User Guide

Overview: This is a user guidance of how to use the application made by team WebScape. The user guidance also explains deeper details of each components that the features provides along with the benefits behind of those components that can provide to users.

Navigating the AMR feature

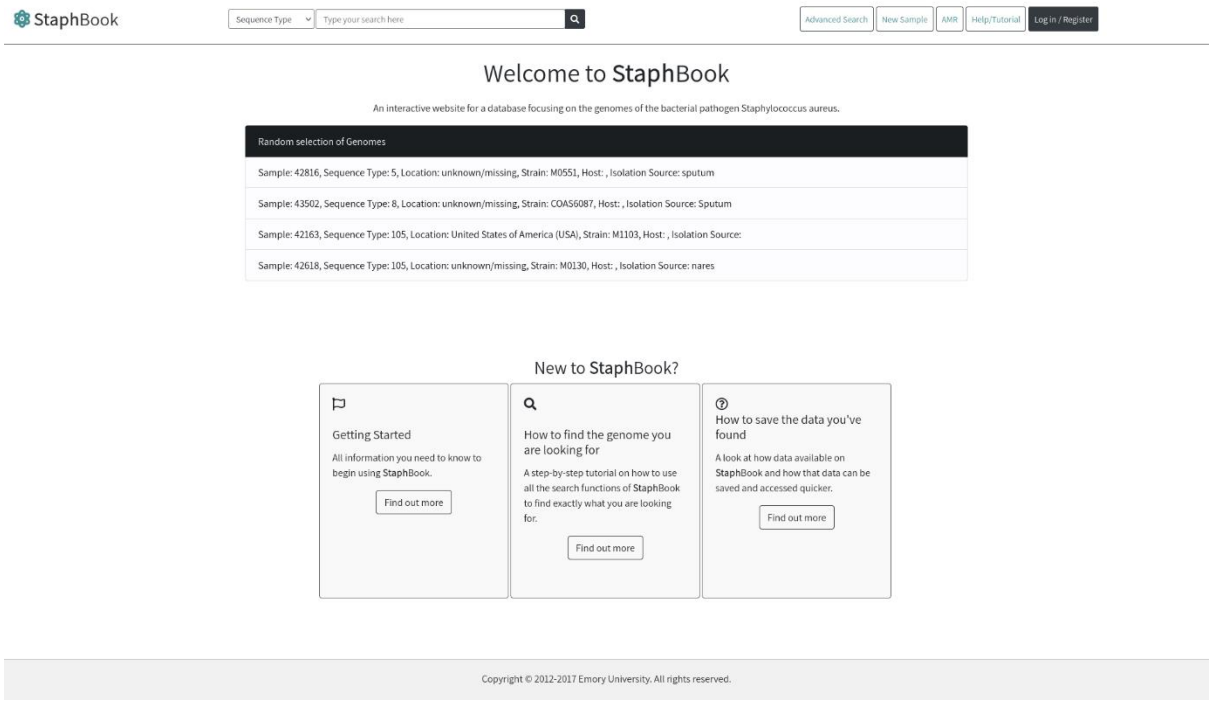

Figure 1: Main Page of the StaphBook

Our team has provided a AMR button on the navigation bar. Users click on that then it takes users the AMR page.

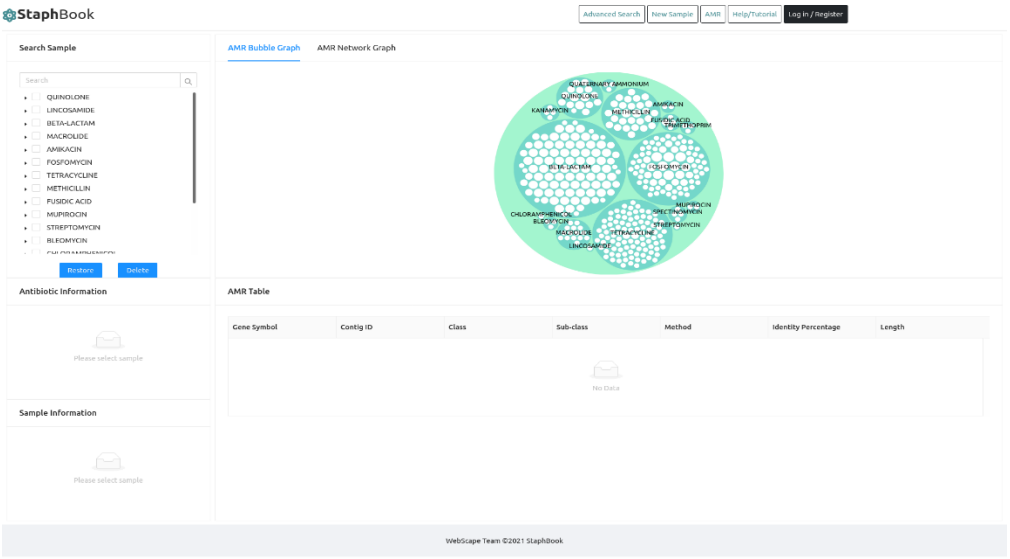

Figure 2: AMR Page

Searching samples

On the left side of the AMR page, there is a list of Anti biotics

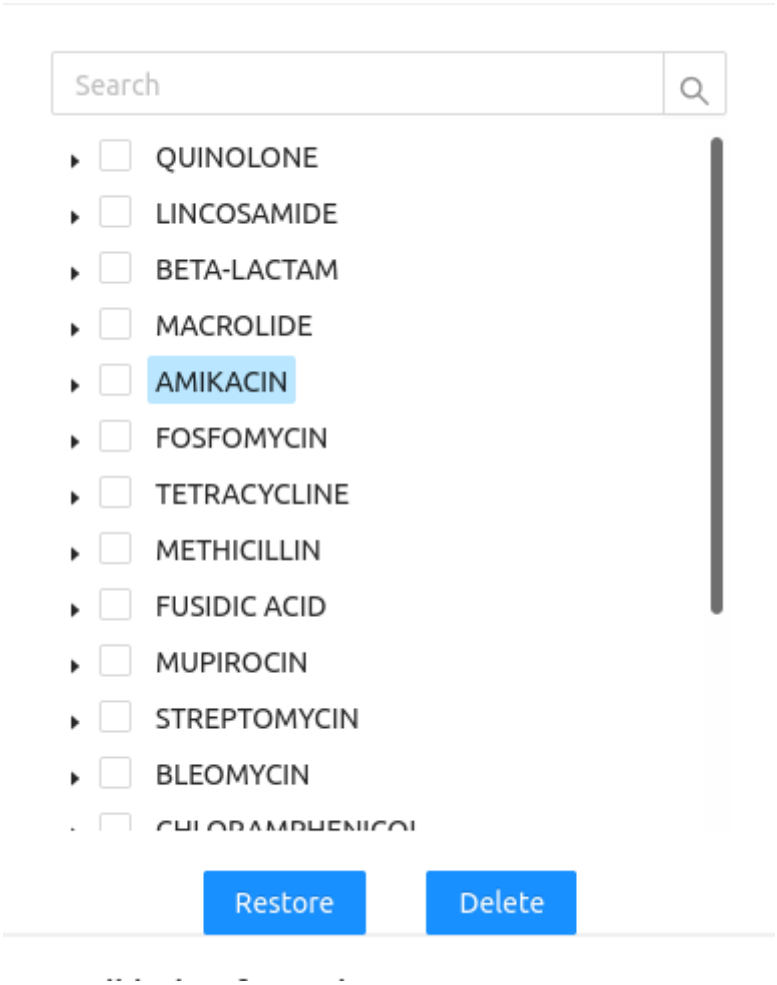

Figure 3: List of samples

When user click on the antibiotics, there is a panel underneath that provides details of the antibiotics such as its structure, its components.

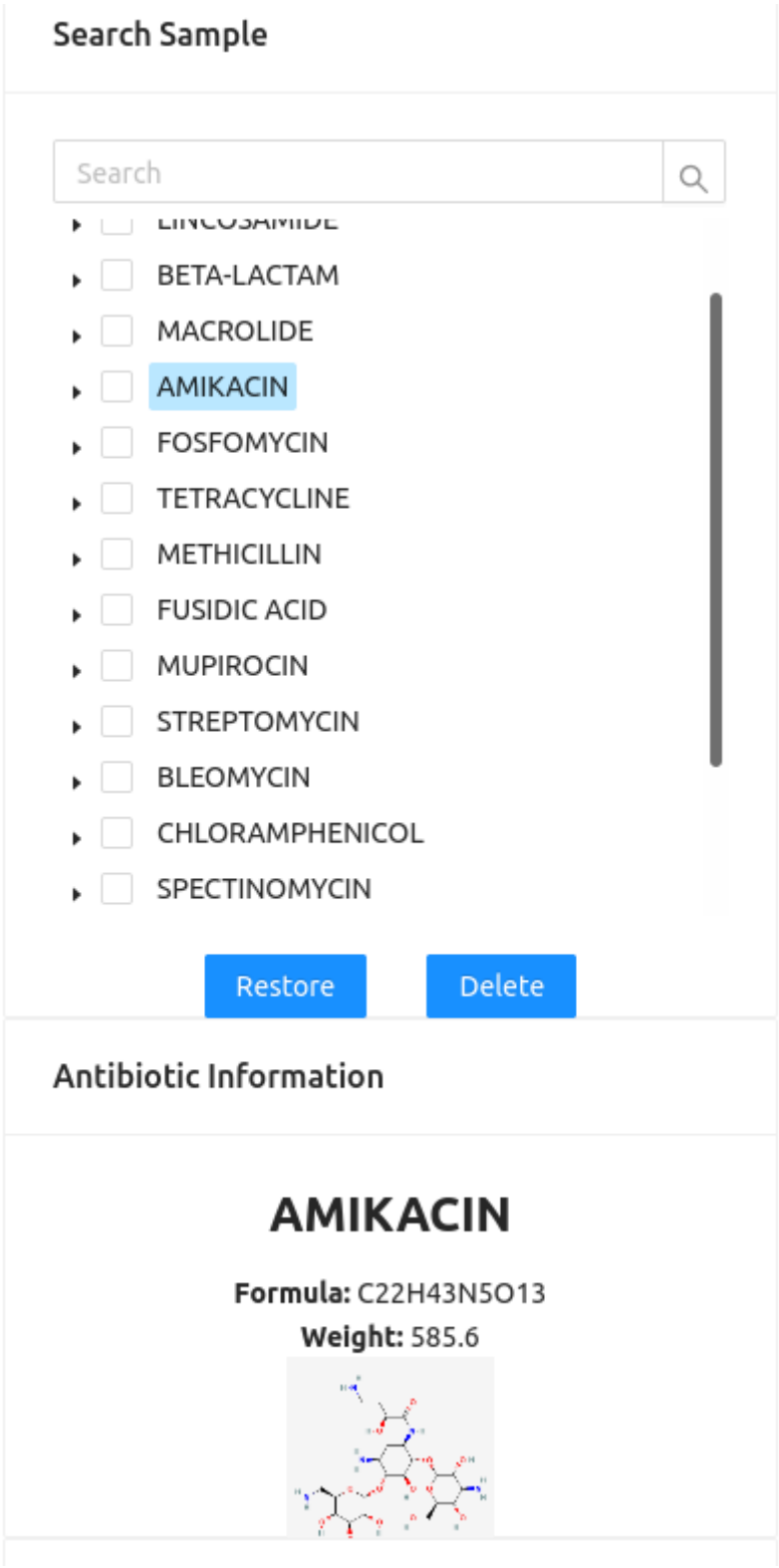

Figure 4: Antibiotic Information

Users can extend the list of samples in each anti biotics, and users can click on the sample to show more information of one particular sample. The sample is the sample of Staphylococcus Aureus that has a sequence or sequences resisting to that particular anti biotics, which is a useful information for biologist. The users are provided more information in the panel at left down corner of the page

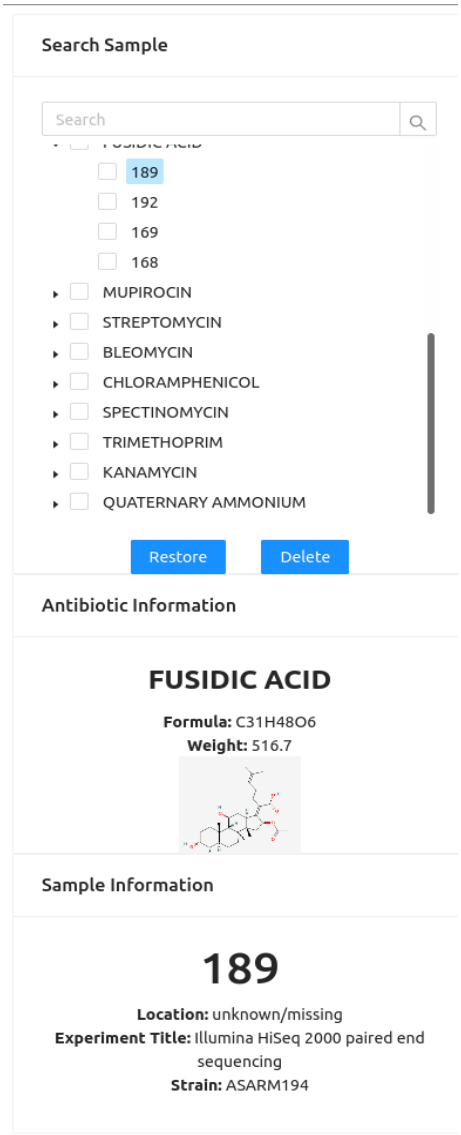

Figure 5 : Sample Information appears

Moreover, there is a table that show sequences information in that sample that has ability resisting to which antibiotics

AMR Table

Gene Symbol	Contig ID	Class	Sub-class	Method	Identity Percentage	Length
mecA	contig39	BETA-LACTAM	METHICILLIN	BLASTX	100	667
mecR1	contig39	BETA-LACTAM	METHICILLIN	PARTIALX	100	325
blaI	contig42	BETA-LACTAM	BETA-LACTAM	BLASTX	99.21	126
blaR1	contig42	BETA-LACTAM	BETA-LACTAM	BLASTX	95.56	585
blaZ	contig42	BETA-LACTAM	BETA-LACTAM	EXACTX	100	281
erm(C)	contig48	MACROLIDE	MACROLIDE	EXACTX	100	244
fusC	contig9	FUSIDIC ACID	FUSIDIC ACID	ALLELEX	100	212
tet(38)	contig9	TETRACYCLINE	TETRACYCLINE	EXACTX	100	450

Figure 6: AMR table

The table is created based on the documentation of the AMRFinderPlus, a tool to predict sequences that resist to particular anti biotics. (https://github.com/ncbi/amr/wiki)

This is info that describe in each column:

- Gene symbol - Gene or gene-family symbol for protein or nucleotide hit. For point mutations it is a combination of the gene symbol and the SNP definition separated by "_"
- Contig id – Contig name in that sample.
- Class - For AMR genes this is the class of drugs that this gene is known to contribute to resistance of.
- Subclass - If more specificity about drugs within the drug class is known it is elaborated here.
- Method - Type of hit found by AMRFinder. A suffix of 'P' or 'X' is appended to "Methods" that could be found by protein or nucleotide.
    - ALLELE - 100% sequence match over 100% of length to a protein named at the allele level in the AMRFinderPlus database.
    - EXACT - 100% sequence match over 100% of length to a protein in the database that is not a named allele.
    - BLAST - BLAST alignment is > 90% of length and > 90% identity to a protein in the AMRFinderPlus database.
    - PARTIAL - BLAST alignment is > 50% of length, but < 90% of length and > 90% identity to the reference, and does not end at a contig boundary.
    - PARTIAL_CONTIG_END - BLAST alignment is > 50% of length, but < 90% of length and > 90% identity to the reference, and the break occurrs at a contig boundary indicating that this gene is more likely to have been split by an assembly issue.
    - HMM - HMM was hit above the cutoff, but there was not a BLAST hit that met standards for BLAST or PARTIAL. This does not have a suffix because only protein sequences are searched by HMM.
    - INTERNAL_STOP - Translated blast reveals a stop codon that occurred before the end of the protein. This can only be assessed if the `-n <nucleotide_fasta>` option is used.
    - POINT - Point mutation identified by blast.
- Length - The length of the query protein or gene. The length will be in amino-acids if the reference sequence is a protein, but nucleotide if the reference sequence is nucleotide.
- Reference sequence length - The length of the Reference protein or nucleotide in the database (NA if HMM-only hit).
- % Identity to reference sequence - % amino-acid identity to reference protein or nucleotide identity for nucleotide reference. (NA if HMM-only hit).

Using the AMR Bubble Graph

At start, the bubble graph renders 100 samples. Each white circle represents one sample and each darker green circle represents of one anti biotics. The white circle inside the darker green circle meaning the sample sequences that resist to that anti-biotics. The white circle is bigger than the others, meaning that the samples has more sequences that resist to the anti-biotics than the other.

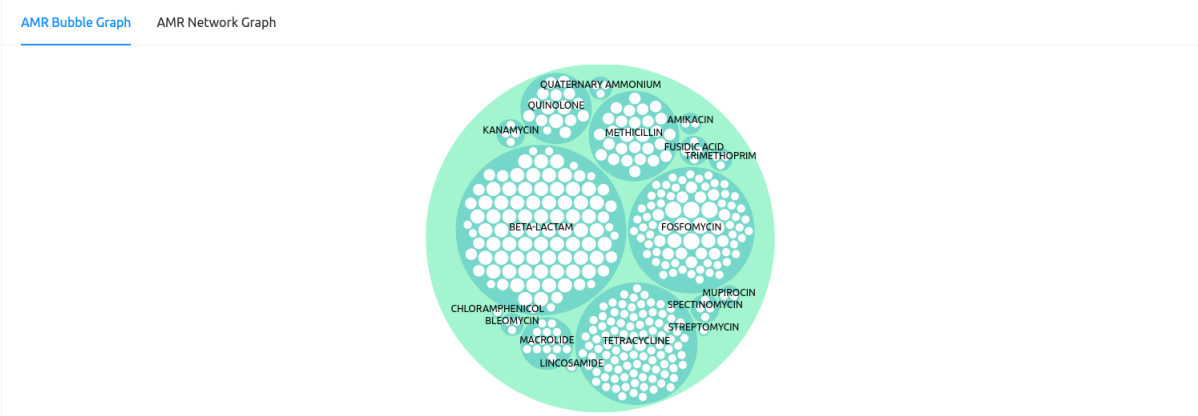

Figure 6: AMR Bubble Graph

Users can click on the the darker green so that the bubble graph zoom in to that circle to see th

Figure 7: Example of clicking the darker circle named Fuscidic Acid

The user can click on the white circle, so that all the panels mention above and the amr table show information of that sample

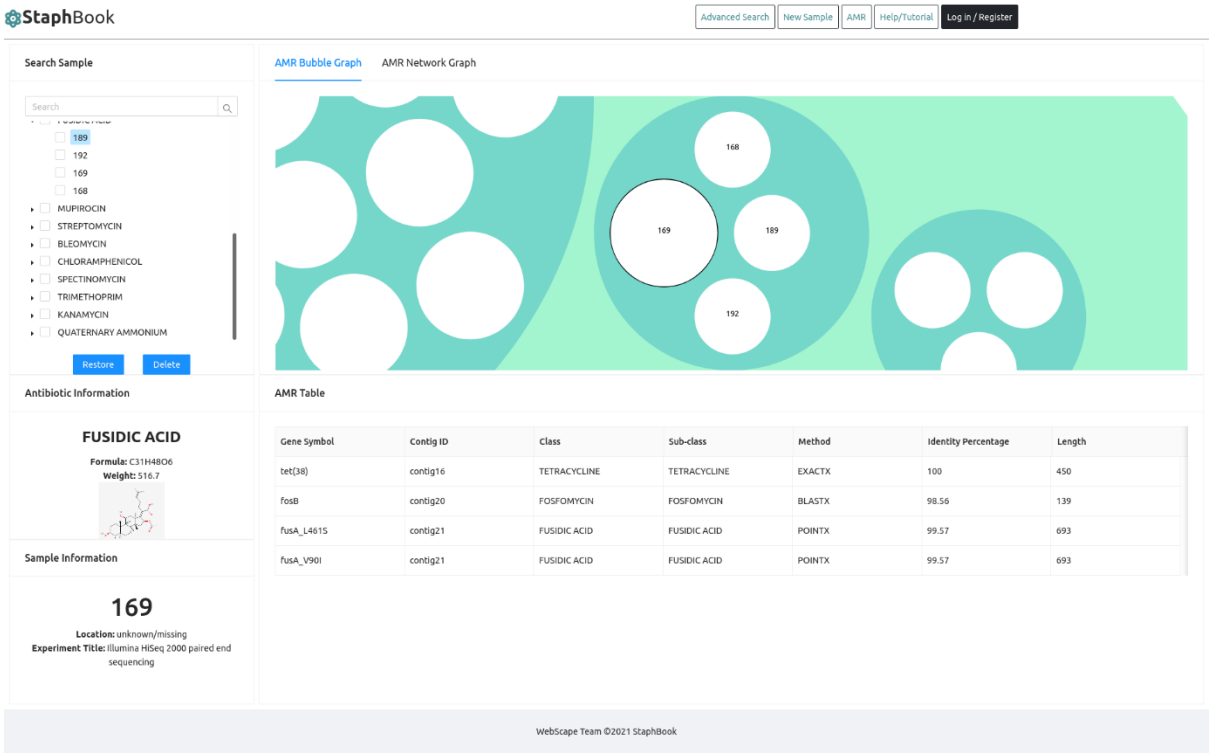

Figure 8: User click on the sample 189

User can filter out darker green circle or white circle by using choosing the samples using the search panel and click on delete button.

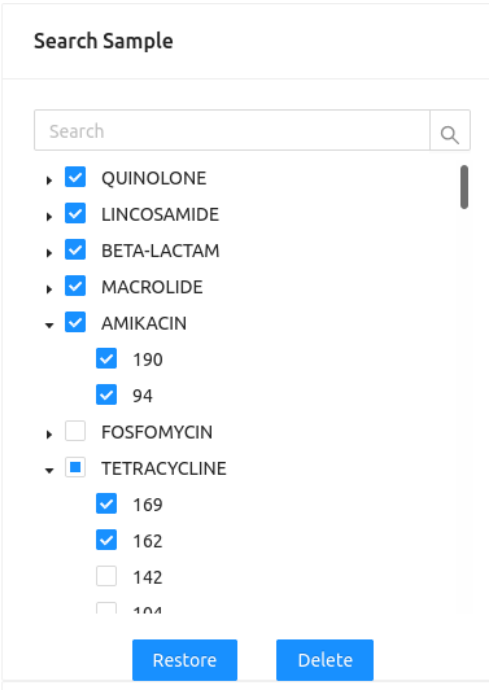

Figure 9: Choosing multiple anti biotics or multiple sample

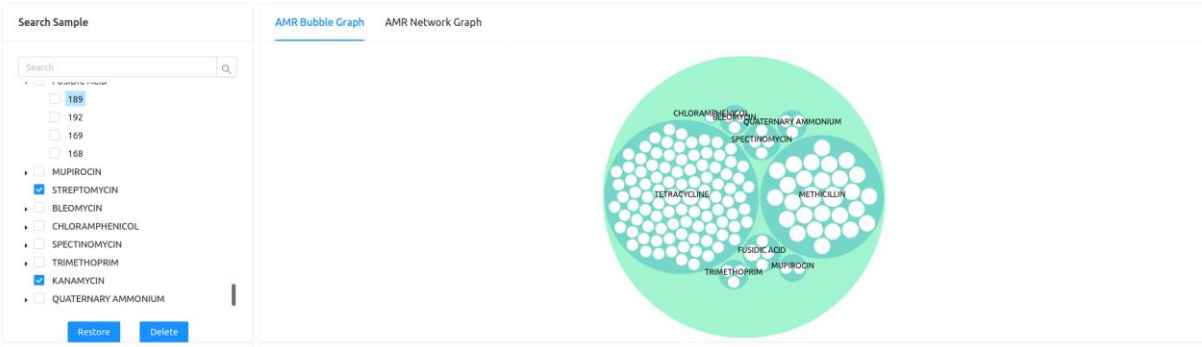

Figure 10: After click the delete button

User can render whole samples again by click the button restore.

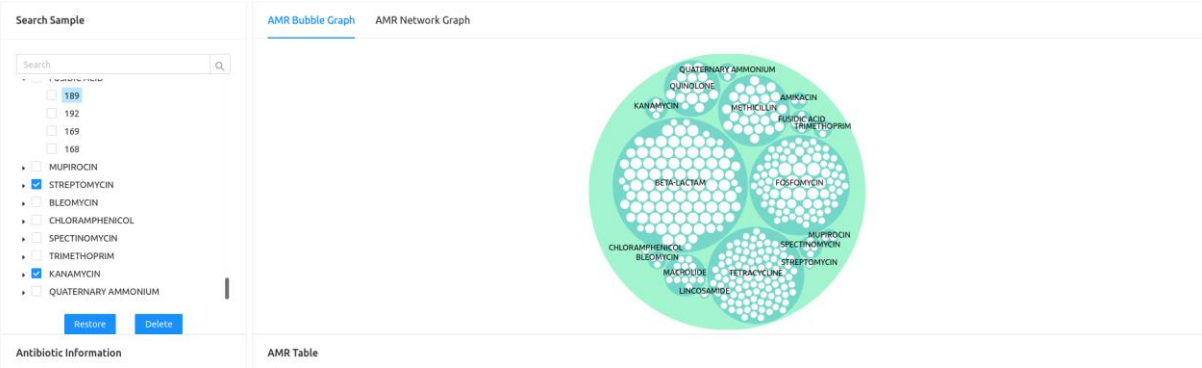

Figure 11: After clicking the button Restore

When the user click on the white circle on the bubble, the circle that has the same sample name are highlighted by increase the border width.

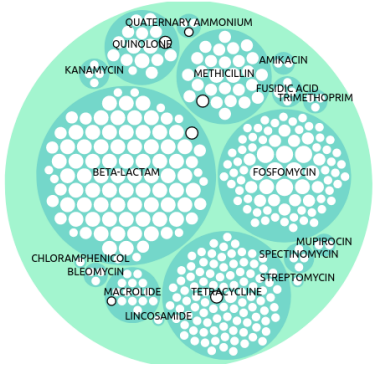

Figure 12: Small white circle with its border width increases after click on one of them

Using the AMR Network Graph

There is a small navigation bar for the user to click to get to the AMR Network Graph

Figure 13: AMR Navigation Bar

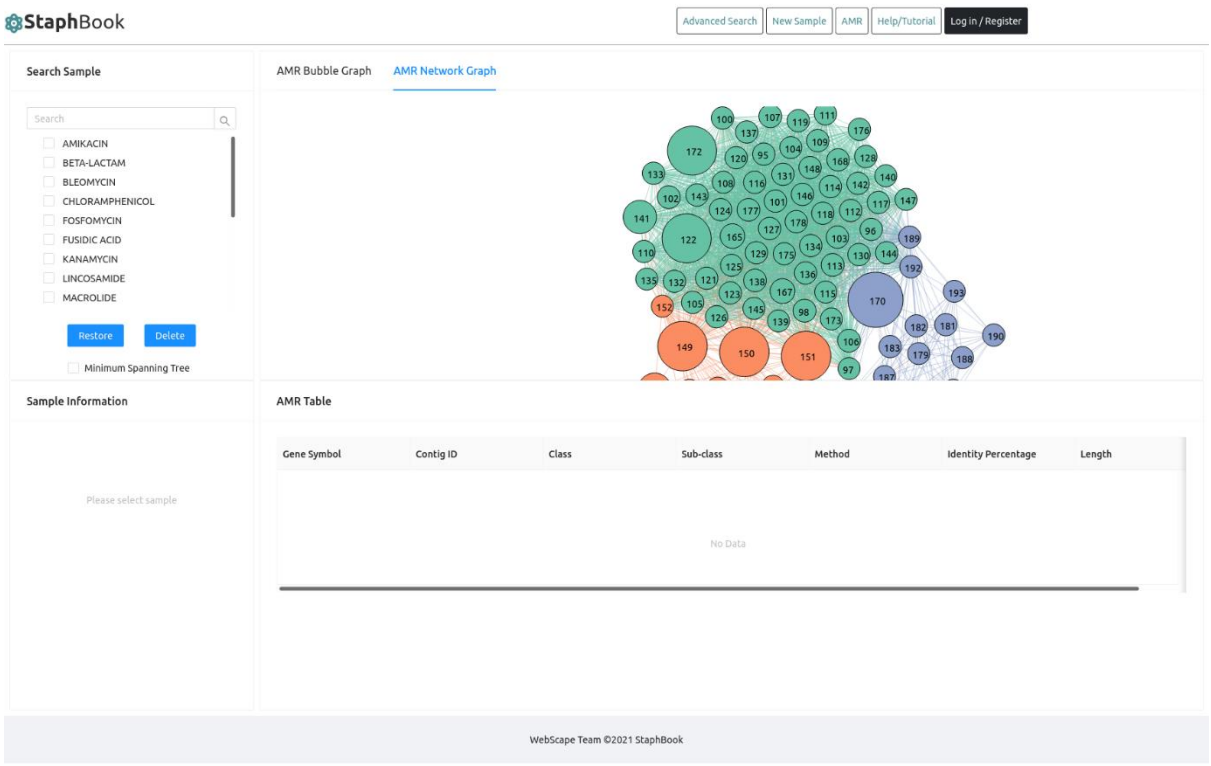

Figure 14: AMR Node Network

Each node in the node network represents a sample. Each line between them is established if the Cosine Similarity between node is greater than the threshold value. The node network illustrate different clusters, each cluster represented by giving different colours. This helps the researcher to find the group having the same behaviour easily.

User can filter out by deleting samples like the bubble graph and the node network will render again with selected samples. Also, the user can restore them.

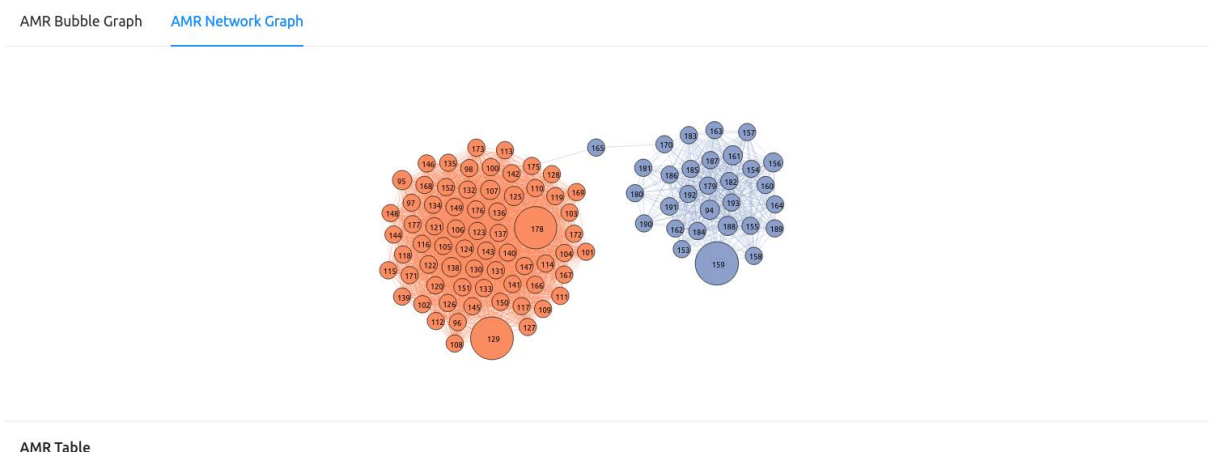

Figure 15: After user delete samples

User can also interact with the node network graph by dragging, hovering, zoom in and zoom out using mouse scroll.

There is another feature called minimum spanning tree and is located in search panel

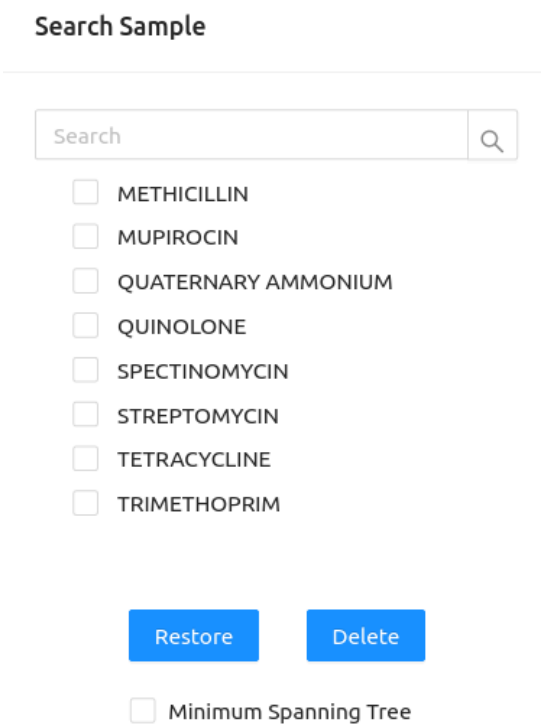

Figure 16: Minimum Spanning is located under.

When the users click on the check box, the node network transform into Minimum Spanning Tree mode, a common type of Node Network that the researchers use. The Minimum Spanning Tree uses Kruskal Algorithm and it is also a common algorithm. The Minimum Spanning Tree mode help users to visualize cluster better without the extra lines between node network.

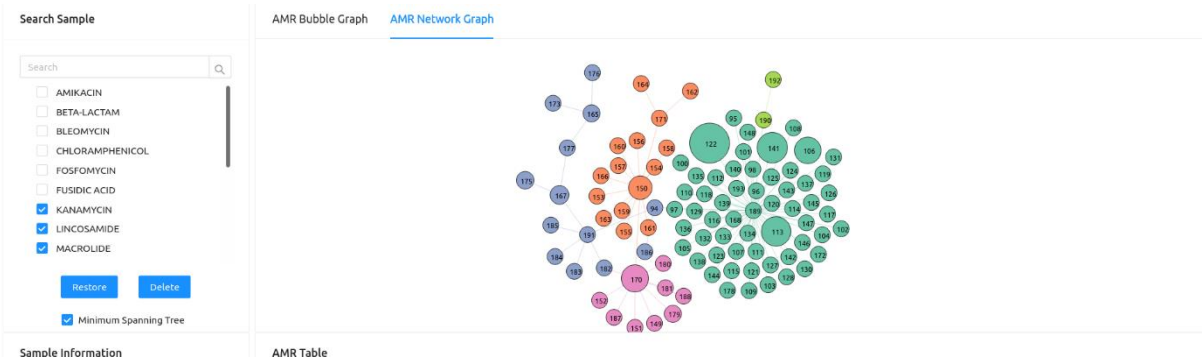

Figure 17: Minimum Spanning Tree Mode of 100 Samples.

Users can also interact with the Minimum Spanning Tree Mode by clicking, dragging, hovering, zoom in and out using mouse scroll. Select sample also gives detail about the samples in Sample Information Panel and AMR Table

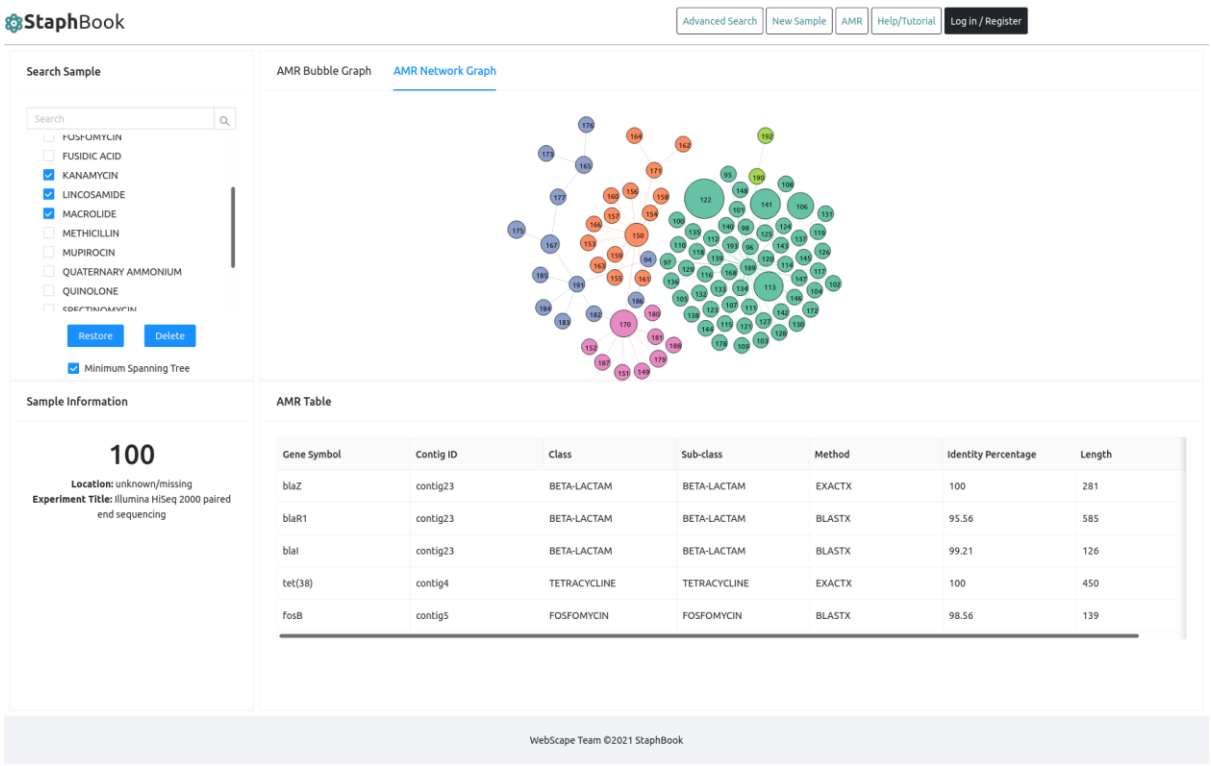

Figure 18: AMR Node Network after user click a sample