

AMR-TV: antimicrobial resistance transmission visualizer

Ivan S Gill

isgill93@student.ubc.ca

Introduction

Antimicrobial resistance, or AMR, describes microorganisms that have evolved resistance against antimicrobial medication (Levy, 1982). The evolution of AMR genes is largely driven by antimicrobial medication overuse, which occurs through channels such as self-medication (Rather *et al.*, 2017), clinical prescriptions (Luyt *et al.*, 2014), livestock management (Tang *et al.*), and the use of antimicrobial pesticides (Ramakrishnan *et al.*, 2019). After AMR evolves in a bacterial population, the associated gene can spread quickly to other populations via horizontal gene transfer, as many AMR genes occur on mobile genetic elements (Moser *et al.*, 2018). Due to both horizontal gene transfer and global antimicrobial misuse, AMR is currently evolving and spreading rapidly, with up to 10 million AMR-attributed deaths expected annually by 2050 (O'Neill, Jim, 2016). One method for mitigating the rate of AMR-induced illnesses and fatalities is through the improvement of AMR surveillance. By tracking the spread of AMR more effectively, clinicians and policymakers can make more informed decisions on the prescription and supply of appropriate antimicrobial medications in response to AMR outbreaks.

We will develop AMR-TV (antimicrobial transmission visualizer) to visualize potential routes of AMR transmission between sampled pathogens. Using data from the NCBI Pathogen Detection Isolates Browser (National Center for Biotechnology Information, 2020), we will visualize shared AMR genes across sampled pathogens, while also encoding information pertaining to sample location, date, and host. We will create an interactive high-level adjacency matrix view that summarizes the number of shared AMR genes across sampled isolates from different organism groups, and also allows users to narrow down the large NCBI dataset to a subset of interest. We will utilize a low-level node-link diagram view to explicitly illustrate potential AMR transmission events between sampled isolates from the user-specified subset.

My personal expertise with AMR is limited, as I have only begun investigating AMR in detail this term. The lab I started my MSc with has done research on AMR in the past, and proposed a project dedicated to visualizing AMR transmission for my thesis. This project I am proposing will be the first component of my MSc thesis. After completing this course project, I will expand AMR-TV in three ways over the course of my thesis. First, I will allow users to upload their own spreadsheets, rather than visualize a static rendition of the NCBI Pathogen Detection Isolates Browser. Second, I will integrate an ontology framework to automatically clean and standardize user-submitted data. Third, I will make changes to the user interface as requested when I present AMR-TV to the appropriate stakeholders, such as clinicians and microbiologists.

Related work

There are other applications that already visualize transmission routes between data points, both within and outside the field of microbiology. Microreact is a web application that visualizes microbial organism transmission across spatial regions (Argimón *et al.*, 2016). Its main interface consists of a map, tree, and time series view for encoding spatial, transmission, and temporal data respectively. Outside microbiology, VAIM is an application that visualizes the spread of influence, or word-of-mouth, within social networks (Arleo *et al.*, 2020). Its main interface allows users to select different influence maximization algorithms, and then observe the spread of influence through a node-link diagram. VAIM also provides a high-level matrix view for detecting clusters of nodes in the lower-level node-link diagram. I will spend the remainder of this section discussing critical differences between these applications and AMR-TV.

Microreact differs from AMR-TV in four critical ways. First, Microreact models microbial organism transmission across spatial regions, while AMR-TV will model gene transmission across microbial organisms. Second, Microreact separates spatial, transmission, and temporal information into separate views, while AMR-TV will encode this information directly into a single node-link diagram. We believe observing trends across attributes, and across data points, will be easier if all relevant information is in the same view. Third, Microreact requires users to interact with a map view to narrow large datasets down to subsets of interest, while AMR-TV will use an adjacency matrix for this purpose. We believe an adjacency matrix is the superior choice for modelling large amounts of data for the user to parse, as it can support many more data points without losing readability to factors such as overlapping data points in a spatial region. Fourth, Microreact hypothesizes microbial transmission routes by constructing a phylogenetic tree from genomic data. As AMR genes can spread horizontally between populations, a phylogenetic tree is not as effective for hypothesizing AMR gene transmission. AMR-TV will be hypothesizing transmission routes by constructing a directed graph, based solely on shared AMR genes, which may contain cycles for the user to interpret.

VAIM differs from AMR-TV in two critical ways. First, VAIM encodes temporal information through animation, with new nodes appearing as the animation proceeds. AMR-TV will not use animation, but will instead encode sampling date information directly into the static lower-level node-link diagram. This static representation can be more effectively employed in scientific papers or presentations by clinicians and policymakers. Second, VAIM's matrix view does not serve the purpose of allowing users to narrow large datasets down into subsets of interest. Instead, its matrix shows the density of node clusters in different regions of the node-link diagram, and allows users to jump to these different regions. AMR-TV's purpose is to monitor potential transmission events that have occurred within AMR outbreaks of interest, so AMR-TV's matrix view is better utilized as a method for narrowing down large amounts of data, than it is to jumping around a gigantic node-link diagram encompassing the entirety of the NCBI Pathogen Detection Isolates Browser.

Data and task abstractions

The NCBI Pathogen Detection Isolates Browser has over 600K rows of data, and seven columns relevant to AMR-TV's goal of visualizing potential AMR transmission events. These columns are "organism group", "isolate", "create date", "location", "isolation source", "host", and "AMR genotypes". There are some missing and unstandardized values in this dataset, particularly under "isolation source" and "host", but data cleaning is not within the scope of this project.

The dataset type we model in our data abstraction will be a network. The items in this network will represent individual rows from the NCBI dataset, identified by a unique isolate value, and the links will represent shared AMR genotypes across rows. There will be four categorical attributes: "organism group", "location", "isolation source", and "host". There will be one ordered and sequential attribute, "create date".

The intended action of AMR-TV will be to analyze and consume. At a high-level, AMR-TV will allow users to both discover and present, by visualizing potential AMR transmission routes between sampled isolates, and also providing static diagrams that can be used in papers and presentations accompanying stakeholder discussions on effectively treating AMR outbreaks. At a mid-level, AMR-TV will allow users to lookup and search a subset of data from the NCBI dataset based on two features: "organism group" and "create date". At a low-level, AMR-TV will allow users to query and summarize, by displaying all proposed AMR transmission routes between isolates from this searched subset of data.

The intended target of AMR-TV will be the paths present in the network data, which will indicate proposed routes of AMR transmission, but there are other features describing all data that will be targeted as well. Encoding "location", "isolate source", and "host" will allow users to consider AMR gene transmission routes in the context of any physical routes that may provide the proximity needed for horizontal gene transfer to occur, such as geographic routes and interspecies relationships. Encoding organism groups, and which AMR genes are transmitted across links, will inform clinicians of symptoms, outcomes, and which antimicrobial medications will not work during AMR outbreaks.

Solution

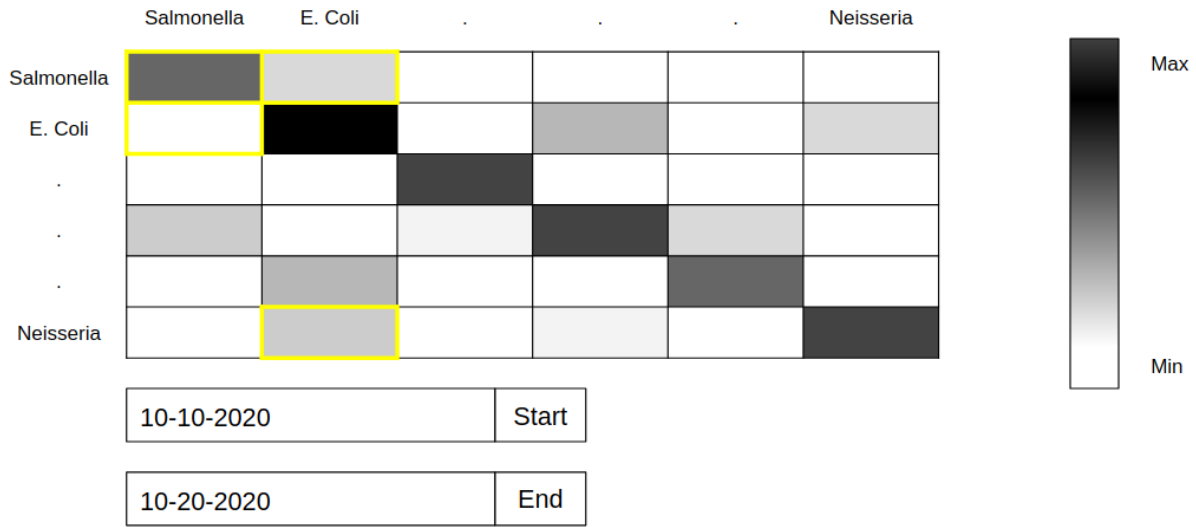
AMR-TV will consist of a high-level and low-level view.

The high-level view will be an adjacency matrix, as seen in Fig. 1A, with rows and columns encoding organism groups such as *Salmonella enterica*, *E.coli* and *Shigella*, and *Neisseria*. Cells will use the luminance channel to encode the number of shared AMR genes between isolates of one or two organism groups, with darker cells representing a larger number of shared AMR genes across all isolates. Users will be able to interact with the matrix in two ways. First, users can specify a date range to narrow down the amount of isolates displayed in the matrix. Second, users can select cells from the matrix to produce a low-level view. To utilize

colour-encoding, users will not be able to select cells from more than 12 organism groups. It is unlikely that users concerned with AMR outbreaks will need to examine more than 12 organism groups, but we may consider removing this limitation and producing the low-level view in black and white.

The low-level view will be a node-link diagram, as seen in Fig. 1B. Nodes will encode sampled isolates. Directed links will encode shared AMR genes, or potential AMR transmission events, and will be labelled with the shared gene names. Thicker links will indicate a greater number of shared AMR genes. Dotted links will indicate genes shared by isolates belonging to the same organism group. The luminance channel will encode the date each isolate was sampled, with lighter nodes representing older isolates. Colour hue, and a three-letter node label, will encode the organism group of isolates. Hovering over a node will display further information on the node, and all immediately connected nodes, including isolate identifier, exact sampling date, host, and isolation source. Clicking nodes will keep this display active.

A)



B)

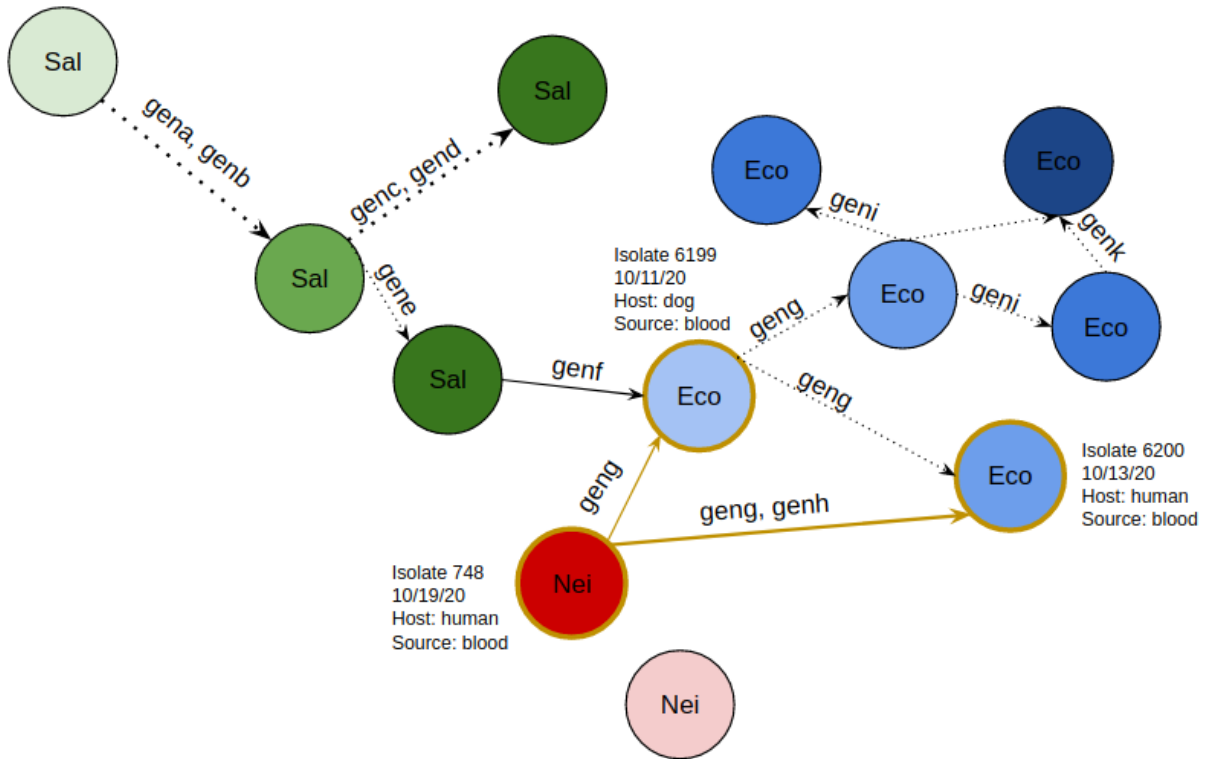


Fig. 1. **A)** High-level adjacency matrix using the luminance channel to illustrate the number of potential AMR transmission events between isolates of varying organism groups, over the period of 10 days in October 2020. In the final result, Max and Min on the colour bar will be replaced by real numbers. The user has selected four cells, highlighted in yellow, which will produce the node-link diagram in Fig. 1B. **B)** Low-level node-link diagram generated from Fig.

1A. Three different organism groups are differentiated by labels and colour. Darker colours represent more recently sampled isolates over the specified time period in Fig. 1A. Edges are labelled with the genes being transferred between isolates. The more recent *Neisseria* isolate node was selected, which automatically selected two connected nodes, displaying detailed information on all three nodes.

Implementation

We will implement AMR-TV using Docker, with a Django container to display the interactive visualizations, and a Postgres container to store the NCBI Pathogen Detection Isolates data. We will use the Python library NetworkX to generate networks. NetworkX contains many different algorithms for generating graphs, but we may try other algorithms in the literature if time permits. We will use the Python library Plotly to draw the high-level adjacency table view, and by using the data supplied by NetworkX, we will use Plotly to draw the low-level node-link diagram view too. Plotly is interactive and reactive, so we should be able to implement the cell and node selection functionalities described in Fig. 1.

Results

Users that want to monitor potential transmission events during an AMR outbreak of interest will begin by inputting start- and end-date values under the high-level adjacency matrix view. Users can then select cells from organism groups they are interested in, or click cells that are dark enough to warrant further exploration. AMR-TV will then generate a low-level node-link diagram view of the user's selected cells. It should be noted that when users consider the darkness of cells, they would expect darker cells on the main diagonal of the adjacency matrix, due to isolates from the same organism group possessing an increased chance of sharing AMR genes due to common ancestry.

Using the node-link diagram, users then begin to infer potential AMR transmission events between isolates. By examining the colour hue of the nodes, and their three letter labels, users can infer transmission events between organism groups. Users can infer the order of transmission, by observing the luminance of nodes, and the directionality of link arrowheads. To infer potential routes of transmission facilitated by geography, or interspecies relationships, users can select nodes for further details on each isolate. Users can more effectively understand the AMR characteristics in their isolates of interest, by reading the labels on each link to understand exactly which genes are being transmitted.

Milestones

Setup Django+Postgres framework in Docker container, with NetworkX and Plotly installed

- Estimated hours: 4 hours
- Estimated completion: Oct 26

Create Django fixture containing NCBI data, and load it into Postgres database

- Estimated hours: 8 hours

- Estimated completion: Nov 3

Create API calls for querying database that will be needed for user-specified visualizations

- Estimated hours: 16
- Estimated completion: Nov 13

Using plotly, create adjacency matrix with shaded cells and colour bar, using data from some stub, static period of time, with stub cell selection ability that doesn't do anything but highlight cells yet

- Estimated hours: 8 hours
- Estimated completion: Nov 16

Set-up time widgets to allow users to select time period of interest, and dynamically update adjacency matrix

- Estimated hours: 8 hours
- Estimated completion: Nov 18

Using selected cells/API calls to get appropriate isolates for nodes, and shared AMR genotypes as links between nodes, generate rudimentary black and white node-link diagram

- Estimated hours: 8 hours
- Estimated completion: Nov 25

Add 3-letter labels to nodes, gene labels to edges, and directionality/width to arrows

- Estimated hours: 8 hours
- Estimated completion: Nov 30

Add colour hue to organism groups, with luminance encoding date too

- Estimated hours: 8 hours
- Estimated completion: Dec 4

Add ability to highlight cells/connected nodes by hovering and clicking

- Estimated hours: 8 hours
- Estimated completion: Dec 8

Bibliography

- Argimón,S. *et al.* (2016) Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb. Genomics*, **2**.
- Arleo,A. *et al.* (2020) VAIM: Visual Analytics for Influence Maximization. *ArXiv200808821 Cs*.
- Levy,S.B. (1982) Microbial resistance to antibiotics. An evolving and persistent problem. *Lancet*, **2**, 83–88.
- Luyt,C.-E. *et al.* (2014) Antibiotic stewardship in the intensive care unit. *Crit. Care*, **18**, 480.
- Moser,K.A. *et al.* (2018) The Role of Mobile Genetic Elements in the Spread of Antimicrobial-Resistant *Escherichia coli* From Chickens to Humans in Small-Scale Production Poultry Operations in Rural Ecuador. *Am. J. Epidemiol.*, **187**, 558–567.
- National Center for Biotechnology Information (2020) Isolates Browser. Bethesda (MD): National Library of Medicine (US).
- O'Neill, Jim (2016) Tackling Drug-resistant Infections Globally: Final Report and Recommendations. HM Government and the Wellcome Trust, London.
- Ramakrishnan,B. *et al.* (2019) Local applications but global implications: Can pesticides drive microorganisms to develop antimicrobial resistance? *Sci. Total Environ.*, **654**, 177–189.
- Rather,I.A. *et al.* (2017) Self-medication and antibiotic resistance: Crisis, current challenges, and prevention. *Saudi J. Biol. Sci.*, **24**, 808.

Tang, K.L. *et al.* Restricting the use of antibiotics in food-producing animals and its associations with antibiotic resistance in food-producing animals and human beings: a systematic review and meta-analysis. *Lancet Planet. Health*, **1**, e316.