Computational Practical 10: International Genomic Surveillance of Antimicrobial Resistance

Module Developers: Dr. Stanford Kwenda and Dr Francesc Coll I Cerezo

Table Of Contents

International Genomic Surveillance of Antimicrobial Resistance

Whole-genome sequencing (WGS) represents a step-change as it overcomes the lack of resolution of current typing methods applied to bacteria such as multi-locus sequences typing (MLST) or serotyping. Multiple publications confirmed the ability of WGS to define transmission dynamics of a single clone at different geographic and temporal scales (Baker et al. 2018; Harris et al. 2018; Snitkin et al. 2012; Eyre et al. 2013; Croucher and Didelot 2015). This has identified global and local transmission routes and, when combined with epidemiological data, can confirm or refute putative bacterial outbreaks. Most studies to date applied sequencing retrospectively in the context of suspected outbreaks wherein prior evidence of an existing outbreak triggered the use of WGS to confirm or refute recent transmission and, in some instances, helped identify the outbreak source and directly informed infection-control interventions. Prospective studies have also been conducted, predominantly in

intensive care units, providing a more typical view of transmission dynamics (Tong et al. 2015; Coll et al. 2017). WGS can also potentially be used for national and local surveillance of antibiotic resistant lineages and comprehensive genomic databases are being built to provide the context that would allow robust epidemiological interferences (Reuter et al. 2016; Aanensen et al. 2016)

Genomic epidemiology is changing the practice of surveillance and outbreak investigation of bacterial pathogens, and it is becoming a useful tool to understand the spread of multidrug resistant pathogens. In this practical you will visualize and interpret the datasets of key studies published on the epidemiology of methicillin-resistant *Staphylococcus aureus* and *Vibrio cholera* at different geographical scales (e.g., within individual hospital wards and internationally). We will make use of Microreact (https://Microreact.org/), a web application that provides an interactive visualization of datasets via phylogenetic trees, maps, timelines, and tables. In the Group activity session at the end of today we will also use Microreact and the epidemiological data collected using Epicollect to draw conclusions on the origin of the extensively Drug-Resistant (XDR) *S. typhi.*

10.1 Genomic history of the seventh pandemic of cholera in Africa

10.1.1 Introduction

Seven cholera pandemics have been recorded since 1817 (Harris *et al*. 2012). The seventh cholera pandemic is the seventh major outbreak of cholera caused by *Vibrio cholerae* O1 biotype El Tor. It occurred from the years 1961 to the 1970s and has continued (though much diminished) to the present. Africa is the continent most affected by the current pandemic, although the origin and propagation routes of the disease remain undefined. The seventh cholera pandemic (7P) began in 1961 in Indonesia, before spreading globally, in particular to South Asia (1963), Africa (1970), Latin America (1991), and the Caribbean (Haiti) (2010).

10.1.2 Available data

We will use published genomic data from 1070 *Vibrio cholerae* O1 isolates, across 45 African countries and over a 49-year period, to identify the origin(s) of the epidemics and the number of introductions into Africa (Weill *et al*. 2017). We will particularly focus on multidrug-resistant sub-lineages. We will make use of a maximum likelihood phylogenetic tree of 1070 *Vibrio cholerae* isolates (Weill2017_tree.nwk) and a metadata file (Weill2017_metadata.csv) with information on the country of origin, year of isolation, antibiotic resistance phenotypes and the presence of antibiotic resistance genetic determinants.

10.1.3 Creating a Microreact project

Start by opening up a new window in Firefox and typing https://Microreact.org/ in the address bar. Click on “Upload” as shown in Figure 1 to create a new project.

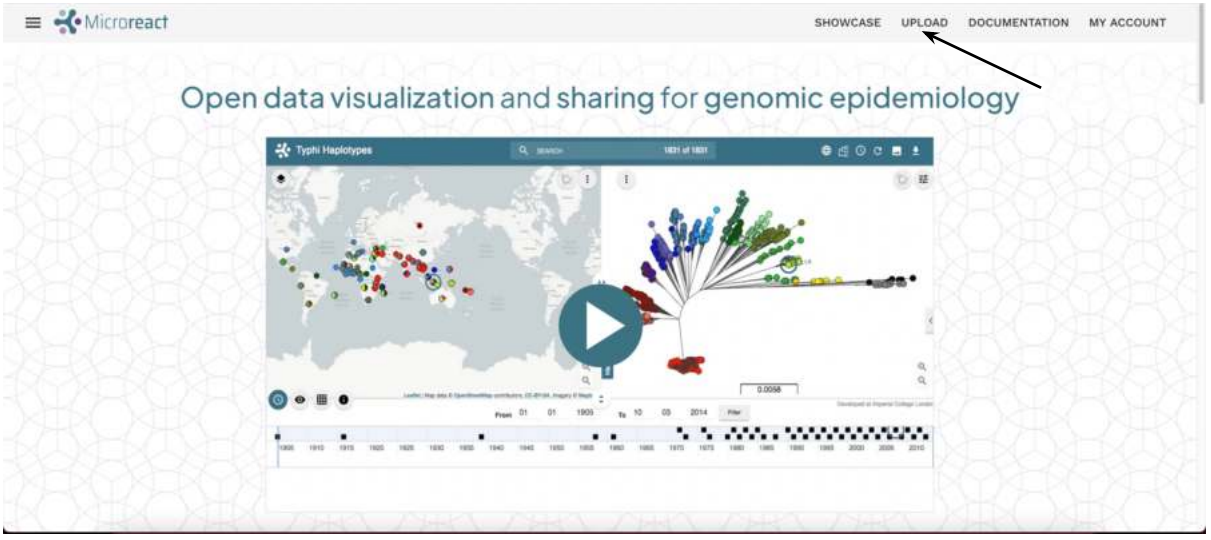

Figure 1 Microreact home page

Drag and drop files “Weill2017_metadata.csv” and “Weill2017_tree.nwk” from your file browser onto your Internet browser (Figure 2).

Figure 2 Microreact upload page

Once the tree and metadata files are loaded you will be directed to a new window where files will be automatically detected as Data (CSV or TSV) file (Weill2017_metadata.csv) and Tree (Newick) file (Weill2017_tree.nwk). In this new window click on ‘Continue’. In the next window (Figure 3), make sure column ‘id’ is selected as the ‘ID column’ and then click on ‘Continue’.

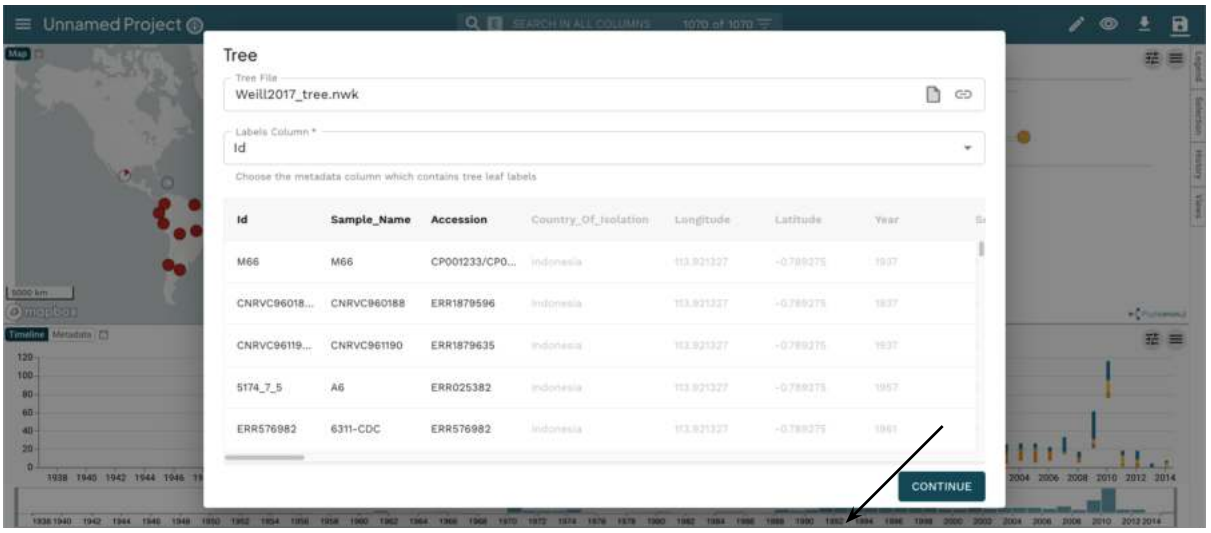

Figure 3 Data Table window in Microreact

Once these forms are completed your data will be utilized to create a Microreact project. You should now have a view similar to the one in Figure 4. You should see a Map, Tree and Timeline panels. You can use click-drag-zoom to navigate both the tree and the map.

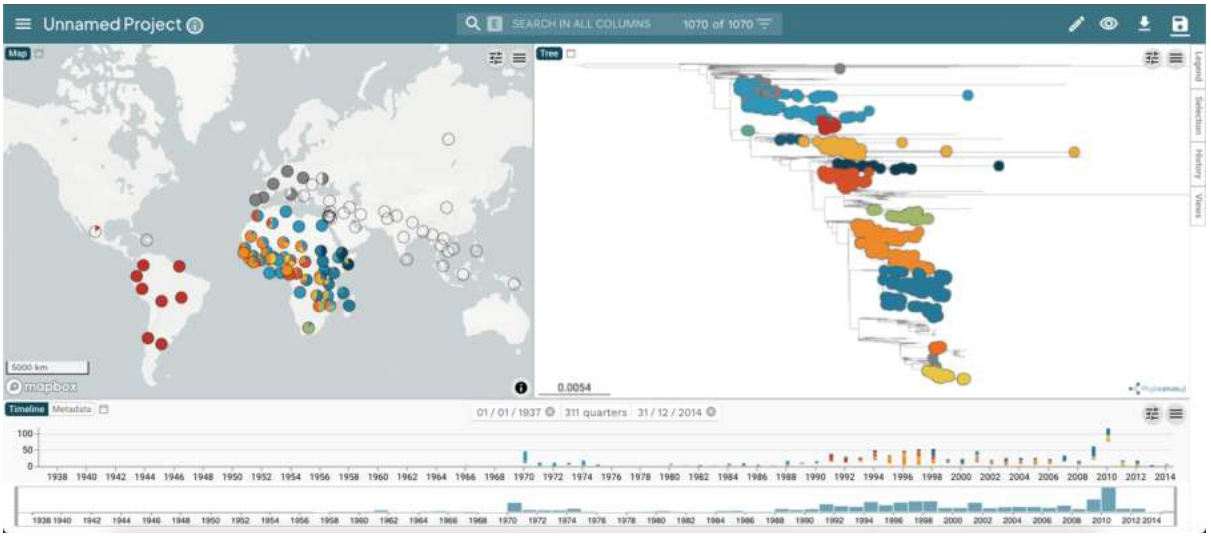

Figure 4 Microreact project page for Weill2017 dataset

10.1.4 Origin of the seventh cholera pandemic in Africa

We will use the phylogeny and the map to identify the origin of several transmission events of the seventh cholera pandemic into Africa.

Click on the 'Show controls' button (arrow 1 in Figure 5) to then select the Hierarchical layout (arrow 2). You may want to decrease the size of 'Leaf nodes'

using the corresponding toggle button  (arrow 3). You can additionally visualise "Internal Labels" (arrow 4), that is, the IDs of internal nodes in the phylogeny.

Figure 5 Displaying transmission events in Microreact

Click on the 'Labels, Colours, and Shapes'  icon and select 'Transmission_Event' both under the lists 'Labels Column' and 'Colour Column' (arrows 1 and 2 in Figure 6) to display the labels and colours of transmission events respectively. Note there are twelve transmission events highlighted on the tree with different colours.

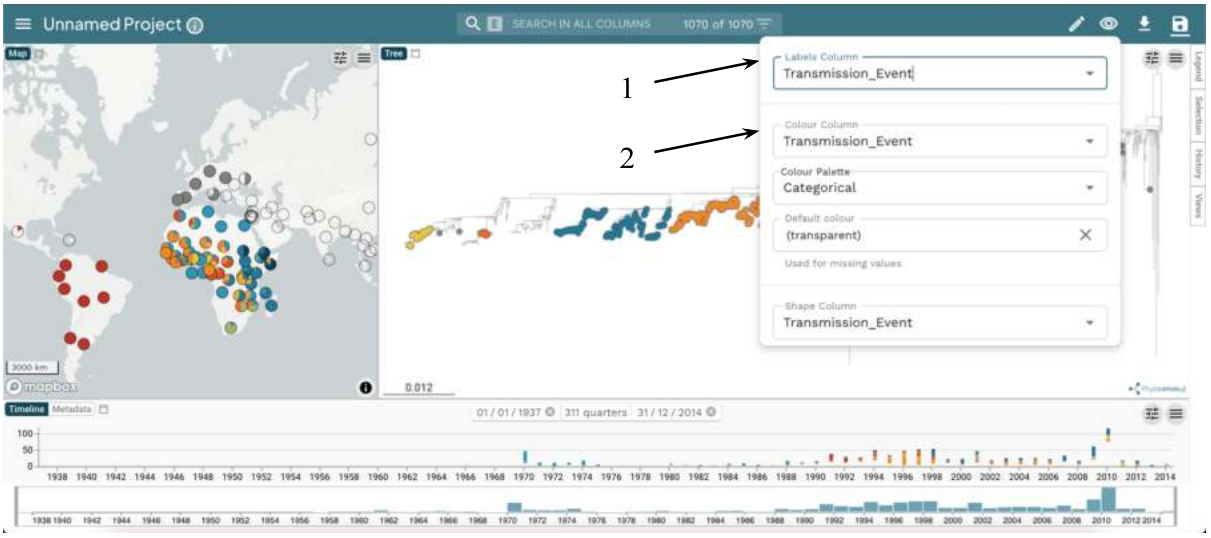

Figure 6 Labelling and colour-coding transmission events on the tree

Search for the two samples from Ethiopia belonging to sub-lineage T3 (arrow 1 in Figure 7) using the Search box at the top of your window (arrow 2).

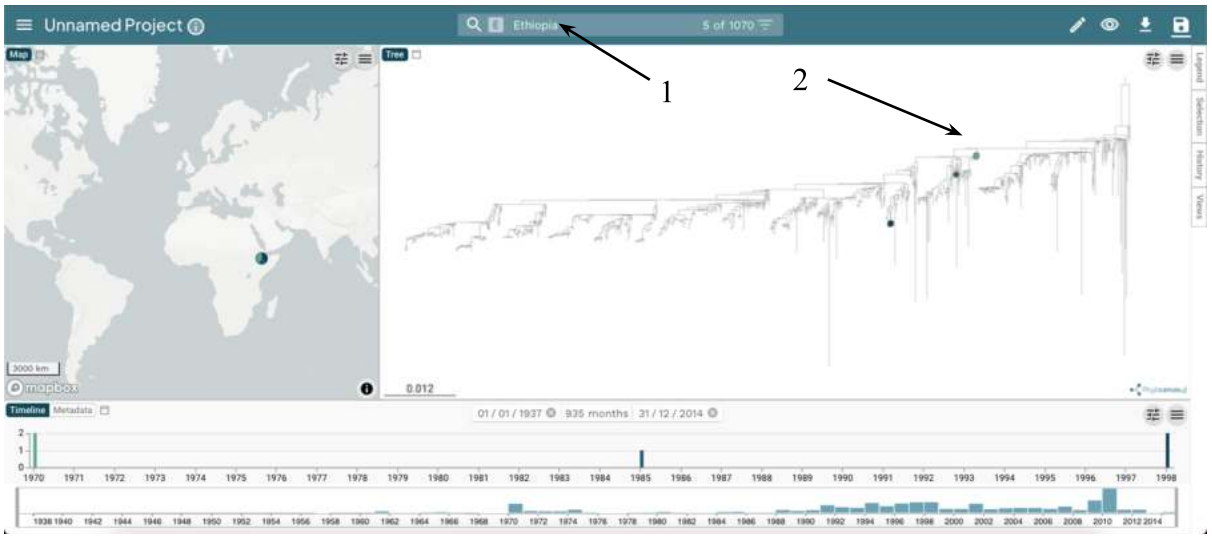

Figure 7 Location of sub-lineage T3 Ethiopian samples on the tree

Make sure you keep labels displayed by clicking on the toggle button next to 'Leaf Labels' and 'Internal Labels' under the 'Nodes & Labels' window (arrows 1-4 in Figure 8) and zoom in enough to make labels visible. Increase the label size if necessary (arrow 3).

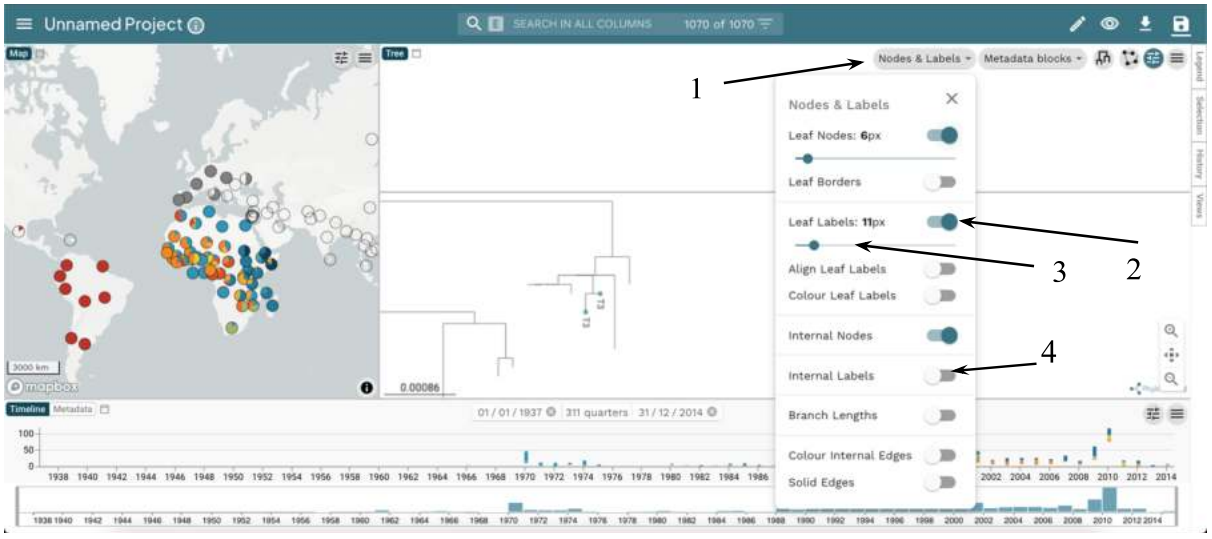

Figure 8 Keeping leaf labels visible on the tree

Select 'Country_Of_Isolation' under 'Labels Column' (arrow 1 in Figure 9) to show what country samples came from. Zoom in to take a closer look at sub-lineage T3 in the Tree panel and click on the internal node225 (arrow 2) to show T3 and its neighbouring samples only on the map. You will observe that isolates from Ethiopia are closely related to isolates from the same year (1970) from Jordan and Israel which suggests an importation of this sub-lineage from the Middle East into Ethiopia. You may want to click on the 'Metadata' panel (row 3) to display isolate metadata, including 'Year' of isolation. If can **double-click** on a specific area of the tree to centre and zoom in your view on this area. Remove any text written in the Search box at the top of your window to display the IDs of all other samples.

Figure 9 Origin of the Ethiopian T3 samples

Zoom out to visualize the whole tree again. Douple-click on a white space on the Tree panel to unselect chosen samples. Click on the 'Legend' button at the top-right corner of your screen (arrow 1 in Figure 10) to display the colour legend. Now click on transmission event 'T2' (arrow 2) to visualise samples labelled with this transmission event on all three panels.

Figure 10 Location of sub-lineage T2 on the tree

Zoom in and click on the internal node 192 (arrow, Figure 11) to highlight only T2 samples on the tree and map. Make sure 'Country_Of_Isolation' is selected under 'Labels column' and that the 'Leaf labels' toggle button is on to display country labels on the tree. This internal node is the most recent common ancestor of all Cholera isolates from the Latin America epidemic in the 1990s.

Figure 11 Close view of sub-lineage T2 on the tree

Select the most immediate ancestral nodes to T2 (i.e. internal nodes 186 and 172) to highlight the location of neighbouring and more ancestral isolates (arrows 1 and 2 in Figure 12). Based on the location of these samples on the map, the authors suggested the Latin American epidemic (T2) resulted from importations from West Africa.

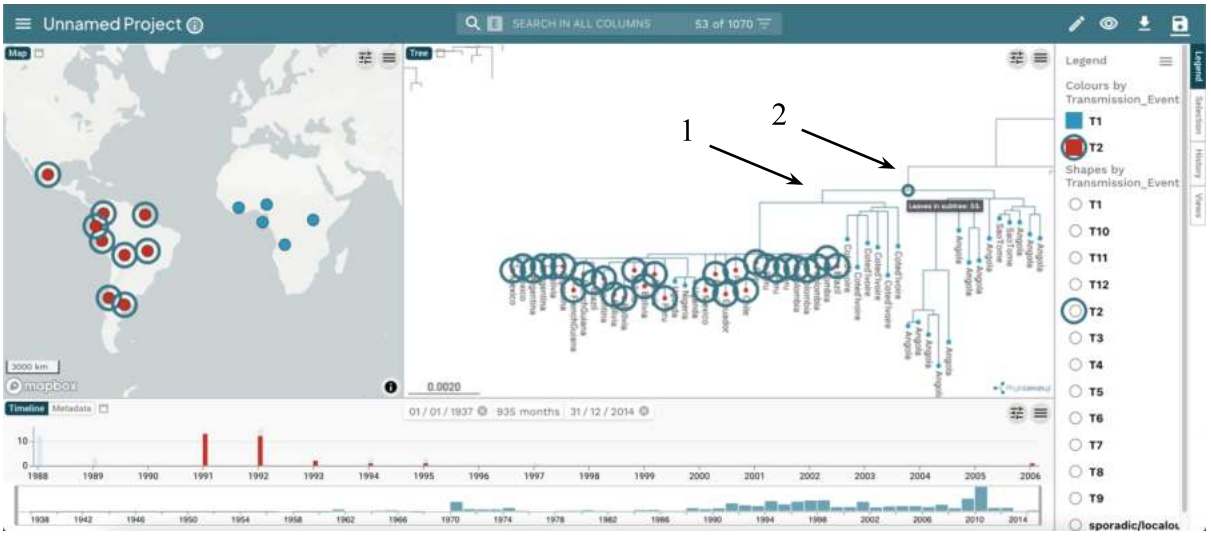

Figure 12 Origin of the Latin American epidemic

The authors noted a recurrent pattern of transmission for cholera epidemics in Africa. Separate sub-lineages of the seventh pandemic (labelled as different transmission events) from Asia were repeatedly introduced into two main regions: West Africa and East/Southern Africa. Epidemic waves then propagated regionally, in some instances spreading to Central Africa, over periods of a few years to 28 years.

Examples of importations to West Africa from Asia include sub-lineages T7 (node 365), T9 (node 603) and T12 (node 1038) (Figure 13).

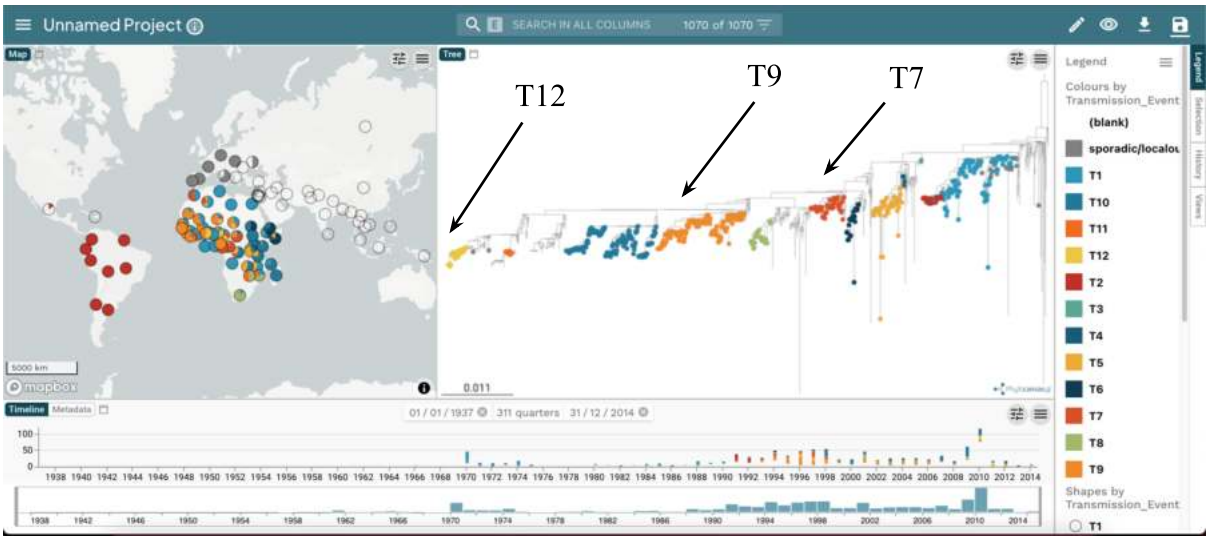

Figure 13 Cholera importations into West Africa

Zoom in and click on the internal node of one of these sub-lineages (e.g., T7, node 365) to visualise the location of these samples on the map (Figure 14). Based on the location and year of isolation of neighbouring and more ancestral isolates, what is the most plausible source of this importation?

Figure 14 Origin of the T7 sub-lineage imported into West Africa

Example of importations to East/Southern Africa include sub-lineages T4 (node 259), T5 (node 262), T6 (node 339), T8 (node 502), T10 (node 700) and T11 (node 959) shown in Figure 15.

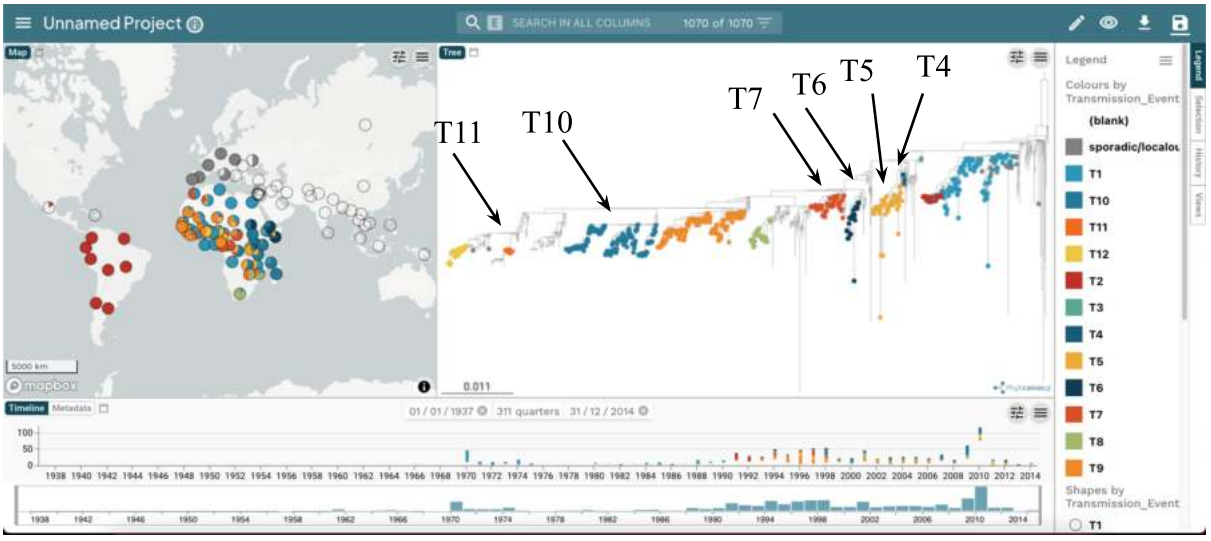

Figure 15 Cholera importations to East/Southern Africa

Zoom in and click on the internal node of one of these sub-lineages (e.g. T10, node 700) to visualise the location of these samples on the map (Figure 16).

Figure 16 T10 sub-lineage is restricted to East Africa

The authors found that multiple sub-lineages (T1, T3, T4, T5, T6, T8, and, possibly, T9) originated from South or East Asia and were circulating in the Middle East before importation into West or East Africa.

10.1.5 Epidemiology of multi-drug resistant *V. cholerae* in Africa

The authors noted that *V. cholerae* African isolates became increasingly resistant to antibiotics over time. The figure below (Supplementary Figure 9 in the original publication) shows an increase in the average number of antimicrobial resistant genes (ARGs) per isolate over time (panel A). The percentage of isolates positive for the IncA/C plasmid and the SXT/R391 genomic island are shown in panel B and C. From these results, the authors hypothesize that the increasing number of ARGs in African isolates (panel A) could be the result of an increasing prevalence of the SXT/R391 genomic island (panel C).

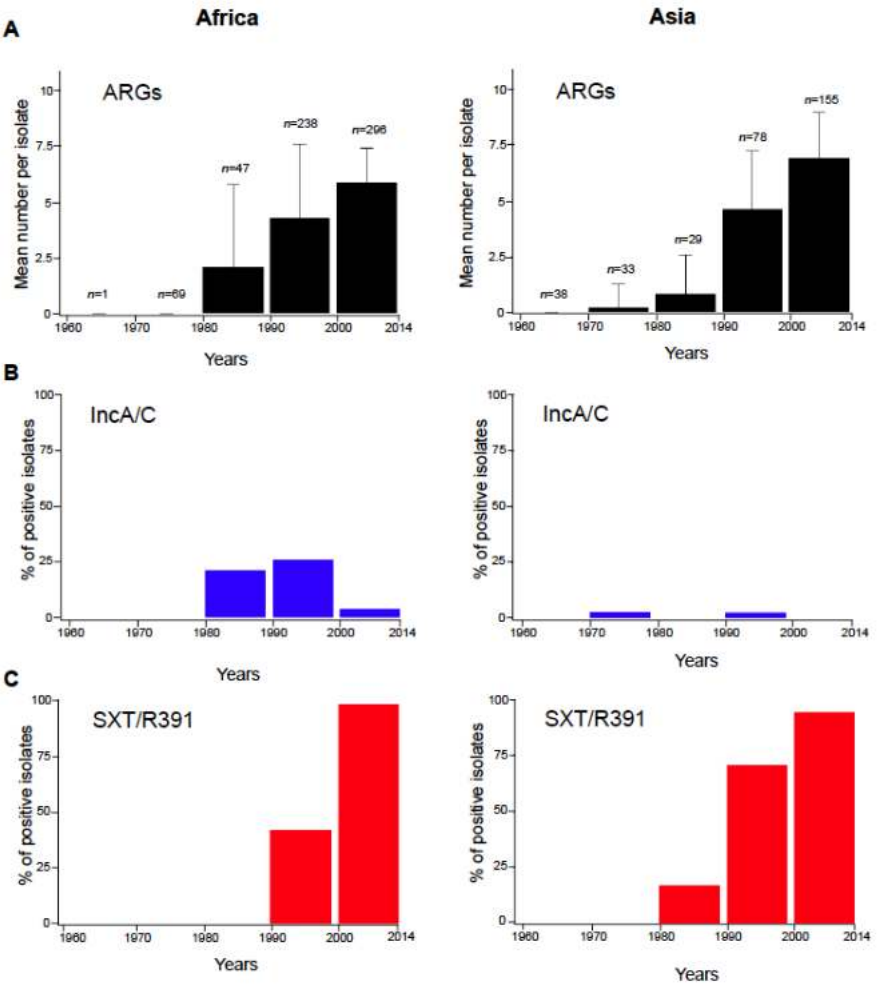

Fig. S9

Evolution of antibiotic resistance determinants over time in 7P *V. cholerae* El Tor isolates from Africa and Asia. Panel "A" shows the change in the mean number of antibiotic resistance genes (ARGs) per isolate over time. The number of studied isolates (*n*) is indicated. Panels "B" and "C" show the percentage of isolates containing an IncA/C plasmid or a SXT genomic island, respectively. The analysis was performed for various time periods: 1960-1969, 1970-1979, 1980-1989, 1990-1999, and 2000-2014.

The SXT-R371 genomic island is a ~100-kb integrative and conjugative element (ICE) encoding resistance to multiple antimicrobials in *Vibrio cholerae*. ICEs are able to transfer by conjugation and integrate into and replicate as part of the bacterial chromosome. Five main groups (i.e. versions) of the SXT-R371 genomic island have been described, being ICEVchInd5 the most frequently represented (Spagnoletti *et al*. 2014). Each group possesses a different complement of antimicrobial resistant genes (ARGs).

We will use Microreact to visualise what sub-lineages carry the SXT-R371 genomic island and the temporal spread of these. Colour-code samples by Transmission_Event (arrow 1 in Figure 17) and show the SXT_R391 label on the tree (arrow 2 in Figure 17).

Figure 17 Labelling samples based on presence of SXT-R391 genomic island

Search for 'ICEV' in the Search Box at the top of your window. This will highlight what samples harbour the SXT-R371 genomic island.

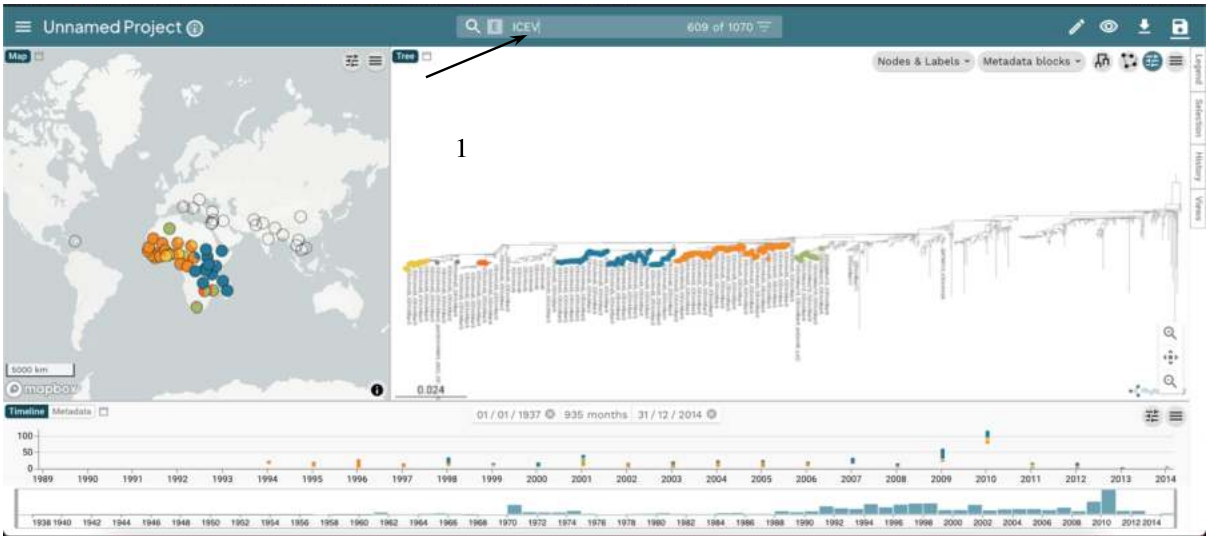

Figure 18 Visualising samples with the SXT-R391 genomic island

You should be able to distinguish what lineages carry the SXT-R371 genomic island (at the left-hand side of the tree) from those that do not (at the right-hand side of the tree). A close inspection will reveal that sub-lineages T8, T9, T10, T11 and T12 all carry this genomic island.

A closer look at these sub-lineages (e.g., T9, node 603) shows how the genomic island has remained very stable within these lineages, i.e., it has rarely been lost. This is characteristic of chromosomally integrated conjugative elements which, once integrated in the chromosome, are inherited vertically.

SXT-R371 typically harbours *floR* (which confers resistance to phenicols), *strA/strB* (aminoglycosides), *sul2* (sulfamethoxazole) and *dfrA1* (trimethoprim), although the number of genes may vary. You can identify what samples have these genes by searching for their name.

Search for *floR*, *strA_strB*, *sul2*, *dfrA1* to identify what samples contain each of these genes (arrow in Figure 19). Note that samples harbouring the SXT-R371 genomic island do also contain these genes.

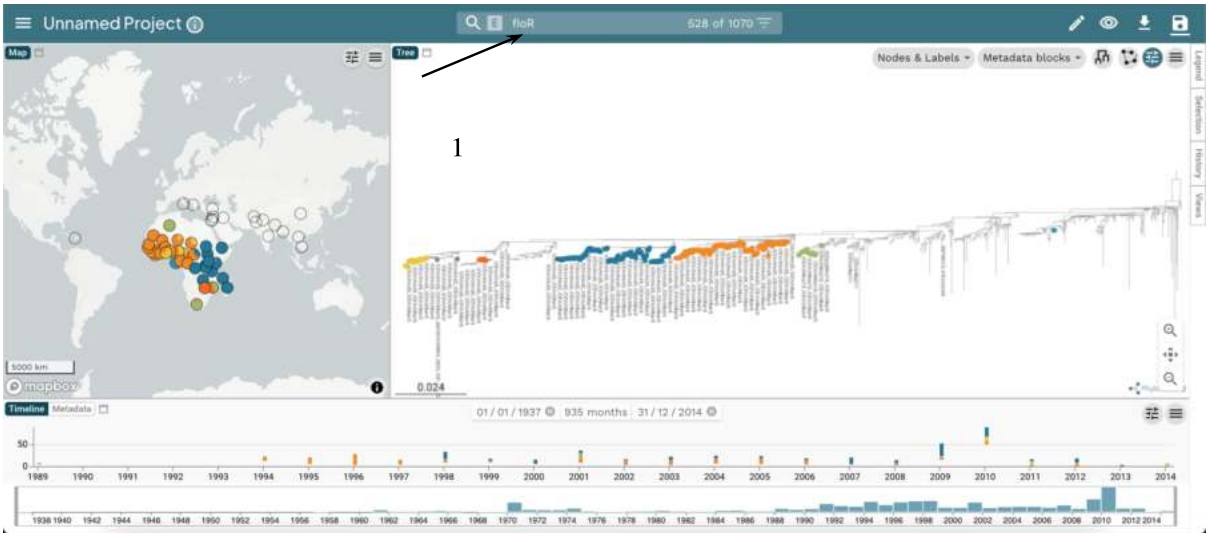

Figure 19 Displaying samples containing different AMR genes on the tree

We will now visualise the temporal and geographical spread of different sub-lineages. First, make sure you can view the whole tree and the Search Box is empty, that is, all nodes in the tree are shown. You may want to turn the ‘Leaf Labels’ toggle button off to hide tip labels on the tree. You should now be able to see the twelve transmission events coloured-coded on the tree and map with different colours (Figure 20).

Figure 20 Choosing a time frame on the Timeline panel

On the Timeline panel (arrow 1 in Figure 20), select ‘Quarter’ (arrow 2) under ‘Unit’, set the play speed to ‘1 quarter per 1 second’ under ‘Speed’ (arrow 3) and, lastly, click on the ‘Play timeline’ button (arrow 4).

This will create a movie where isolates will be shown both on the Tree and Map panels according to their year of isolation. The first isolates to appear correspond to the earliest (1960s) and more ancestral isolates collected from Central and South-east Asia, the origin of the seventh pandemic lineage (circled on the tree in Figure 21).

Figure 21 The earliest isolates from the seventh pandemic originated in Central and South-East Asia

After this, isolates from the T1 sub-lineage appeared on the 1970s mostly clustered in North Africa, West Africa (in blue) and Europe (in grey) (circled on the tree in

Figure 22). This sub-lineage continued to spread throughout Southern Africa during the 1970s, and it circulated in this region until the early 1990s.

Figure 22 Sub-lineage T1 emerged in the 1970s

Isolates from the Latin American epidemic (sub-lineage T2, in red) emerged in the 1990s (Figure 23).

Figure 23 Sub-lineage T2 emerged in the 1990s

The antibiotic-resistant sub-lineages in Africa (T8, T9 and T10) started emerging in the 1990s, followed by sub-lineages T11 and T12, all five accounting for most of the *V. cholerae* cases sampled in the 2000s and 2010s. If you want to highlight where the most recently sampled *V. cholerae* isolates are located in the tree, set the 'From' date bar to 2000-01-01 (arrow 1 in Figure 24) and the 'To' bar to 2014-01-01 (arrow 2) in the Timeline panel.

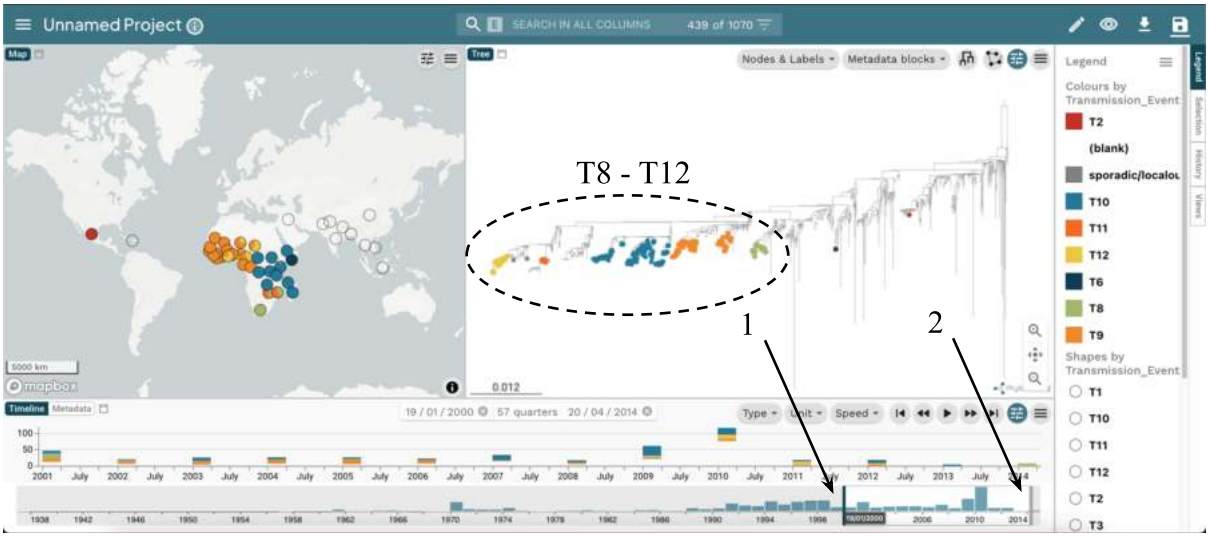

Figure 24 The most recent cholera cases are attributable to sub-lineages T8 to T12

The recent increase in antibiotic resistance among *V. cholerae* isolates in Africa have been largely driven by resistance determinants (the SXT/R391 genome island) already present within the genomes of imported sub-lineages (T8-T12) from South Asia, rather than by the independent local acquisition of resistance within this region.

Local Genomic Surveillance of Antimicrobial Resistance

Whole-genome sequencing (WGS) overcomes the lack of resolution of current bacterial typing methods. Several recent publications have confirmed the ability of WGS to define transmission dynamics of a single clone at different geographic and temporal scales (Alam *et al.* 2015; Aanensen *et al.* 2016; Harris *et al.* 2010, 2013; Köser *et al.* 2012). This has identified global and local transmission routes and, when combined with epidemiological data, can confirm, or refute putative outbreaks.

Genomic epidemiology is changing the practice of surveillance and outbreak investigation of nosocomial and other pathogens. Genomic surveillance is becoming increasingly high profile as a mechanism to understand the spread of multidrug resistant pathogens.

Within this practical you will visualize and interpret the datasets of key studies published on the epidemiology of methicillin-resistant *Staphylococcus aureus* (MRSA). We will focus on this bacterium to exemplify different applications of WGS. We will make use of Microreact (https://Microreact.org/), a web application that

provides an interactive visualization of datasets via phylogenetic trees, maps, timelines and tables.

10.2 Hospital transmission of MRSA

10.2.1 Introduction

Whole-genome sequencing of MRSA has been used to define phylogeny and transmission in healthcare settings. Here, we will investigate the genetic diversity of MRSA within wards and individual patients in a hospital in Thailand and identify the source of numerous transmission events. All patients in two intensive care units were screened for MRSA carriage over a 3-months period. All MRSA belonged to multi-locus sequence type 239 (ST 239). A total of 79 isolates from 46 patients and five members of staff, including the first MRSA-positive screen isolates and up to two repeat isolates where available (Tong *et al.* 2015).

10.2.2 Available data

In this section, we will make use of a maximum likelihood phylogenetic tree of the 171 ST239 isolates (Tong2015_tree.nwk) and a metadata file with information on the patient Id, ward, time of isolation and antibiotic resistance phenotypes (Tong2015_metadata.csv).

10.2.3 Creating a Microreact project

Start by opening up a new window in Firefox and typing https://Microreact.org/ in the address bar. Click on “Upload” as shown in Figure 1 to create a new project.

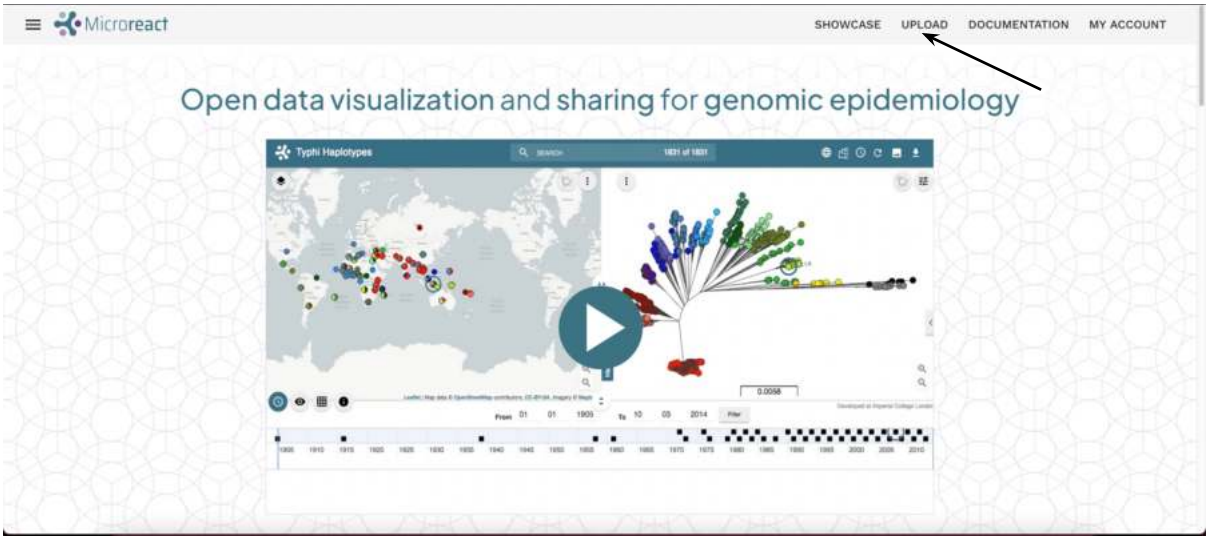

Figure 1 Microreact home page

Drag and drop files “Tong2015_metadata.csv” and “Tong2015_tree.nwk” from your file browser onto your Internet browser (Figure 2).

Figure 2 Microreact upload page

Once the tree and metadata files are loaded you will be directed to a new window where files will be automatically detected as Data (CSV or TSV) file (Tong2015_metadata.csv) and Tree (Newick) file (Tong2015_tree.nwk). In this new window click on 'Continue'. In the next window (Figure 3), make sure column 'id' is selected as the 'ID column' and then click on 'Continue'.

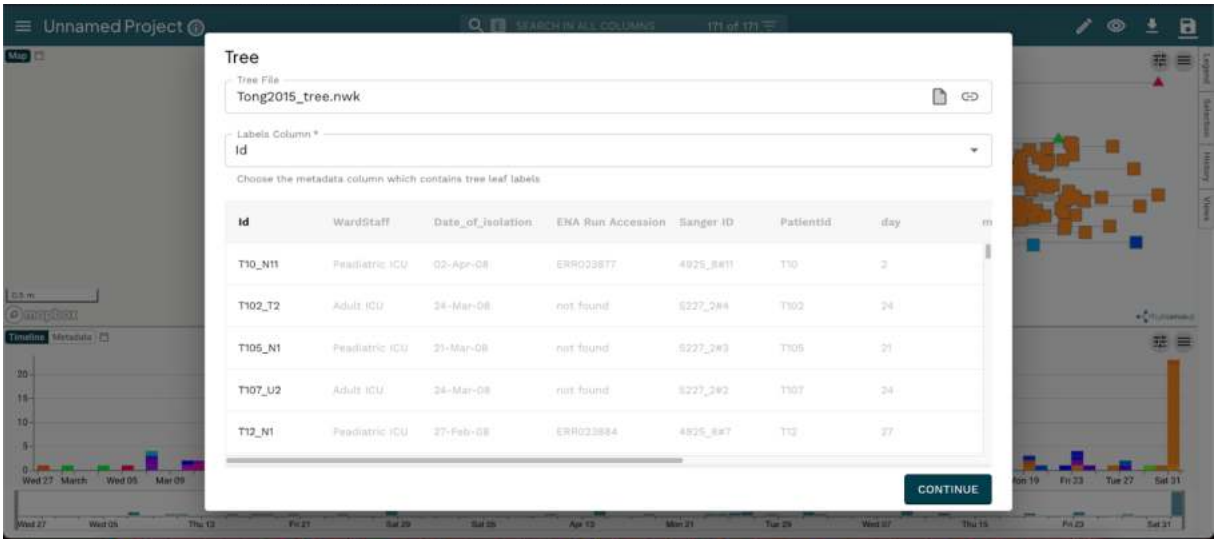

Figure 3 Data Table window in Microreact

As the country of origin is irrelevant in this investigation (because all isolates come from the same country and hospital) click on the 'add or edit panels' button  at the top-right corner of the screen (arrow 1 in Figure 4), and then on 'Edit Existing Panels' (arrow 2).

1 ⟶

Figure 4 Editing existing panels in Microreact

In the newly opened window (Figure 5), click on ‘Map’ and ‘Remove Map’ to hide the Map panel and allow more space for the phylogenic tree.

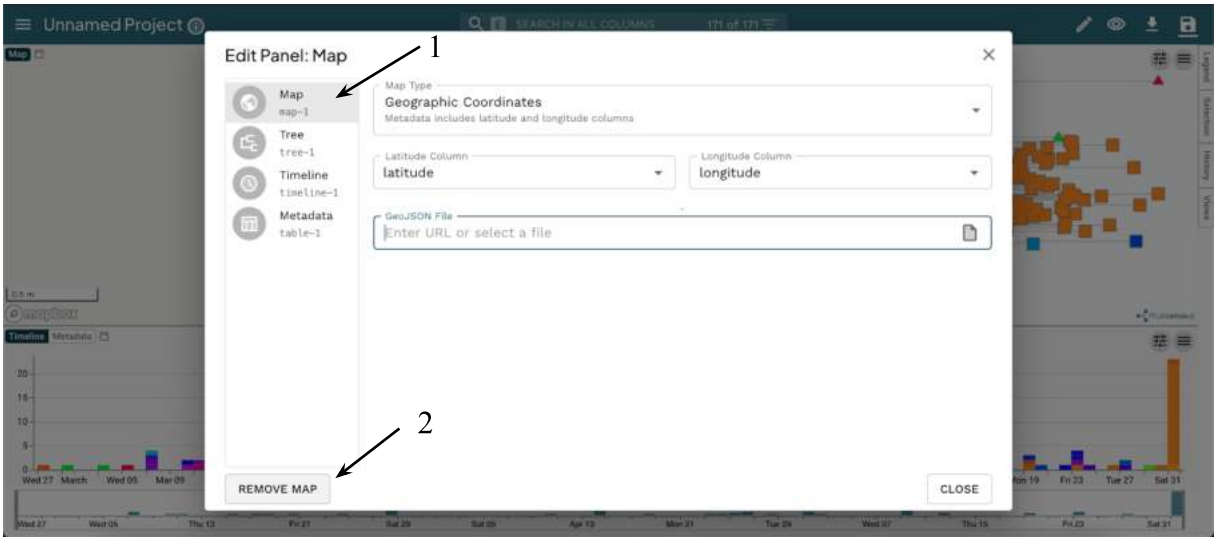

Figure 5 Removing Map panel in Microreact

Click on the configuration button  (arrow in Figure 6) and under ‘Tree Type’ select ‘Hierarchical’  .

Figure 6 Changing phylogenetic tree layout to hierarchical

1

Note that colours in the tips of the tree represent different hosts, whereas shapes represent the type of host: adult patients are represented by squares, paediatric patients as triangles and nurses as circles (Figure 6).

Do you spot any clustering by ward? in other words, is there any evidence of ward-specific clades? Is there evidence of intra-ward transmission and/or inter-ward transmission?

There is evidence of intra-ward transmission revealed by the phylogenetic clustering of paediatric patients in one clade (defined by node 22, dotted red rectangle in Figure 6) and adult patients in another separate clade (defined by node 54, dotted blue rectangle in Figure 6). There is also evidence of inter-ward transmission revealed by clades containing both paediatric and adult patients (e.g., those defined by nodes 165 and 2).

Click on the ‘Legend’ button at the right-hand side of the window (arrow 1 in Figure 7). A list of patient Ids will be displayed with a colour legend. Click on different patient Ids to highlight all isolates from the same patient in the tree and timeline panels as shown in Figure 7 for patient T126 (arrow 2).

Figure 7 Isolates from patient T126 are highlighted in the tree and timeline.

Now, we are going to investigate cases of patient-to-patient transmission within the same ward. The most obvious example is that of patient T126 who had multiple isolates of their bacterial clone sequenced over time as highlighted in Figure 7. The fact that all isolates from T126 belong to the same monophyletic clade with multiple variants (i.e., branches originating from the same internal node) points to the presence of a genetic 'cloud of diversity' as opposed to colonisation with multiple clones (isolates present in different phylogenetic clades).

Try to identify patients with multiple isolates in the tree. Specifically, compare patients T20, T35, T99, T126, T188 and T234. From these, try to differentiate patient colonised with multiple bacterial clones (mixed colonisation) from patients with variants of the same clone (also refer as to 'cloud of diversity'). Patients T126, T188 or T20 have evidence of a cloud of diversity whereas T234, T99 or T35 have evidence of multiple clones (mixed colonisation).

Comparison of isolates from patient T126 with isolates from other patients within the same phylogenetic clade (i.e., T327, T192, T301, T234, T303, T271, T335, T232) provides evidence that T126 acted as the donor for MRSA transmission to these other patients, who were colonized with isolates derived from variants within the T126 cloud. In other words, the fact that other patients are nested within the diversity of patient T126 in the tree supports the hypothesis that T126 acted as a source.

The  icon on the top-right corner of the window (arrow 1 in Figure 8) will allow you to colour and label the tips of the tree based on different metadata fields. Here, in the drop-down list under 'Labels Column' select 'PatientId' to label the tips (or leaves) based on this field. Tip nodes are already colour-coded by PatientId too.

Figure 8 Labelling tips with patient ids

Next, browse to the part of the tree where the isolates of most paediatric case cluster. Make sure tip labels (i.e. patient ids) are visible by clicking on the 'Leaf Labels' toggle button (arrow 2, Figure 9) under the 'Nodes & Labels' window (arrow 1). Highlight isolates from the following paediatric patients with more than one isolate in the tree: T12, T183 and T188 (arrow 3).

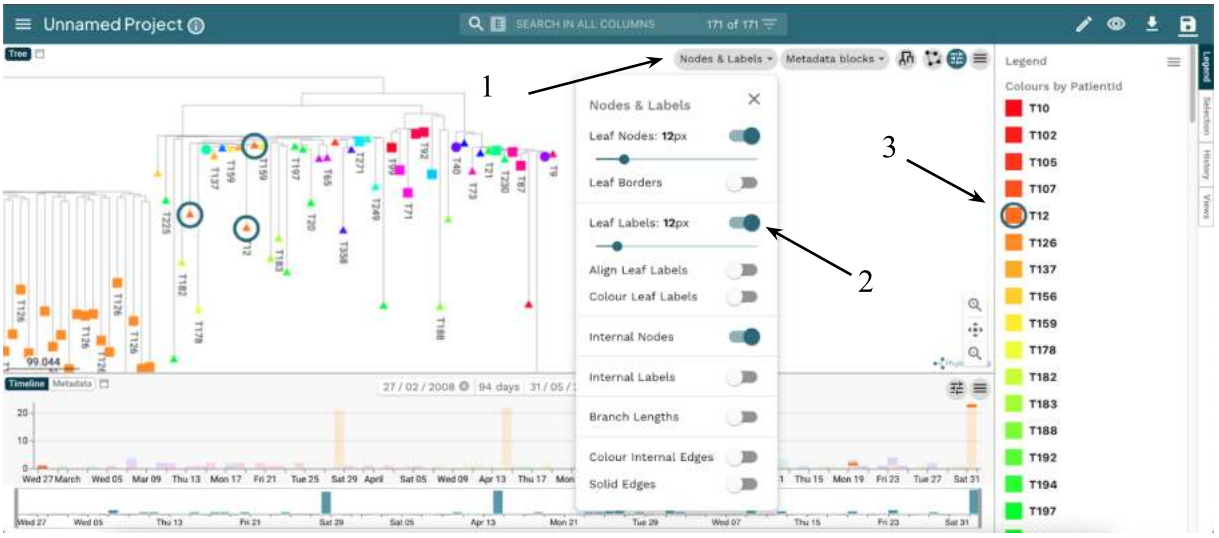

Figure 9 Investigation intra-patient diversity in paediatric patients

Is there any evidence of any of these patients acting as a donor in a transmission event? A group of six patients seem to have acquired their strain from patient T12 (see clade defined by node 40). Patients T183 and T188 do not have evidence of onward transmission.

10.3 Tracing the origin and spread of a methicillin-resistant *Staphylococcus aureus* epidemic clone (optional)

10.3.1 Introduction

The widespread use of antibiotics in association with high-density clinical care has driven the emergence of drug-resistant bacteria that are adapted to thrive in hospitalized patients. Of particular concern are globally disseminated methicillin-resistant *Staphylococcus aureus* (MRSA) clones that cause outbreaks and epidemics associated with health care. The most rapidly spreading and tenacious health-care-associated clone in Europe currently is EMRSA-15. In this section we will investigate the origin and spread of this clone using the genome sequences of 193 *S. aureus* isolates from sequence type (ST) 22 (Holden et al. 2013).

10.3.2 Available data

In this section we will make use of a maximum likelihood phylogenetic tree of the 193 ST22 isolates (Holden2013_tree.nwk) and a metadata file with information on the country of origin, time of isolation, antibiotic resistance phenotypes and the presence of antibiotic resistance genetic determinants (Holden2013_metadata.csv).

10.3.3 Creating a Microreact project

Start by opening a new window in Firefox and typing https://Microreact.org in the address bar. Click on “Upload” as shown in Figure 1 to create a new project.

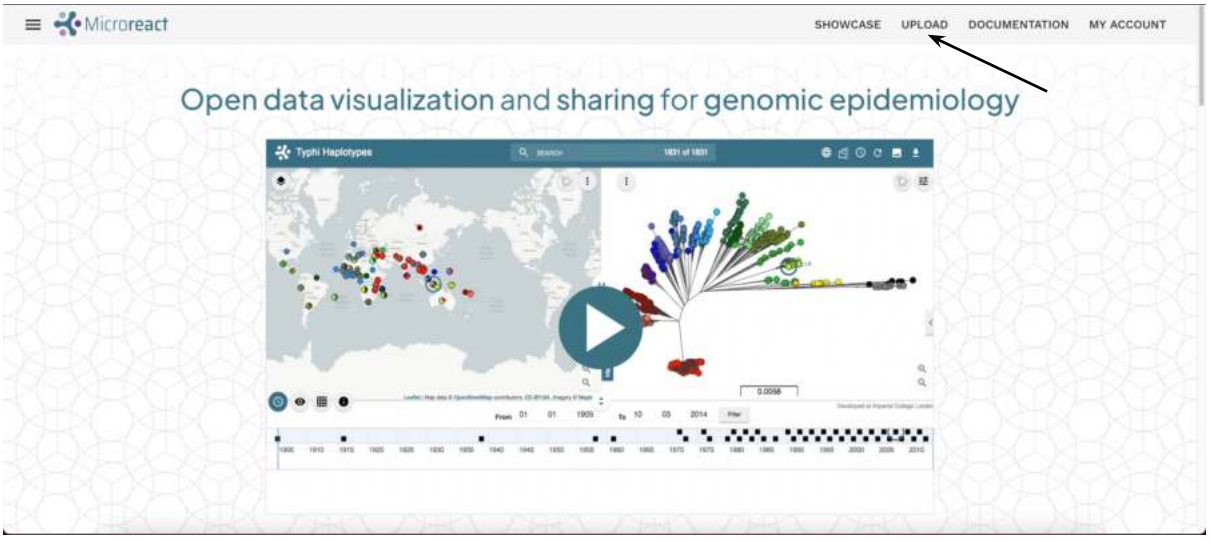

Figure 1 Microreact home page

Drag and drop files “Holden2013_metadata.csv” and “Holden2013_tree.nwk” from your file browser onto your Internet browser (Figure 2).

Figure 2 Microreact upload page

Once the tree and metadata files are loaded you will be directed to a new window where files will be automatically detected as Data (CSV or TSV) file (Holden2013_metadata.csv) and Tree (Newick) file (Holden2013_tree.nwk). In this new window click on 'Continue'. In the next window (Figure 3), make sure column 'id' is selected as the 'ID column' and then click on 'Continue'.

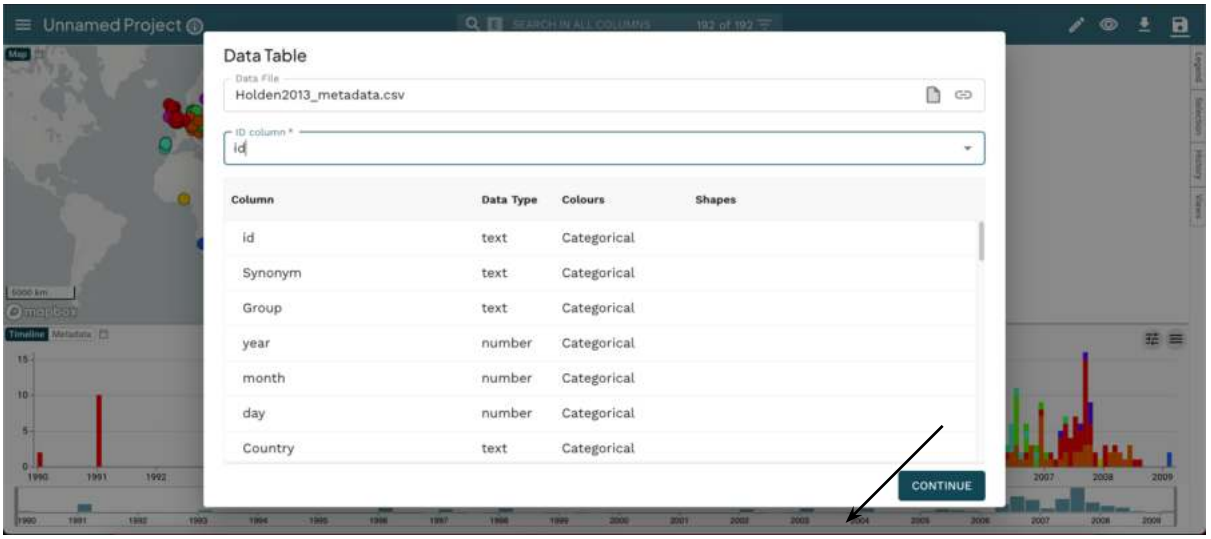

Figure 3 Data Table window in Microreact

Once these forms are completed your data will be utilized to create a Microreact project. You should now have a view similar to the one in Figure 4. You should see a map view, tree view and metadata view. You can use click-drag-zoom to navigate both the tree and the map.

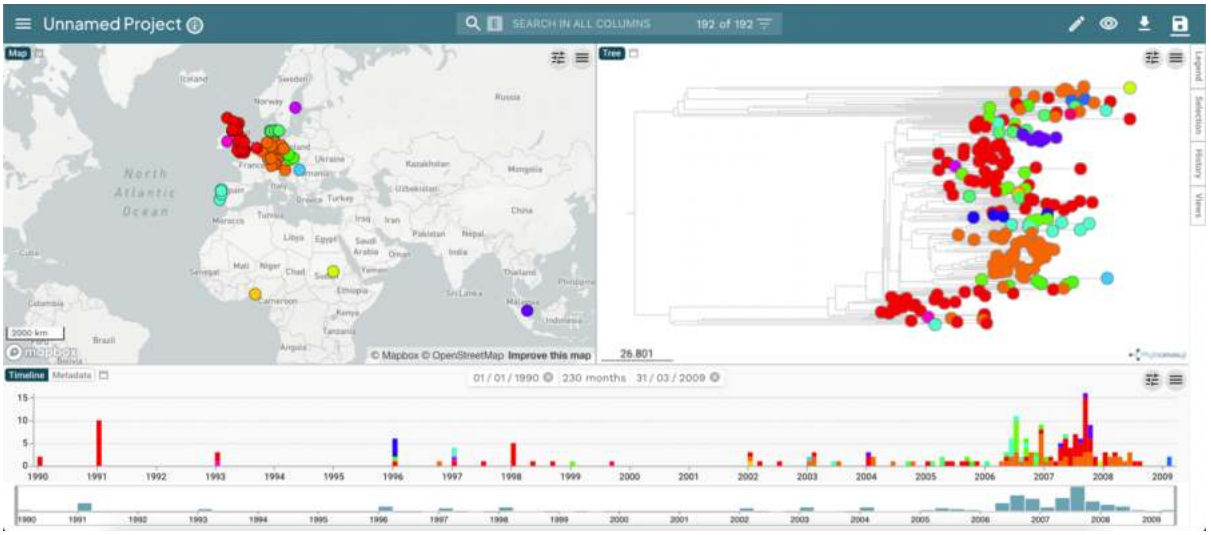

Figure 4 Microreact view of the Holden 2013 dataset

Clicking on ‘Show Controls’ (arrow 1 in Figure 5) button in the tree window allows you to view different kinds of trees and change text/node size. You can change the layout of the tree as pointed by arrow 2. Keep the default Rectangular layout of the tree. Click on the button ‘Hide Controls’ (arrow 1) to hide controls.

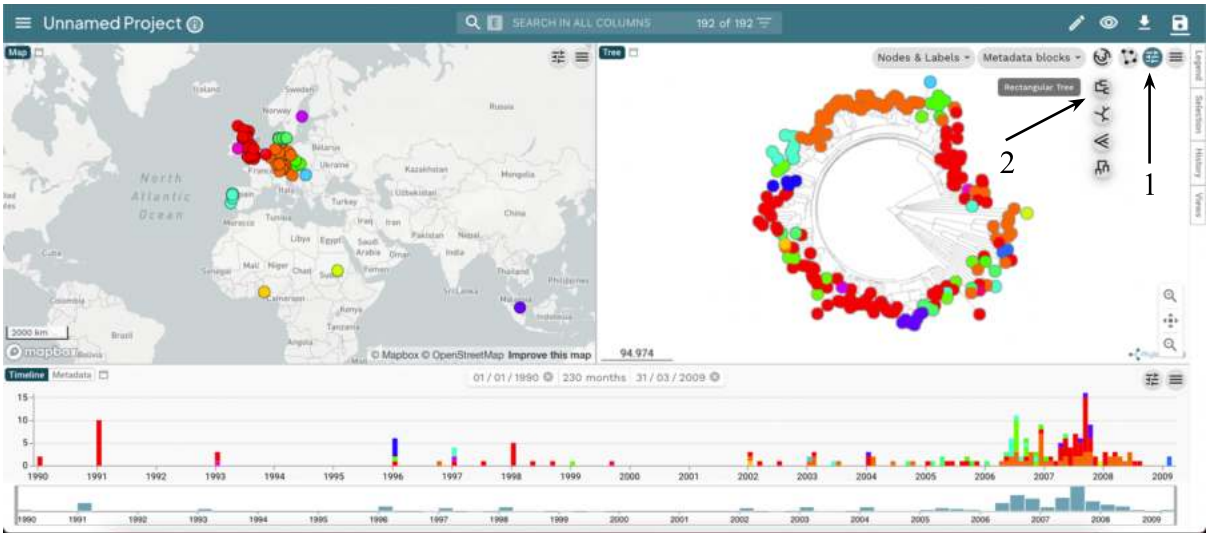

Figure 5 Circular view of the phylogenetic tree

10.3.4 Investigating the country of origin of an epidemic MRSA clone

We are interested in investigating the country of origin of EMRSA15, labelled as group ST22-A2 in the tree. By default, the tips of the tree are coloured-coded by country of origin. Click on the ‘Legend’ button at the top-right corner of your screen (indicated by the arrow in Figure 6) to display the colour legend.

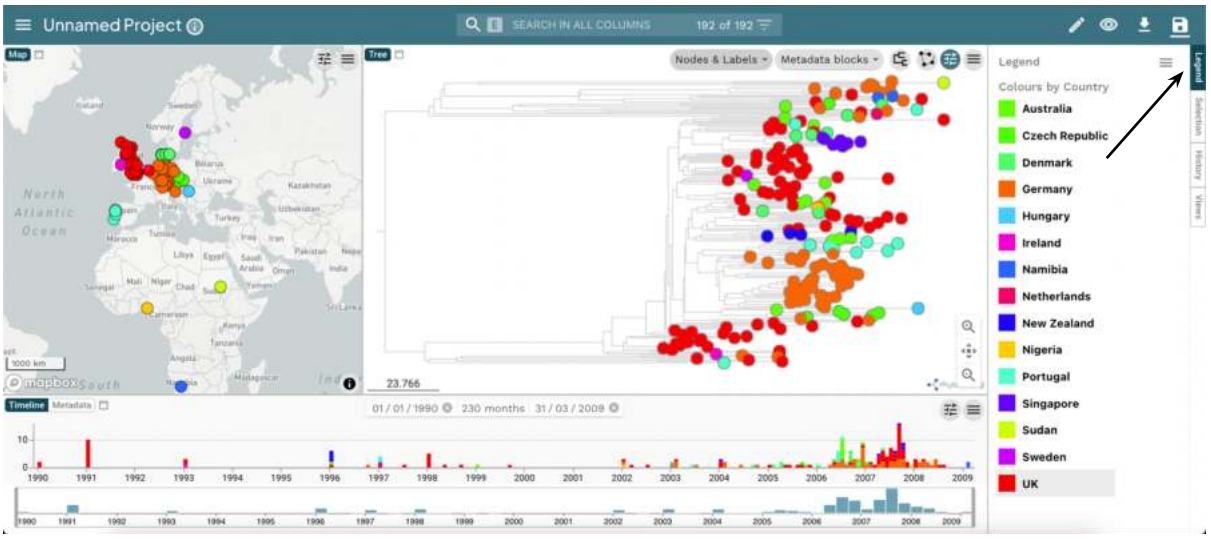

Figure 6 Tips are colour coded by Country of origin

The  icon on the top-right corner of the window (arrow 1 in Figure 7) will allow you to colour and label the tips of the tree based on different metadata fields. Here, in the drop-down list under 'Labels Column' select 'Group' to label the tips (or leaves) based on this field, which indicates the phylogenetic group or clade of ST22 isolates. Here, also make sure tips remain coloured by country of origin by keeping the option 'Country' under 'Colour column'. Remember you can use the settings 'Labels, Colors, and Shapes' to change the way metadata fields are displayed on the tree.

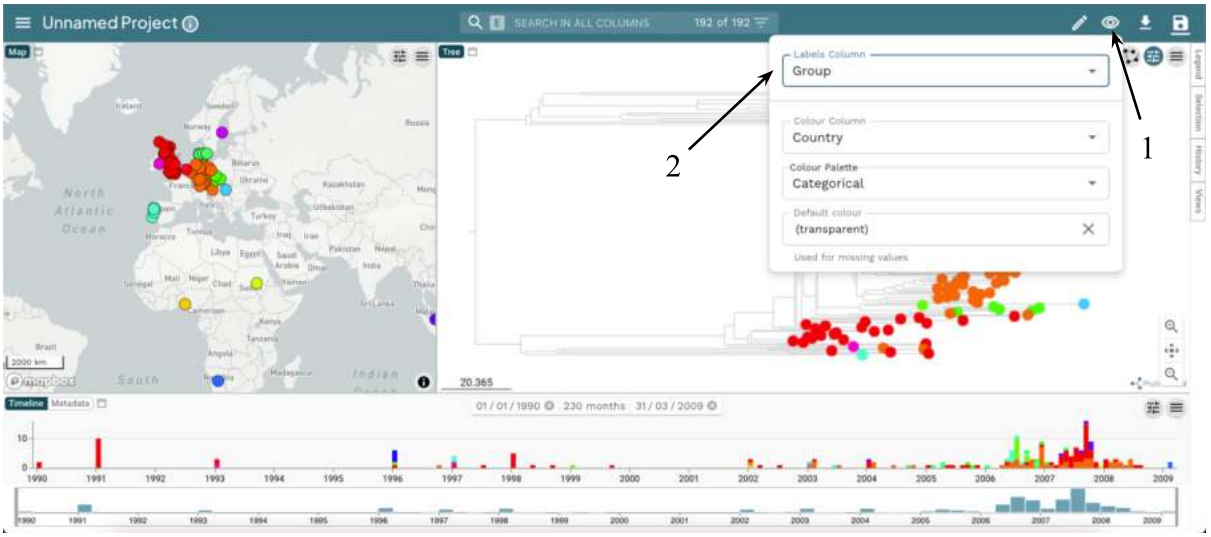

Figure 7 Tips are labelled by Group

Expand the Tree panel at the expense of the Map panel to allow for more space to visualise the phylogenetic tree on your screen (arrow 1 in Figure 8). Click on the 'Show controls' button  (arrow 2), open the 'Nodes & Labels' window (arrow 3), and click on the toggle button  next to 'Leaf Labels' (arrow 4) to ensure tip labels are visible. Here you can also change the font size of tip labels with the corresponding bar.

Make sure the branches/internal edges of your phylogenetic tree are coloured-coded by Country by selecting the toggle button 'Colour Internal Edges' (arrow 5). You will need to zoom in to be able to see the labels of leaves/tips on the tree. Once the Group labels (e.g. non-ST22-A, ST22-A1) are visible on your screen, close the 'Nodes & Labels' window and 'Hide controls' (arrow 2).

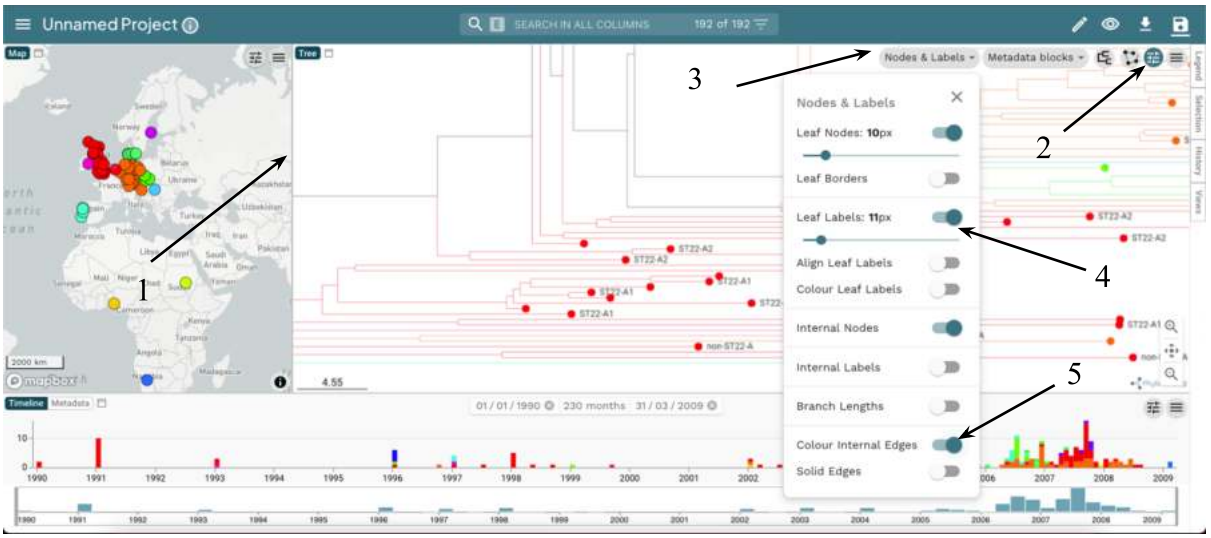

Figure 8 Changing Nodes and Labels on the phylogenetic tree

Now, zoom in to identify the internal node that represents the common ancestor of all ST22-A isolates, that is the internal node that splits groups ST22-A and non-ST22, as pointed by the arrow in Figure 9 (i.e., node 26). Next, identify the internal node representing the split between ST22-A1 and ST22-A2 (i.e., node 28). Here note that clade ST22-A2 is a sub-clade within the broader ST22-A1 clade.

Figure 9 Common ancestor of all ST22-A isolates

Now click on the 'Timeline' button at the left corner of the bottom panel (pointed by arrow 1 in Figure 10) to display the time of isolation of samples in the tree.

Investigate the year and country of isolation of ST22-A1 isolates, the most basal group to ST22-A2 isolates, by clicking on the internal node that represents the common ancestor of all ST22-A1 isolates (arrow 2).

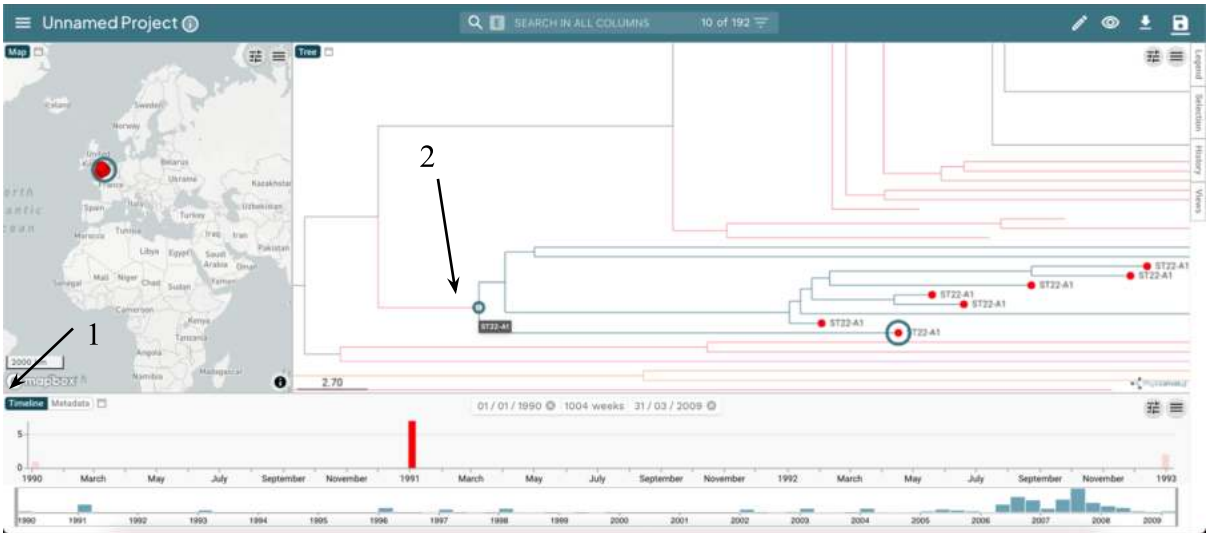

Figure 10 Geographical location and time of ST22-A1 isolates

Based on the location of ST22-A1 isolates on the map and their year of isolation, the most likely origin of clone ST22-A2 (commonly referred as to EMRSA15) was the United Kingdom during the 1990s.

10.3.5 Evidence of regional spread of an epidemic MRSA clone

Click on the 'Legend' button at the top-right corner of the screen as pointed by arrow 1 in Figure 11. A list of countries will be displayed with a colour legend under the list 'Colours by Country'. Click on 'Portugal' (arrow 2) to highlight all isolates from this country on the Tree, Map (arrow 3) and Timeline panels (Figure 11).

Figure 11 Filtering and highlighting isolates per country

You will notice that most isolates from Portugal are part of the same monophyletic clade in the phylogenetic tree defined by node 121 (arrow 1 in Figure 12). Make sure the branches/internal edges of the phylogenetic tree are coloured-coded by Country by selecting the option 'Colour Internal Edges' (arrow 5 in Figure 8) in the 'Nodes & Labels' window. Colour-coded internal branches/edges can help you identify the country of origin of samples from the same clade and identify putative migration events.

Figure 12 Colour-coded internal branches/edges help identify migration events.

Country-specific clades like this one result from point introductions and later diversification of bacterial clones within the same country/region. The earliest instance of this clone in Portugal is in 1997 and the latest in 2016.

The UK has by far the greatest number of clades represented in the ST22-A2 clade. This further supports the UK as the origin of the ST22-A2 (EMRSA15) clone, with multiple introductions from the UK into Germany, Denmark, and Australia, possibly via migration.

10.3.6 Tracing the origin and spread of antibiotic resistance.

Click on the 'Labels, Colours, and Shapes'  icon (arrow 1 in Figure 13) at the top-right corner of the window to change the colour and labels of tips in the tree. Click on 'Ciprofloxacin' in the drop-down list under 'Colour Column' (arrow 2 in Figure 13).

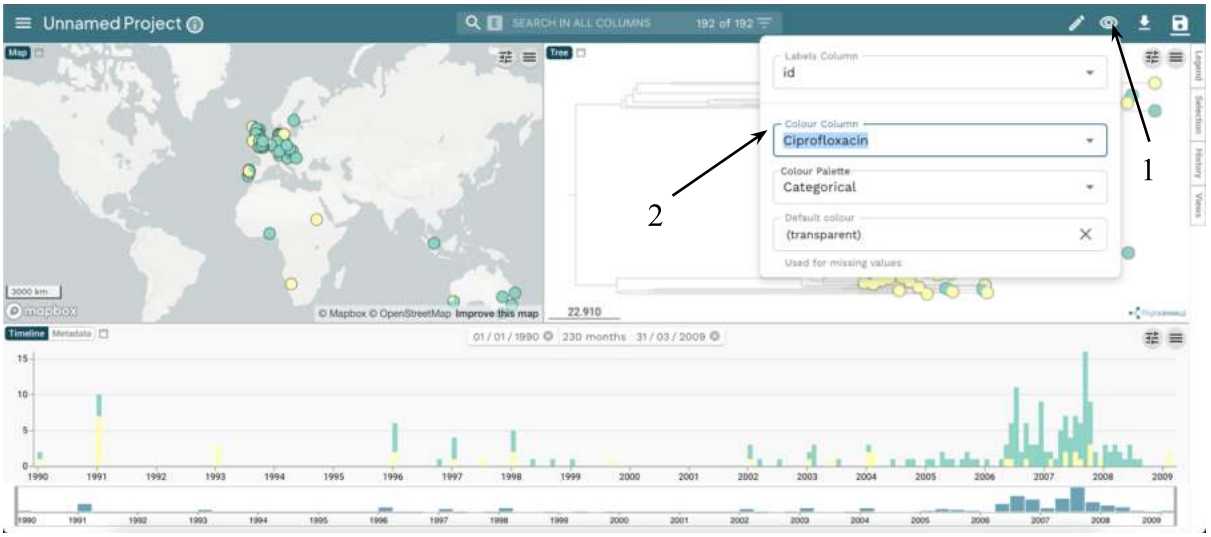

Figure 13 Geographical distribution of ciprofloxacin resistance

Ciprofloxacin resistance seems to have emerged four times independently in the phylogenetic tree (in node 189, node 28, sample 06-00896 and sample 08_5828_K). The map shows the global distribution of ciprofloxacin resistance. Rather than being restricted to a particular country, ciprofloxacin resistance is present in countries from all sampled continents. Ciprofloxacin resistance results from two point mutations generating amino acid substitutions Ser80Phe in topoisomerase IV (GrlA) and Ser84Leu in gyrase A (GyrA). These two mutations are found in the ST22-A2 clade (defined by node 28) that spread throughout the UK and globally. This is a clear example of clone-mediated intercontinental transmission of antibiotic resistance.

Click gain on the ‘Labels, Colours, and Shapes’  icon and select ‘Clindamycin in the drop-down list under ‘Colour Column’.

Figure 14 Geographical distribution of clindamycin resistance

In contrast to the global dissemination of ciprofloxacin resistance, clindamycin resistance seems to be restricted to central Europe, particularly to Germany. The

authors (Holden *et al.* 2013) hypothesized that country-specific differences in antibiotic usage could be causing these differences in the prevalence of clindamycin resistance. They showed that prescription levels of clindamycin in Germany are much higher than in the UK which would result in a higher selective pressure for maintaining clindamycin resistance.

References:

Aanensen DM, Feil EJ, Holden MTG, Dordel J, Yeats CA, Fedosejev A, Goater R, Castillo-Ramírez S, Corander J, Colijn C, et al. 2016. Whole-Genome Sequencing for Routine Pathogen Surveillance in Public Health: a Population Snapshot of Invasive Staphylococcus aureus in Europe. *mBio* **7**: e00444-16.

Baker S, Thomson N, Weill F-X, Holt KE. 2018. Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens. *Science* **360**: 733–738.

Coll F, Harrison EM, Toleman MS, Reuter S, Raven KE, Blane B, Palmer B, Kappeler ARM, Brown NM, Török ME, et al. 2017. Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community. *Science Translational Medicine* **9**: eaak9745.

Croucher NJ, Didelot X. 2015. The application of genomics to tracing bacterial pathogen transmission. *Current Opinion in Microbiology* **23**: 62–67.

Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, Ip CLC, Golubchik T, Batty EM, Finney JM, et al. 2013. Diverse sources of C. difficile infection identified on whole-genome sequencing. *The New England Journal of Medicine* **369**: 1195–205.

Harris JB, LaRocque RC, Qadri F, Ryan ET, Calderwood SB. 2012. Cholera. *The Lancet* **379**: 2466–2476.

Harris SR, Cole MJ, Spiteri G, Sánchez-Busó L, Golparian D, Jacobsson S, Goater R, Abudahab K, Yeats CA, Bercot B, et al. 2018. Public health surveillance of multidrug-resistant clones of Neisseria gonorrhoeae in Europe: a genomic survey. *The Lancet Infectious Diseases* **1407**: 1–11.

Holden MTG, Hsu L-Y, Kurt K, Weinert L a, Mather AE, Harris SR, Strommenger B, Layer F, Witte W, de Lencastre H, et al. 2013. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus aureus pandemic. *Genome Research* **23**: 653–664.

Reuter S, Török ME, Holden MT, Reynolds R, Raven KE, Blane B, Donker T, Bentley SD, Aanensen DM, Grundmann H, et al. 2016. Building a genomic framework for prospective MRSA surveillance in the United Kingdom and the Republic of Ireland. *Genome Research* **26**: 263–270.

Snitkin ES, Zelazny a. M, Thomas PJ, Stock F, Henderson DK, Palmore TN, Segre J a. 2012. Tracking a Hospital Outbreak of Carbapenem-Resistant Klebsiella pneumoniae with Whole-Genome Sequencing. *Science Translational Medicine* **4**: 148ra116.

Spagnoletti M, Ceccarelli D, Rieux A, Fondi M, Taviani E, Fani R, Colombo MM, Colwell RR, Balloux F. 2014. Acquisition and Evolution of SXT-R391 Integrative Conjugative Elements in the Seventh-Pandemic Vibrio cholerae Lineage. *mBio* **5**: e01356-14-e01356-14.

Tong SYC, Holden MTG, Nickerson EK, Cooper BS, Köser CU, Cori A, Jombart T, Cauchemez S, Fraser C, Wuthiekanun V, et al. 2015. Genome sequencing defines phylogeny and spread of methicillin-resistant Staphylococcus aureus in a high transmission setting. *Genome Res* **25**: 111–118.

Weill F, Domman D, Njamkepo E, Tarr C, Rauzier J, Fawal N, Keddy KH, Salje H, Moore S, Mukhopadhyay AK, et al. 2017. Genomic history of the seventh pandemic of cholera in Africa. *Science* **358**: 785–789.