group	
☑ occupation	
☑ residential_status	
☑ vaccination_status	
☑ county	
☑ postcode	
☐ DATE	
☐ strain	
☐ virulence_factor	
☑ latitude	
☑ longitude	
☐ type	
☐ new_date	

Show the legend

Click the "Legend" tab on the side panel to show the legend. This will show all active colours and shapes on the dashboard.

After importing our data, we see that the Map, Network and Tree files are all displaying data by sex (notice the consistent green and yellow throughout the dashboard)

Legend

Colours by sex

🟩 F

🟨 M

Legend · Selection · History · Views

Centre for **Genomic Pathogen Surveillance**

Change the colour scheme

Click on the eye icon in the upper right corner of the dashboard.

The **Colour Column** is the field used to colour map markers, tree nodes, chart features and timeline blocks. Any column in the metadata can be used to colour the panels.

To change the colour scheme of our Tree Panel to match the colour scheme of our Bar Chart Panel (see Panel 3 later), select "residential_status" as the **Colour Column**.

Labels Column
sample_id

Colour Column
residential_status

Autocolour

sample_id

sex

severity_of_illness

age_group

occupation

residential_status

vaccination_status

county

postcode
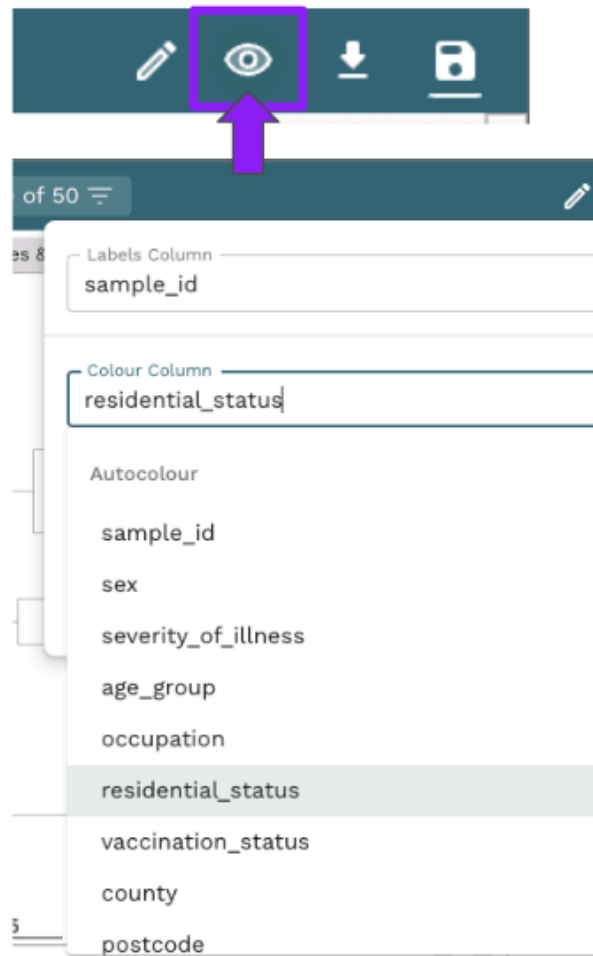
Creating dashboard views

Each project has a side panel which will show the Legend, Selections, History and Views for the entire project. In this demonstration, we will focus on the **Legend** and **Views** panes as these are the panes you will use the most.

The **Legend** reflects the colours and labels used in the project.

The **Views** pane shows a list of saved views associated with the current project. Think of a view as a snapshot of your dashboard with a particular filter, sample selection, or group of visuals that you want to refer back to. You can create multiple views within a project to tell a story or explore a subset of geographies or cases. Each saved view has its own URL, so an individual URL can be shared to bring others directly to a specific view. Whatever you are looking at on your dashboard will become your view. Set your dashboard as you like, add a view, and give it a name. If you wish to change what you have done, manipulate the dashboard, and select "update view", then save the entire project. When updating, make sure you select the View you wish to change or else a different View may be overwritten.

ⓘ Learn more about the side panel options such as Views here:

Centre for **Genomic Pathogen Surveillance**

https://docs.microreact.org/instructions/interacting-with-microreact-projects

Create a View

Click the "Views" tab on the side panel (located beneath the Legend tab).

Click the plus button to save this view or snapshot of the dashboard. This will create a small screenshot of your dashboard, initially named "Unnamed View"

Hover over the image until a circle icon appears ⋮ . Click on the circle icon, select Rename View and name the view "Overview - All Samples"
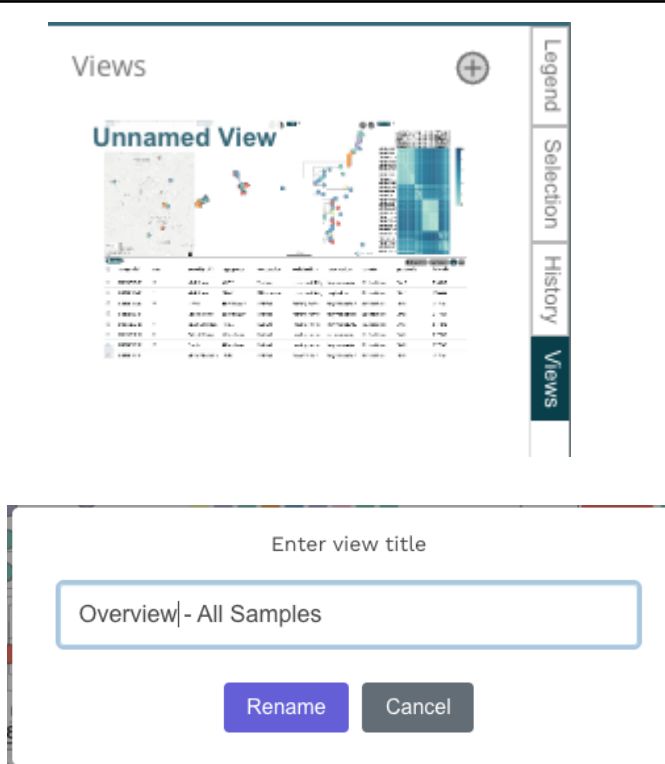
Save the View

Hover over the image until a circle icon appears ⋮ . Click on the circle icon, select Update View. This saves the view in your side panel.

Any time you make adjustments to your dashboard that you would like to save, return to this menu to update the View.

💡 Tip - Don't navigate away from a View until you've selected "Update View" to save any changes, and don't forget to save the whole project after making new Views or changes to Views.

💡 Tip - Saving a View will also save all associated formatting. Feel free to resize or reorder the panels and then update the view to lock in those changes.
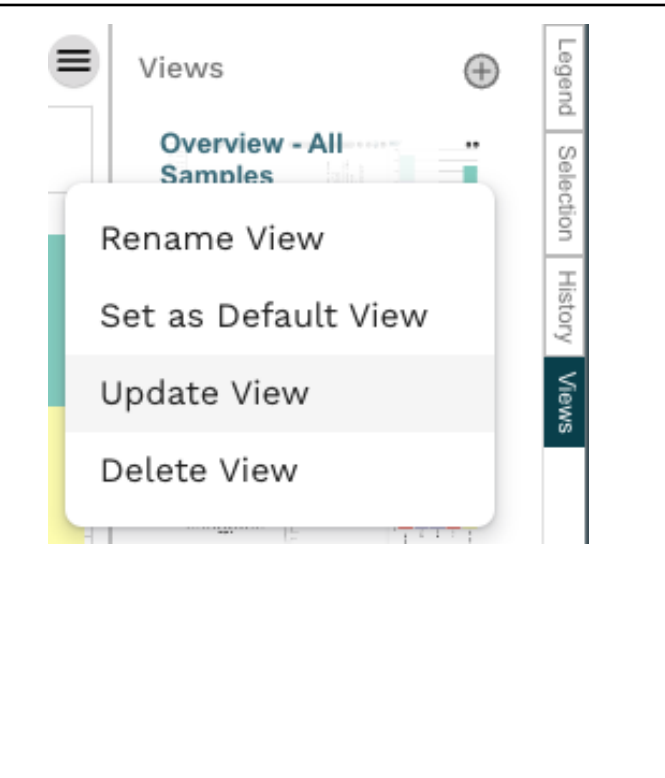
Panel 2: Timeline panel	
Create a timeline panel Locate the [pencil icon] icon in the upper right corner of the screen. Click "Create New Timeline".	Create new Chart Create new Data Slicer Create new Matrix Create new Map Create new Network Create new Note Create new Table Create new Timeline Create new Tree Edit Existing Panels
Drag and drop the Timeline Panel onto your dashboard Use your mouse to drag the Timeline Panel so that it overlays the Metadata Panel. Once the panel is in place, click your mouse to insert the panel.	
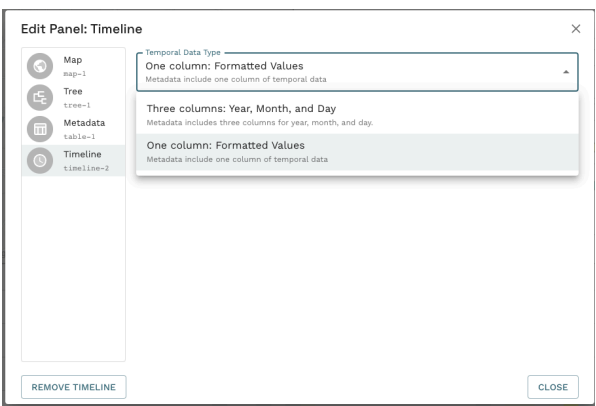
Centre for Genomic Pathogen Surveillance

Specify the temporal data type

In the dialog box that appears, specify whether the temporal data is stored in 1 column or 3 separate columns.

For this exercise, our temporal data appears in one column. Select "One column: formatted values" from the drop down menu.

Specify the temporal data column

Use the drop down menu to select the "new_date" column from your metadata that will be used to create the timeline.

Final result

The resulting Timeline Panel should look like this figure.

Notice how you are now able to toggle between the Metadata and Timeline panels.

Update View

As a best practice, after you are happy with the updates you've made to a panel, go ahead and update the View to save your changes.

Hover over the sidebar View image until a

circle icon appears . Click on the circle icon, select Update View. This saves the view in your side panel.

Centre for Genomic Pathogen Surveillance

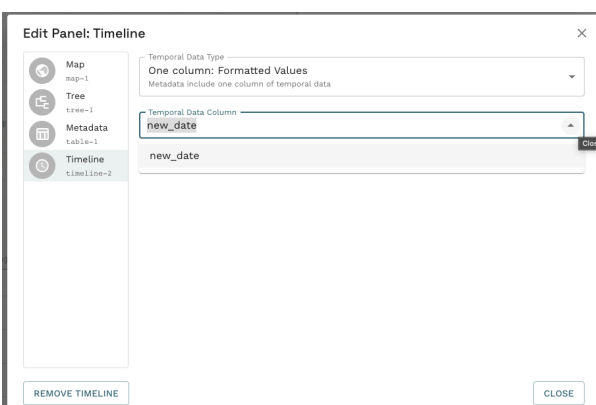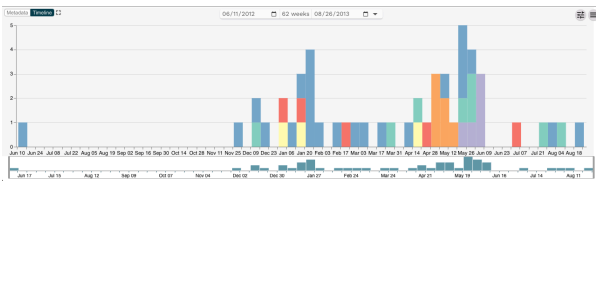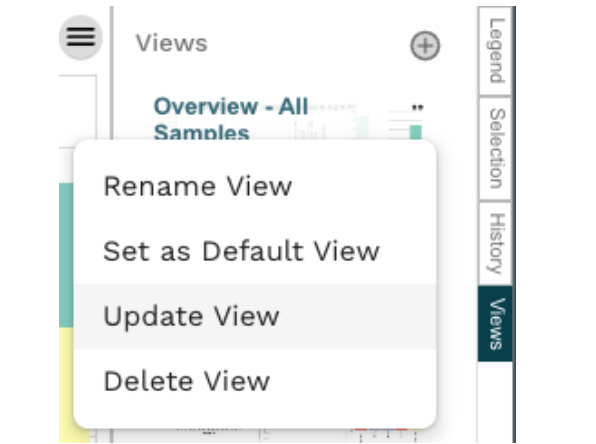| Panel 3: Chart panel	
Create a chart panel Locate the [pencil icon] icon in the upper right corner of the screen. Click "Create New Chart".	📊 Create new Chart ▽ Create new Data Slicer ▦ Create new Matrix 🌐 Create new Map △ Create new Network ▭ Create new Note ▥ Create new Table 🕐 Create new Timeline ⼚ Create new Tree Edit Existing Panels
Drag and drop the Chart Panel on your dashboard Use your mouse to drag the Chart Panel so that it overlays the Timeline Panel. Once the panel is in place, click your mouse to insert the panel. A white canvas will appear for you to start building the chart.	
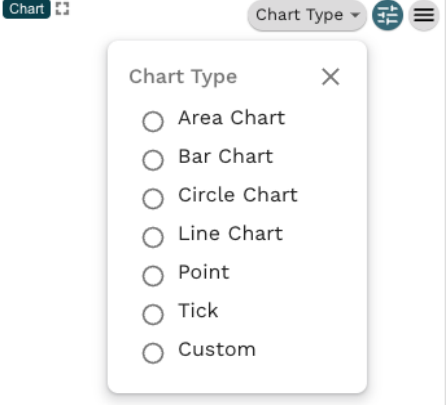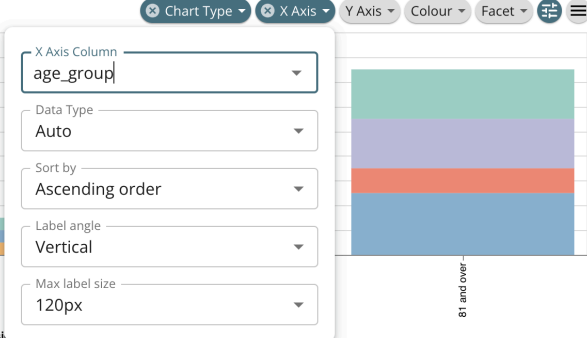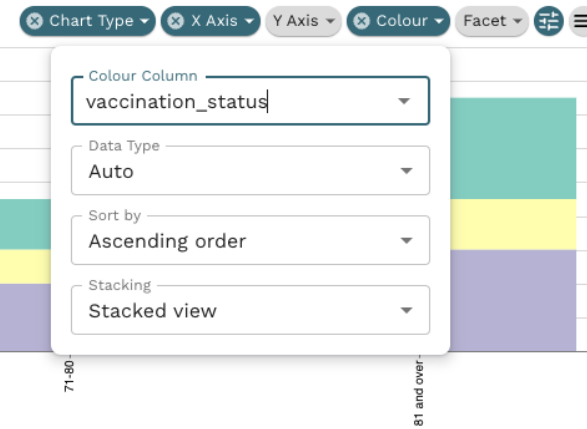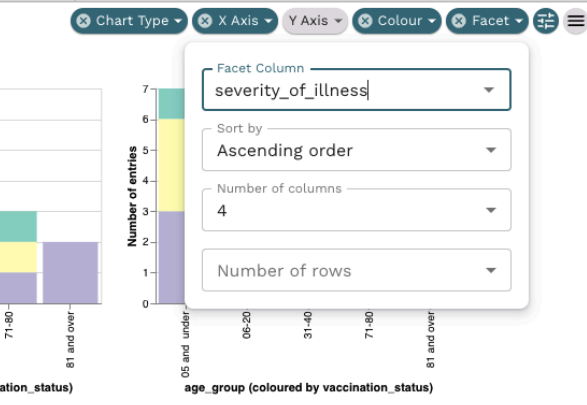
Centre for Genomic Pathogen Surveillance

Select your chart type Navigate to the "Chart Type" lozenge. Select "Bar Chart".	
Change the x-axis Navigate to the "X Axis Column" lozenge. Use the drop down arrow to specify "age_group" as the x-axis value.	
Change the criteria for column colour Navigate to the Colour lozenge. Select "vaccination_status" as the Colour Column.	
Change the chart facet Navigate to the Facet lozenge. Select "severity_of_illness" as the Facet Column and change the number of columns to 4. Tip - Click the icon again to make the menu options disappear. This helps to visually tidy your dashboard!	

Centre for Genomic Pathogen Surveillance

Final result The resulting Chart Panel should look like this figure. Note, your Metadata and Timeline Panels are still accessible. Select the title lozenges to toggle between them.	
Update View If your chart panel matches the above, go ahead and update the View to save your changes. Hover over the sidebar View image until a circle icon appears ⋮ . Click on the circle icon and select Update View. This saves the view in your side panel.	

Panel 4: Tree panel

Edit the Tree Panel Metadata Blocks

Metadata blocks are coloured blocks that sit next to each tip on a phylogenetic tree and are coloured based on a selected column in the metadata. These are a powerful tool because you can view multiple columns of data at once.

Select the Metadata Blocks lozenge.

All metadata columns appear in this menu. Tick the boxes of the metadata fields you would like to appear on your Tree Panel.

For this exercise, select the following blocks:
- Sex
- Severity_of_illness
- Age_group
- Vaccination_status
- Postcode

💡 Tip - Click the ⇄ icon again to make the menu options disappear.

Final result

The resulting Tree Panel should look like this figure.

Update View

If your tree panel matches the above, go ahead and update the View to save your changes.

Hover over the sidebar View image until a circle icon appears ⋮ . Click on the circle icon, select Update View. This saves the view in your side panel.

Create a second View

Now that we have saved our main view, we may have other ideas for additional views to create from our dashboard. Let's create a second View to show cases involved in an imaginary outbreak.

First, toggle back to our Chart Panel.

Click the "Views" tab on the side panel, and then click the plus button.

This will create a small screenshot of your dashboard, initially named "Unnamed View"

Hover over the image until a circle icon appears ⋮ . Click on the circle icon, select Rename View and name the View "Outbreak Cluster"

Now, let's add some additional modifications to this second View.
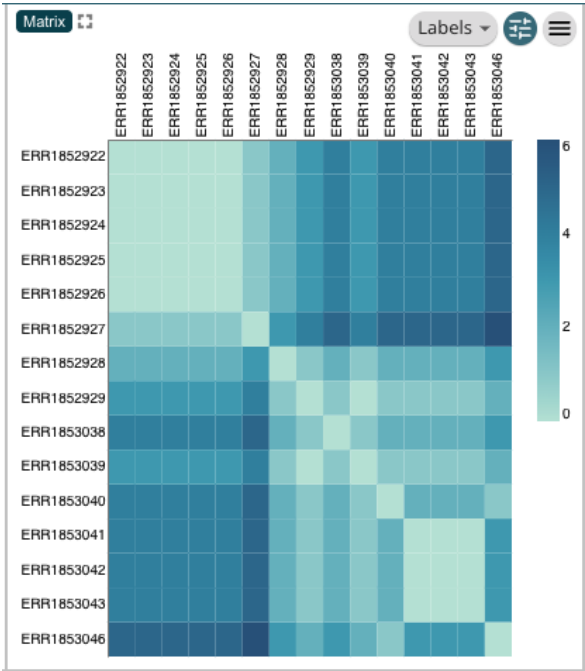
Changing our colour scheme

For this view, we want to investigate a potential outbreak. In this example, we will next view our data by virulence factor to see if any patterns in our data emerge.

Navigate to and click on the eye icon in the upper right corner of the dashboard.

Change the colour column of our view to "virulence_factor".

Viewing sub-trees

After viewing our data by virulence factor, a distinct cluster appears in our tree panel.

Right click on the node of the top branch and select "View Subtree".

This action not only zooms in on this portion of the tree but it also filters the remaining panels to show only the samples in this cluster.

Update the View

Now that we are visualising data from a specific cluster to further evaluate our outbreak, we are ready to update this cluster investigation View to share with others or refer back to later.

Hover over the sidebar View image and click the circle icon, then click Update View.

💡 Tip - You will see the changes we made to the View represented in the sidebar image.

Centre for Genomic
Pathogen Surveillance
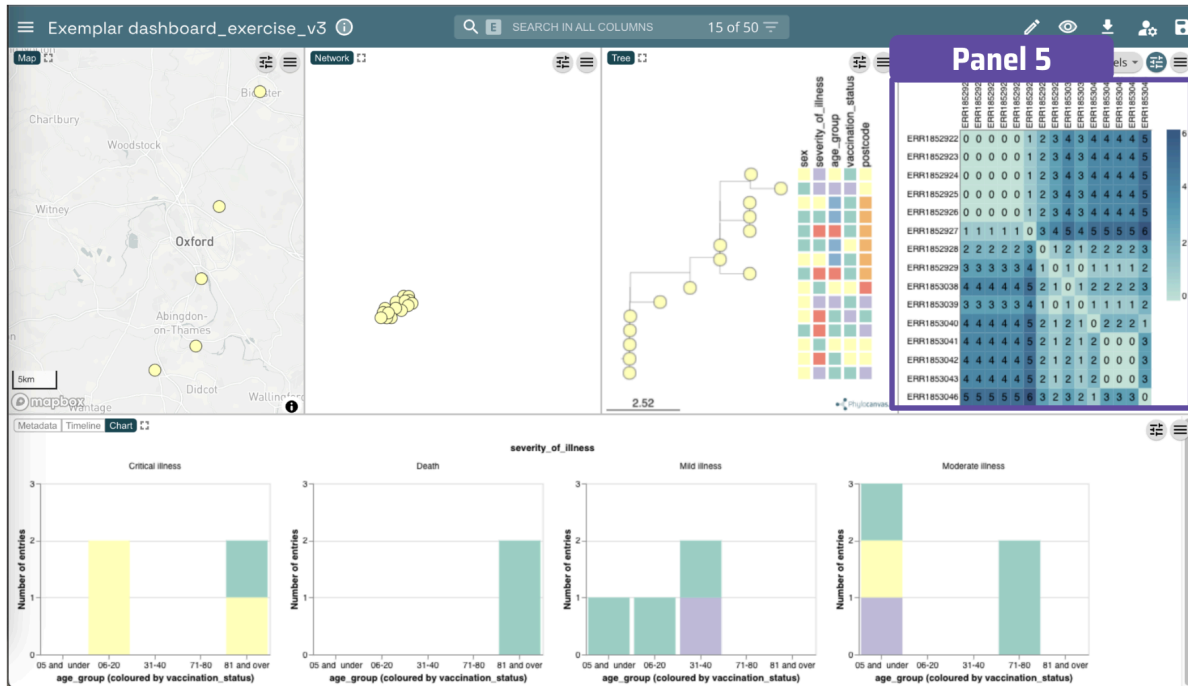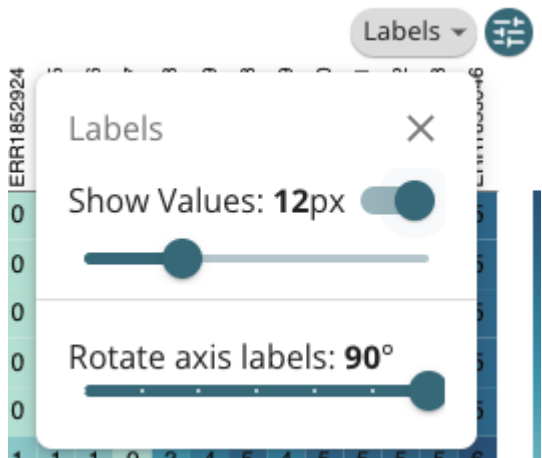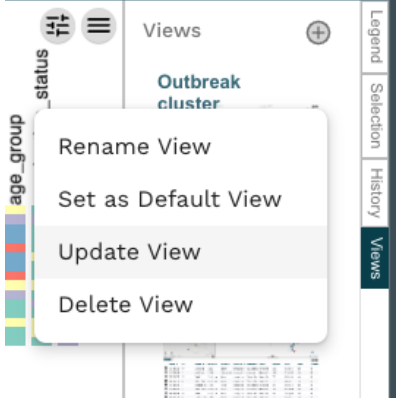
Panel 5: Matrix panel

Add matrix labels

Navigate to the matrix panel control menu

by clicking the ⚙ icon. Notice the Labels lozenge appears.
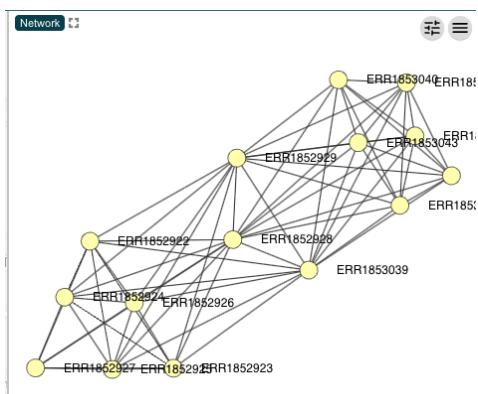
Adjust value size Click the Labels lozenge and toggle on the Show Values feature. This will show any specific row/column relationships if highlighted in your matrix csv file.	
Update the View Hover over the sidebar View image and click the circle icon then click Update View.	

Genomic Pathogen Surveillance - Tools & Interpretation

Panel 6: Network panel
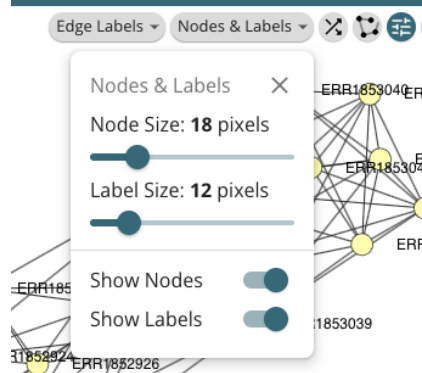
Zoom into the network

Use your mouse scroller or the zoom

buttons ⊕ 🔍 in the network panel to zoom in on the cluster.
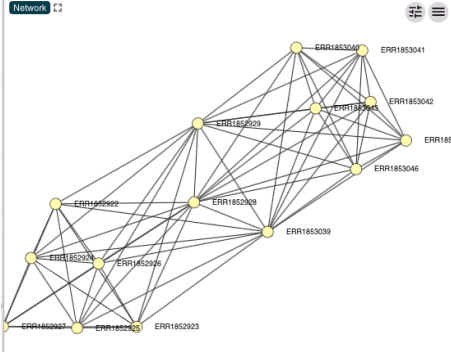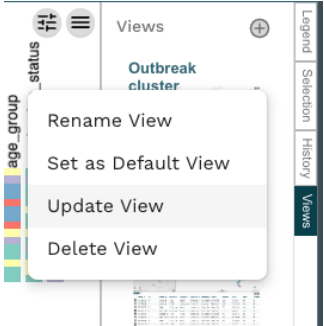
Add network labels

Navigate to the network panel control menu

by clicking the ⚙ icon.

Centre for Genomic Pathogen Surveillance

Adjust the label size Use the scroll bar to adjust the size of the labels and/or nodes until you are happy with how your network is displayed.	
Update the View Hover over the sidebar View image and click the circle icon, then click Update View.	

Panel 7: Map panel

Add GeoJSON map features Navigate to the pencil icon at the top right of the screen and select Edit Existing Panels	
Locate the Edit map panel Click the Map panel from the left column of project panels. Here you have the option to add in a GeoJSON file which will add boundaries to your map. Click the upload icon and locate the file provided to you for this exercise. Click Close.	
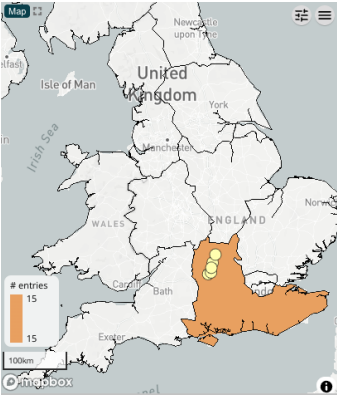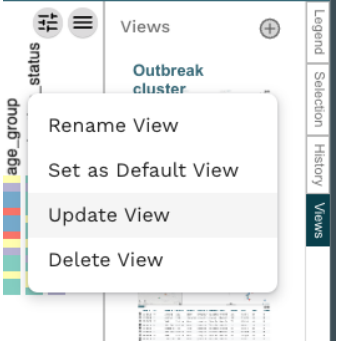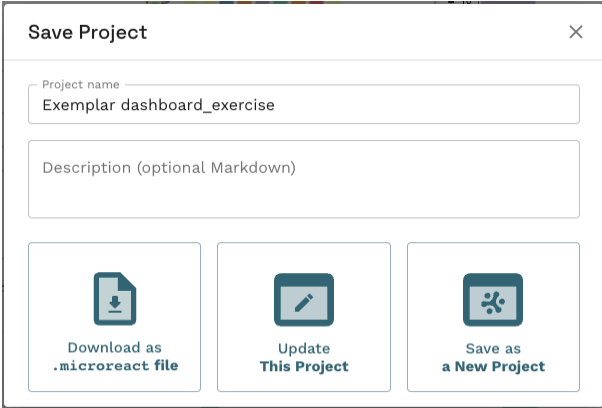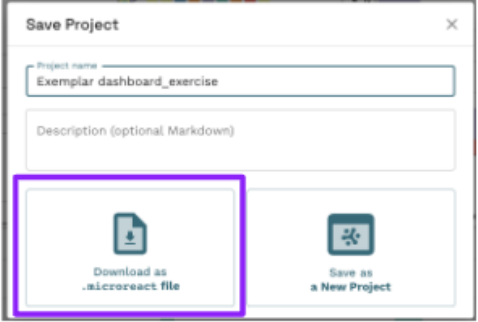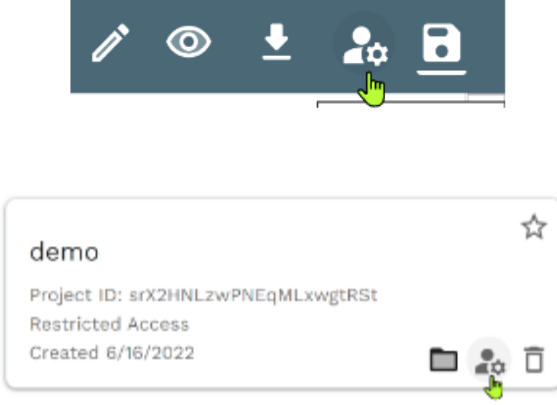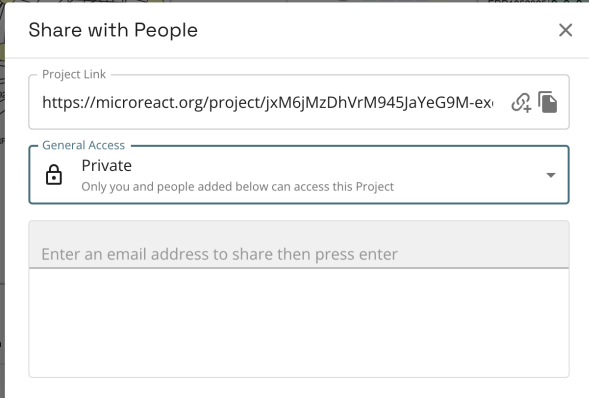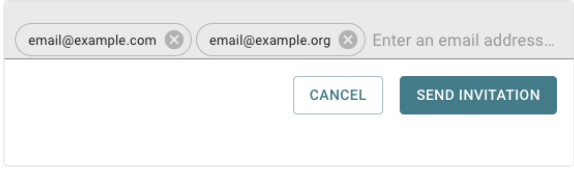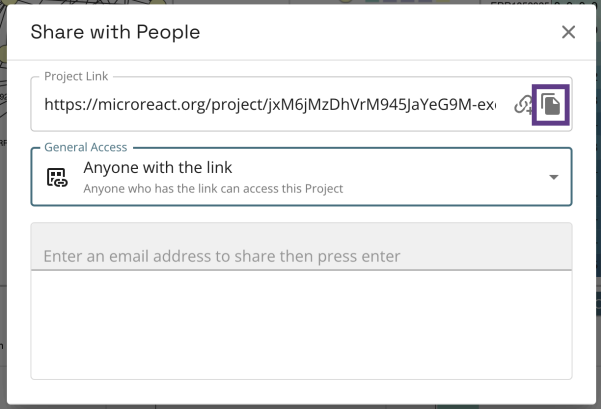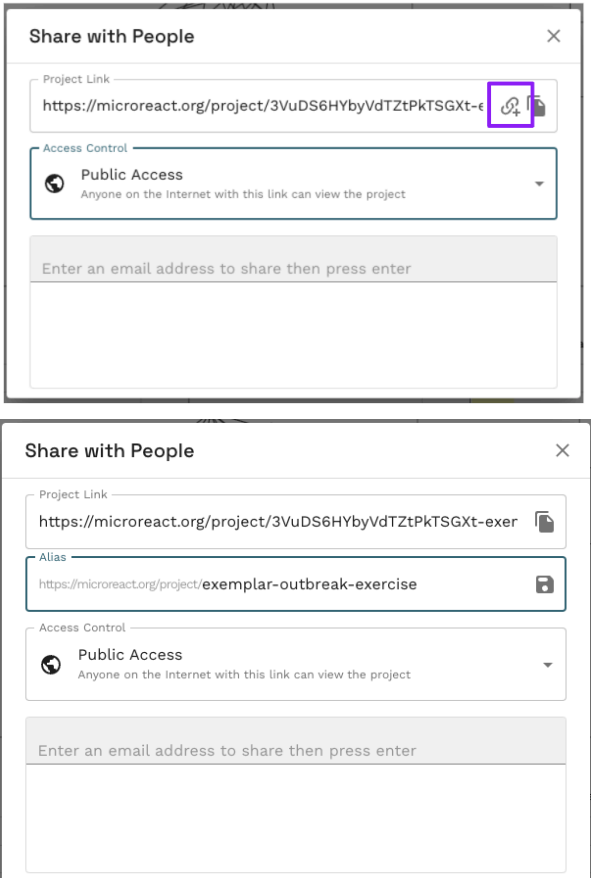
Centre for Genomic Pathogen Surveillance

Resize the map panel Microreact will return to the Outbreak Cluster view. If needed, use the scroll tool on your mouse to zoom out so you can see the entire UK map. With this feature, it is clear that all isolates in this cluster were living in the same geographic area.	
Update the View Hover over the sidebar View image and click the circle icon, then click Update View.	
Update the project After you save the project, you may continue to make changes. Be sure to return to the icon and select "Update this Project" to save those changes.	

Access permissions and data sharing

Sharing your project There are several ways to share your work with others. **Option 1:** First, you may download the .microreact file and send it as an attachment to an email or upload it to a shared drive. **Option 2:** Another way is to navigate to the Project Access icon [icon] in the upper right corner of the screen. Clicking this button will bring you to the Share with People menu.	**Option 1:** **Option 2:**
Navigate to the Project Access menu Select the Project Access menu icon at the top of the screen or within the project folder on your Account page. By default, a saved project on microreact.org can only be accessed by the user who created it. Access must be explicitly granted for others to view it.	
Give a project Private Access You have the option to restrict access to the project to a limited number of other users. Select Private from the drop down menu in the Share with People screen. You can grant others access to your project by providing their email addresses (up to 30 invitees).	

Centre for Genomic
Pathogen Surveillance

Share with anyone with the link

Select "Anyone with the link" from the drop down menu on the Share with People screen.

Click the copy icon to the right of the Project Link provided to copy the link and share your project with whom you choose.

💡Tip - If you have a locally installed version of Microreact, "Public Access" will be limited to the user accounts within your network.

Adding an alias before sending your projects

Before sharing your project link with others, it may be helpful to edit the URL to a shortened and/or more meaningful version.

Click the alias icon to the left of the Project Link provided.

An Alias menu appears for you to edit the URL. Click the save icon when you are finished to save your changes.

Centre for Genomic
Pathogen Surveillance

Downloading projects, views and panels

Each of the original files used to create a project can be downloaded from the Download Project Files menu.

Data can be downloaded by anyone with access to the project.

Individual panel visuals can be downloaded by clicking the ☰ icon in each panel. You have the option of downloading the original file format or a static image (PNG or SVG).

Original project files or image files can then be shared with others.

Centre for **Genomic Pathogen Surveillance**
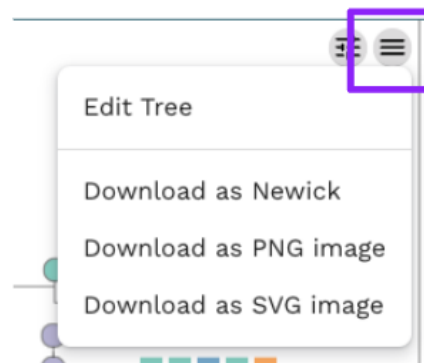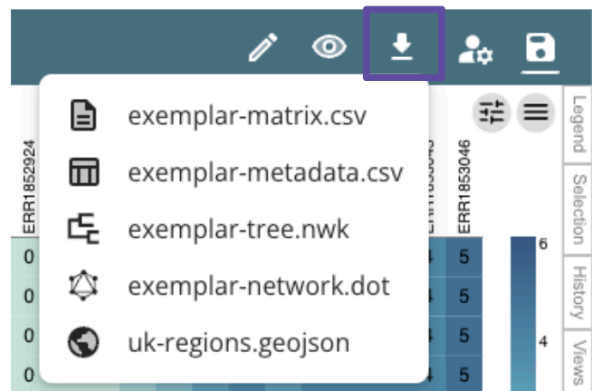
Exercise 3: Investigating a hospital outbreak of MDR *Klebsiella pneumoniae*

In this exercise, we will perform a genomic epidemiology investigation using the Pathogenwatch and Microreact tools that we explored previously.

You can work in groups of 2-3 people to complete the exercise.

Note that all events and settings in this exercise are fictional and any resemblance to a real-life situation is coincidental.

Scenario

You are a genomic epidemiologist working at the National Agency for Public Health in Lagos, Nigeria. In the recent days and weeks, a number of *Klebsiella pneumoniae* isolates have been sent to your reference laboratory from Kano Central Hospital, northern Nigeria, for further characterisation. The hospital has reported a cluster of cases, some involving bloodstream infections, in its neonatal unit. All isolates are resistant to third-generation cephalosporins while some are also resistant to carbapenems.

Six isolates from this cluster have been whole genome sequenced in your laboratory using Illumina technology. The sequence reads have been assembled and the genome assemblies are available for analysis.

The following metadata relating to these six isolates has been provided to you by Kano Central Hospital.

Isolate ID	Patient	Sampling date	Sample type	Phenotypic resistance?
36773_1#336	Patient 1	28/08/2024	Blood	Third-generation cephalosporins
37775_1#167	Patient 2	20/08/2024	Blood	Third-generation cephalosporins
35876_2#299	Patient 3	30/08/2024	Blood	Third-generation cephalosporins
38089_1#139	Patient 4	23/08/2024	Sputum	Third-generation cephalosporins, carbapenems
38089_1#134	Patient 5	03/09/2024	Urine	Third-generation cephalosporins
40660_1#353	Patient 6	05/09/2024	Faeces (carriage)	Third-generation cephalosporins, carbapenems

Centre for Genomic
Pathogen Surveillance

Key questions:

- How related are the different isolates to each other at the genomic level, and do the results support the occurrence of nosocomial transmission?
- Which sequence type(s) (ST) do the isolates belong to?
- Which antimicrobial resistance (AMR) markers are these isolates carrying?
- Do the isolates carry any acquired virulence markers?
- Are any of the isolates similar to other previously-sequenced isolates from Nigeria or elsewhere?
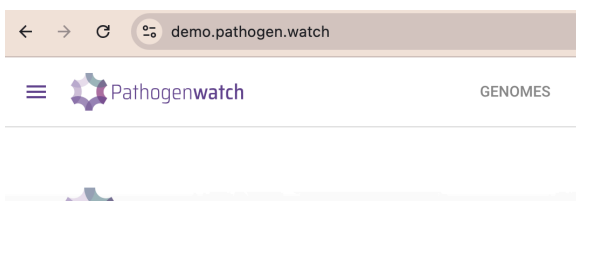
At the end of the exercise, you will compile a brief report to send back to the hospital. This will be presented back to the group.

Exercise data

All data for this exercise can be found here:
https://drive.google.com/drive/folders/1XzXRVDPy70fboBAIa1PE0av6J7TY771o

Characterisation of the outbreak isolates in Pathogenwatch

Login to the PW-demo site For this exercise we will use the URL **demo.pathogen.watch** This is a "light" version of Pathogenwatch, used for training, but it looks and feels like the main Pathogenwatch site. Any data in your PW-demo account will be removed after the training session. Never upload confidential metadata to the demo site. Only https://pathogen.watch is appropriate for viewing such metadata. It is also worth looking at private metadata browsing if it is not possible to upload your metadata at all, e.g. due to personally identifying information (PII) or patient confidentiality: https://cgps.gitbook.io/pathogenwatch/how-to-use-pathogenwatch/private-metadata	← → C demo.pathogen.watch ≡ Pathogen**watch** GENOMES

Upload outbreak data to Pathogenwatch

Upload the six genome assemblies (*.fasta) listed in the table above and the associated metadata file (.csv). These are in the folder "Pathogenwatch Outbreak Data".

Remember that bacterial genomes are typically "Single Genome FASTAs".

Note that one isolate is a *Klebsiella variicola*. Often laboratory methods cannot distinguish between different species of the *Klebsiella pneumoniae* species complex (KpSC).
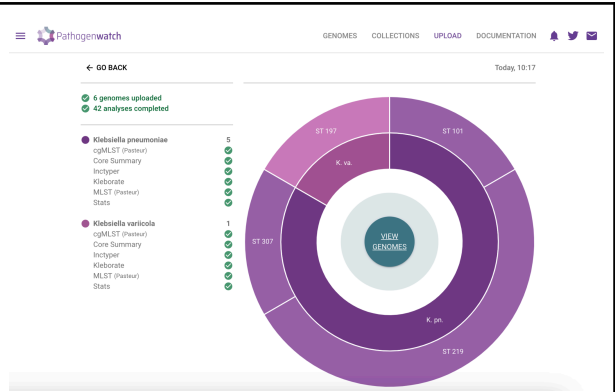
View the QC statistics

Click "View Genomes" and your six uploaded genomes will be listed in the Genomes page.

Click the "Stats" tab to view the QC information.

For *Klebsiella* genomes, a good rule of thumb is an assembly length of 5-6.5Mb and <300 contigs.
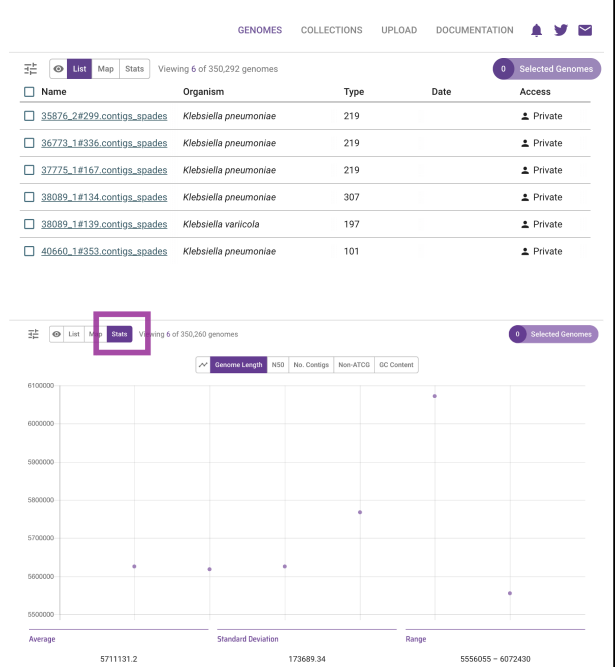
Explore the Genome Reports

Go back to the "List" tab where your six uploaded genomes are listed.

View the Genome Report for each isolate. In another document, make a table similar to the one below and start populating it with the available information.

Do the resistance markers explain the phenotypic resistances reported by the hospital? Are there any discrepancies?
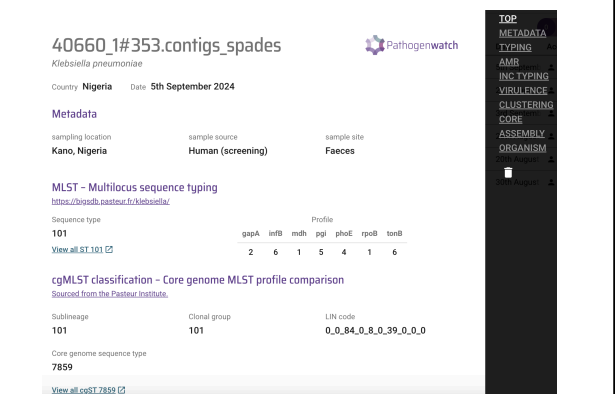
Characteristics of the outbreak isolates.

Isolate	Patient	ST	LINcode	Resistance markers (3GC and carbapenems only)	Virulence markers
36773_1#336	Patient 1				
37775_1#167	Patient 2				
35876_2#299	Patient 3				
38089_1#139	Patient 4				
38089_1#134	Patient 5				
40660_1#353	Patient 6				

You have now discovered that three of your isolates belong to the same ST, ST219, and share similar resistance and virulence marker profiles. ST219 has been associated with MDR infections in multiple countries including South Africa (Marais et al. 2024), Türkiye (Kurekci et al. 2024) and Germany (Xanthopoulou et al. 2022).

You search your own database at the National Agency for Public Health and discover that you have previously sequenced two *K. pneumoniae* ST219 isolates, one from Kano in 2022 and another from Lagos in 2020. You wonder whether there could be any connection between either of these two isolates and the current outbreak isolates.

The following metadata for these two historical isolates is available from your database.

Isolate ID	Sampling date	Sampling location	Sample type	Phenotypic resistance?
36773_1#371	07/03/2022	Kano, Nigeria	Urine	Third-generation cephalosporins
38089_1#189	25/10/2020	Lagos, Nigeria	Blood	Third-generation cephalosporins

To investigate the genomic relatedness of these two isolates with your contemporary ST219 isolates, you next upload these genome assemblies into Pathogenwatch (see below).

Comparison of historical and outbreak isolates in Pathogenwatch

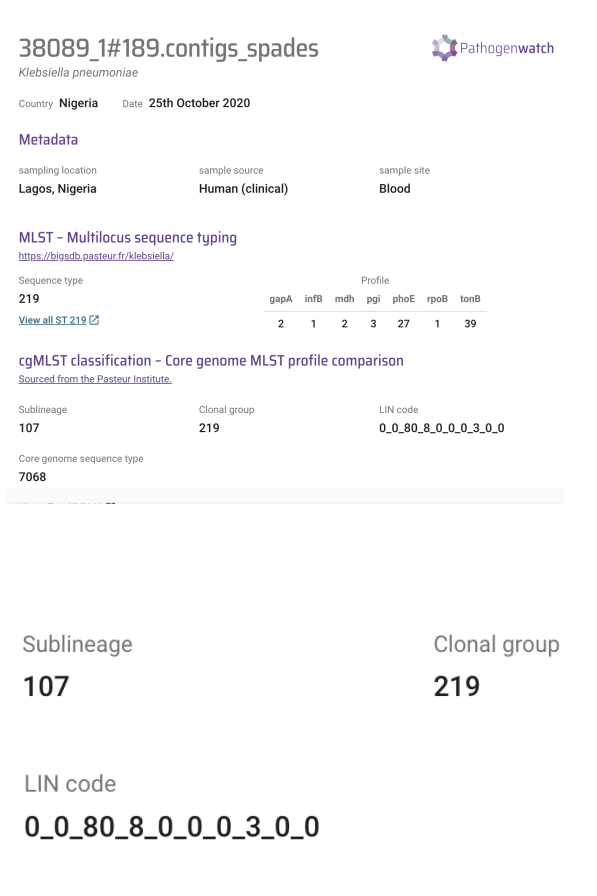|	
Upload historical data to Pathogenwatch Upload the two genome assemblies (*.fasta) listed above and the associated metadata file (.csv). These are in the folder "Pathogenwatch Historical Data". As previously, check the QC information in the "Stats" tab of the Genomes page.	

Centre for Genomic Pathogen Surveillance

Compare the historical genomes with contemporary genomes

Explore the Genome Reports of the two historical ST219 genomes. How related are these to the contemporary ST219 outbreak genomes?

💡 Tip. cgMLST profiles and LIN codes provide a rapid way of assessing genome similarity from assemblies, and provide increased resolution compared to 7-gene MLST.

LIN code prefixes are used to assign genomes to a "Sublineage" and "Clonal group", and can also be used to estimate the number of allelic differences between genomes. Identical LIN codes mean that two genomes share all 629 cgMLST alleles.

For more information on the cgMLST/LIN code approach for *K. pneumoniae*, see **https://www.biorxiv.org/content/10.1101/2024.03.11.584534v1**

38089_1#189.contigs_spades
Klebsiella pneumoniae

Country **Nigeria**   Date **25th October 2020**

Metadata

sampling location	sample source	sample site
Lagos, Nigeria	Human (clinical)	Blood

MLST – Multilocus sequence typing
https://bigsdb.pasteur.fr/klebsiella/

Sequence type   Profile
219

	gapA	infB	mdh	pgi	phoE	rpoB	tonB
View all ST 219	2	1	2	3	27	1	39

cgMLST classification – Core genome MLST profile comparison
Sourced from the Pasteur Institute.

Sublineage	Clonal group	LIN code
107	219	0_0_80_8_0_0_0_3_0_0

Core genome sequence type
7068

Sublineage	Clonal group
107	**219**

LIN code

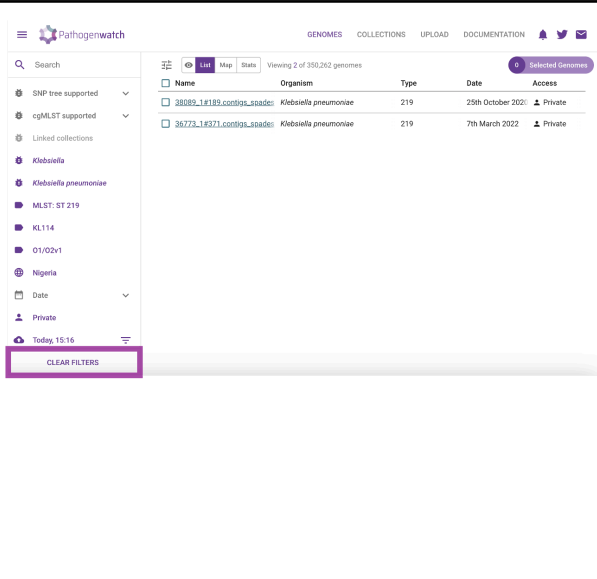**0_0_80_8_0_0_0_3_0_0**

Building a Pathogenwatch collection of private (uploaded) and public genomes

---

Build a Collection of ST219 genomes

Let's now build a combined Collection of all public *K. pneumoniae* ST219 genomes in Pathogenwatch and your 5 uploaded ST219 genomes (3 contemporary, 2 historical).

Note that the demo.pathogen.watch site only includes a small subset of public genomes compared with the main pathogen.watch site.
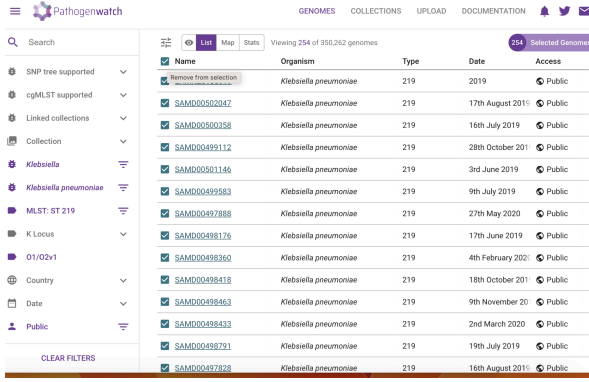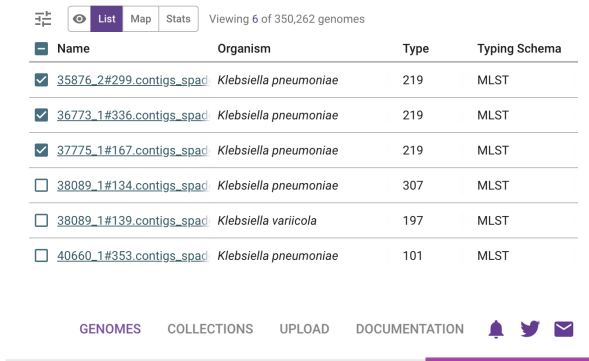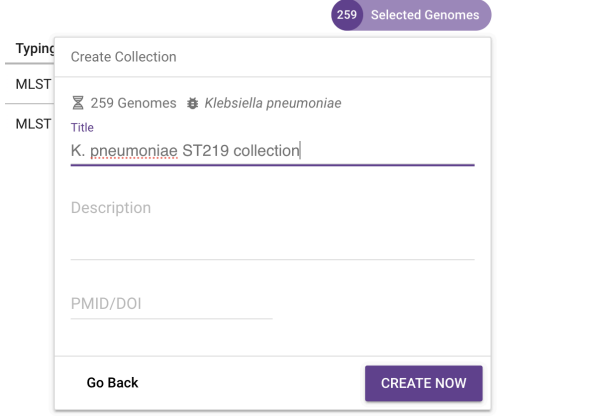
Start by clicking "Clear Filters" in the bottom left corner of the page. Also ensure that no genomes are selected yet in the "Selected Genomes" tab on the top right of the page.
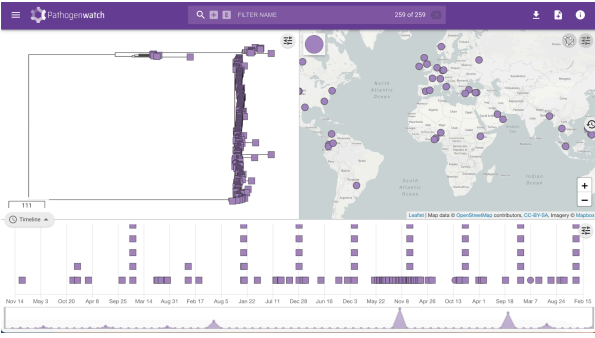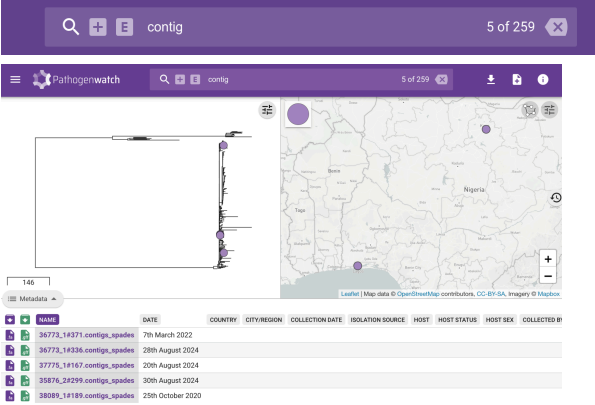
Centre for Genomic Pathogen Surveillance

Select all public ST219 genomes

Use the following filters:
Genus → *Klebsiella*
Species → *Klebsiella pneumoniae*
MLST Pasteur → ST219
Access → Public

This should list 254 genomes in the Genomes page.

Click the checkbox to the left of "Name" to select all genomes in the list. You should now see "254 Selected Genomes" in the tab in the top right corner.

Select your uploaded ST219 genomes

Click "Clear Filters" in the bottom left of the page again.

Now use the "Uploaded at" filter on the left hand side to bring up your 6 recently-uploaded outbreak genomes in the Genomes page.

In the list of 6 genomes, select the 3 ST219 genomes only using the individual tick boxes. You should see that you now have "257 Selected Genomes" in the tab in the top right corner of the page.

Now click "Clear Filters" again, and this time bring up your two recently-uploaded historical ST219 genomes. Select both of these using the tick boxes and you should now have "259 Selected Genomes".

Create a Collection

Click on the "259 Selected Genomes" tab and then click "Create Collection".

Give the collection a title, e.g. "K. pneumoniae ST219 collection", and then click "Create Now".

If you have any problems creating the Collection, you can use the following URL to view a publicly-accessible version:
https://pathogen.watch/collection/eyev4yq jshc2-k-pneumoniae-st219-collection

Centre for Genomic Pathogen Surveillance

Exploring the Pathogenwatch Collection

View the Collection Once the phylogenetic tree has been computed, you can click "View Tree". Your Collection View should now look like the image on the right. Note the wide geographic distribution of public ST219 genomes. Locate the different tables that sit behind the timeline: Metadata, Typing, Stats, Antibiotics, AMR Determinants.	 Collection View of 259 ST219 genomes.
Locate your outbreak and historical genomes Locate your five uploaded samples in the phylogenetic tree. What can you conclude from their positions in the tree about their relatedness? 🔦 Tip. You could use the filter bar at the top to quickly find your uploaded genomes. Searching for the string, "contigs", which is found in all five isolate IDs, will bring up all five at the same time. 🔦 Tip. You should see that the five uploaded genomes all fall into the main dominant clade of the ST219 tree. You can click the parent node of this clade and then "View Subtree" to explore the clade with greater resolution.	 Identification of uploaded genomes via the filter bar.
Identify other public genomes from Nigeria Now let's clear all filtering by clicking the "X" at the right of the filter bar, and also right click on the white canvas of the tree panel and select "Redraw original tree". Use the lasso tool in the map panel to select all genomes from Nigeria. Locate the two public genomes from Nigeria in the phylogenetic tree. What can you infer about their relatedness to your contemporary outbreak isolates?	 Map panel filtering genomes from Nigeria.

Centre for Genomic Pathogen Surveillance

Tip. Uploaded genomes have circular tips in the phylogenetic tree while public genomes are shown with square tips.

Investigate AMR dynamics in ST219

Clear all filters again.

Now go to the "Antibiotics" table and click on the header "Cephalosporins (3rd gen.)".

This shows us that most genomes from the dominant ST219 clade carry a resistance marker for this class of antibiotics.

This is a typical pattern within *K. pneumoniae* - large clonal expansions are frequently associated with acquisition of resistance to key antibiotic classes.

Take a look at the other antibiotics in the table and observe any patterns.

ST219 tree showing predicted resistance to Cephalosporins (3rd gen.).

Identify key resistance markers in ST219

Take a look at the "AMR Determinants" table which shows you the specific resistance markers identified via Kleborate.

In *K. pneumoniae*, *bla*CTX-M-15 is the most common AMR gene conferring resistance to 3rd gen. cephalosporins.

Note also the acquisition of the *bla*NDM-1 carbapenemase by a sub-clade, as well as the truncation of the *ompK35* porin in a sub-clade of this (both associated with carbapenem resistance).

Cumulative resistance (in this case *bla*CTX-M-15 → *bla*NDM-1 → *ompK35* truncation) is another typical feature of MDR *K. pneumoniae* lineages.

Dominant clade (subtree) of ST219 tree showing the presence of *bla*NDM-1 in a sub-clade.

Centre for Genomic Pathogen Surveillance

Assess the distribution of virulence markers in ST219

Let's now look at the distribution of yersiniabactin, previously identified in our uploaded genomes, across the entire ST219 collection.

Clear all filters again and redraw the original tree if you have zoomed into any subtrees.

Click on the + button in the top filter bar, then choose the "Yersiniabactin" column to filter on. There are many different yersiniabactin variants, but we can type "ybt" in the filter bar (then hit enter) to encompass and select all of these.

The genomes will now be filtered in all panels to include only those that carry yersiniabactin.

Re-clear the filters and now try filtering the genomes by the presence of other virulence markers. Are any other virulence markers found among ST219 genomes?

ST219 tree highlighting genomes carrying any yersiniabactin locus.

Visualisation of the ST219 genome collection using Microreact

For the next part of our investigation, we will further explore the results obtained from Pathogenwatch using the Microreact tool, which allows for more customisable visualisations of the data. As illustrated in Exercise 1, all results from Pathogenwatch are downloadable including the phylogenetic tree (.nwk file) from the Collection View, together with associated metadata and genotypic data (.csv files). In a typical workflow, you can then collate the required information from the CSV files into a single worksheet and ensure that it meets the input requirements for Microreact (see https://docs.microreact.org/instructions/creating-a-microreact-project/metadata-column-types). This can be done simply in Excel, but if you find yourself repeating a similar process on multiple occasions, you may want to consider using the data-flo tool for designing and re-using automated workflows. More information on data-flo can be found here: https://www.data-flo.io/

To speed up the outbreak investigation on this occasion, we have already prepared the files for you to upload to Microreact, which include the ST219 phylogenetic tree (*.nwk) and an excel file (.xlsx) with collated information from the "Metadata", "Typing" and "Kleborate" CSV outputs of Pathogenwatch. These are available in the "Microreact data" of this folder: **https://drive.google.com/drive/folders/1XzXRVDPy70fboBAIa1PE0av6J7TY771o**

Centre for Genomic Pathogen Surveillance

Upload files to Microreact

Ensure that you are logged into Microreact (https://microreact.org/).

Navigate to the Upload page and upload the .nwk and .xlsx files provided.

During the upload process, check that both files have been uploaded and the detected file type is correct. Ensure that the ID column is set to "id" in the Metadata table and the Labels column in the Tree is set to "id".

If you have problems with the upload, an equivalent Microreact project can be accessed here: https://microreact.org/project/kpneumoniae-st 219 Feel free to save a copy of this project in your own account so that you can make edits.

💡 Tip - Once you have loaded a Microreact project into your internet browser, it is possible to explore this offline (apart from the map panel) although any updates to the project would require an online connection again.

| Add Files or URLs | ✕ |

File	File kind
ST219-metadata.xlsx ✕	Data (Excel or ODS Spe... ▾
ST219-tree.nwk ✕	Tree (Newick) ▾
Enter URL	⊕ ADD MORE FILES

CANCEL   CONTINUE

Data Table

Data File
ST219_metadata.xlsx

ID column *
id

Column	Data Type	Colours	Shapes
id	text	Categorical	
Run accession	text	Categorical	
Data source	text	Categorical	
Date	text	Categorical	
day	number	Categorical	
month	number	Categorical	
year	number	Categorical	

CONTINUE

Explore the Microreact dashboard

The initial dashboard should look like the image on the right hand side.

Save your project using the [💾] icon and give it a name.
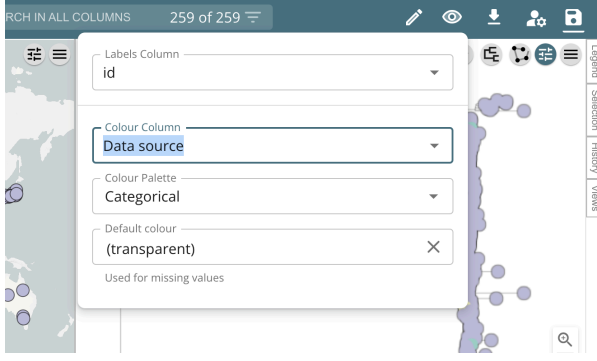
Explore the dashboard!

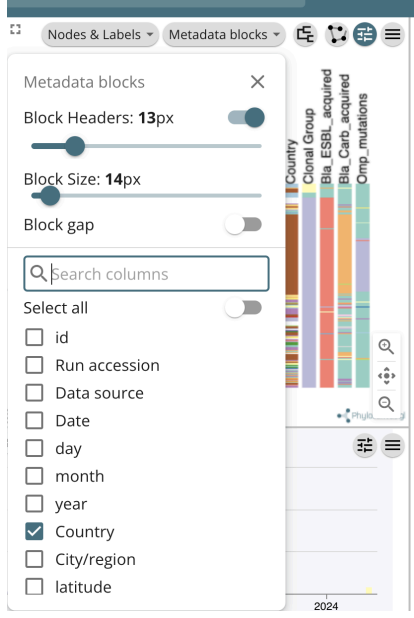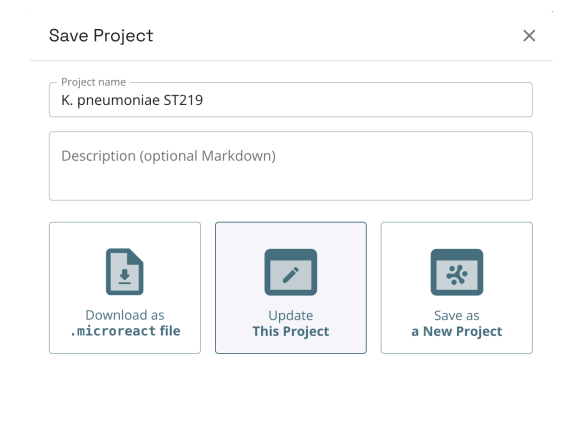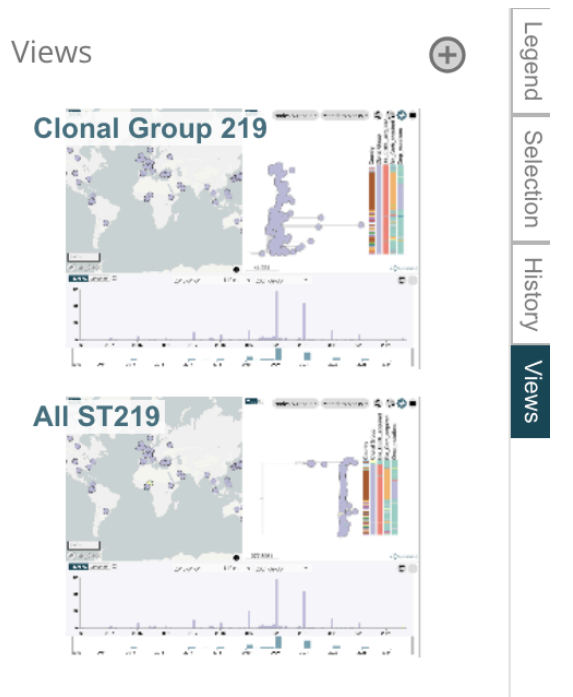The steps below can be used as a guide but feel free to follow your own intuition.

💡 Tip - Recall that you can toggle between the timeline and metadata. The metadata table provides excellent filtering options to explore.

Centre for Genomic
Pathogen Surveillance

Colour isolates by metadata columns

Recall that you can use the 👁 button in the top right to colour isolates across all panels, including the tips of the phylogenetic tree. Select the required column to colour by using the "Colour Column" drop-down.

Click "Legend" on the right hand panel to bring this up.

Colour by "Data source" to separate the isolates into "Public", "Outbreak" and "Historical".

Add metadata blocks

Add metadata blocks of your choice next to the tree to further facilitate your understanding of the ST219 genome collection.

Toggle on "Block Headers" to add the column names above the metadata blocks.

Questions of interest

Try addressing these questions to guide your exploration of the dashboard:

- What inferences can you draw from the country clustering?
- How many genomes have no country information?
- Where are the genomes that are most closely-related to your outbreak isolates from?
- How many genomes in the collection were sampled in the African continent?
- How many genomes in the collection are from 2022 onwards?
- Can you identify any significant clustering by isolation source?
- What is the diversity of K- and O- types across the ST219 collection?
- How many genomes carry an extended-spectrum beta-lactamase (ESBL) gene?
- How many different carbapenemase gene variants are found among ST219 genomes?
- What type of porin (*omp*) mutations are found among ST219 genomes?
- How many ST219 genomes carry colistin resistance mutations? What are these?

Centre for Genomic Pathogen Surveillance

Saving your Microreact dashboard When you are happy with changes to your customised dashboard, remember to click the [save icon] icon on the top right of the page, and click "Update this Project". Any annotations (e.g. metadata blocks) added prior to saving/updating the project will be maintained when reloading or sharing the URL.	Save Project ✕ Project name K. pneumoniae ST219 Description (optional Markdown) Download as .microreact file · Update This Project · Save as a New Project
Saving multiple "Views" As described in Exercise 2, you can also make and save multiple "Views" using the "Views" tab on the right hand side of the page. Remember to update your entire project each time you have made a new View that you are happy with. Try creating one View of the entire ST219 tree (with additional annotations/colours of your choice), and a second View of the major ST219 clade (Clonal Group 219).	Views ⊕ Clonal Group 219 All ST219 Legend | Selection | History | Views

Reporting back to the hospital

Congratulations! Your data analysis and interpretation are now complete.

Work with another team to summarise your findings into a brief report in the form of Powerpoint Slides to provide to the hospital.

You will present this to the class (max. 5 minutes) in our final session of this exercise.

Focus on details that are relevant to public health and clinical decision-making. You should also consider any limitations of the data and analytics used.

Centre for Genomic Pathogen Surveillance