

# An Analysis of Antimicrobial Resistance of Clinical Pathogens from Historical Samples for Six Countries

## **Authors:**

Karen Li, Joanna Zheng, Thomas Deng, James Peng, Dagmar Daniel, Qian Jia, Zuyi Huang

*Date Submitted:* 2020-01-20

*Keywords:* hierarchical clustering, data analysis, principal component analysis, clinical pathogens, antimicrobial resistance

## **Abstract:**

The spread of antimicrobial resistance pathogens in humans has increasingly become an issue that threatens public health. While the NCBI Pathogen Detection Isolates Browser (NPDIB) database has been collecting clinical isolate samples over time for various countries, few studies have been done to identify genes and pathogens responsible for the antimicrobial resistance in clinical settings. This study conducted the first multivariate statistical analysis of the high-dimensional historical data from the NPDIB database for six different countries from majorly inhabited landmasses, including Australia, Brazil, China, South Africa, the UK, and the US. The similarities among different countries in terms of genes and pathogens were investigated to understand the potential avenues for antimicrobial-resistance gene spreading. The genes and pathogens that were closely involved in antimicrobial resistance were further studied temporally by plotting time profiles of their frequency to evaluate the trend of antimicrobial resistance. It was found that several of these significant genes (i.e., *aph(3?)-Ib*, *aph(6)-Id*, *blaTEM-1*, and *qacEdelta1*) are shared among all six countries studied. Based on the time profiles, a large number of genes and pathogens showed an increasing occurrence. The most shared pathogens responsible for carrying the most important genes in the six countries in the clinical setting were *Acinetobacter baumannii*, *E. coli* and *Shigella*, *Klebsiella pneumoniae* and *Salmonella enterica*. South Africa carried the least similar antimicrobial genes to the other countries in clinical isolates.

*Record Type:* Published Article

*Submitted To:* LAPSE (Living Archive for Process Systems Engineering)

*Citation (overall record, always the latest version):*

LAPSE:2020.0105

*Citation (this specific file, latest version):*

LAPSE:2020.0105-1

*Citation (this specific file, this version):*

LAPSE:2020.0105-1v1

*DOI of Published Version:* <https://doi.org/10.3390/pr7120964>

*License:* Creative Commons Attribution 4.0 International (CC BY 4.0)

Article

# An Analysis of Antimicrobial Resistance of Clinical Pathogens from Historical Samples for Six Countries

Karen Li <sup>1</sup>, Joanna Zheng <sup>1</sup>, Thomas Deng <sup>1</sup>, James Peng <sup>1</sup>, Dagmar Daniel <sup>1</sup>, Qian Jia <sup>2</sup> and Zuyi Huang <sup>1,\*</sup>

<sup>1</sup> Department of Chemical Engineering, Villanova University, Villanova, PA 19085, USA; karenjennyli@gmail.com (K.L.); joannazheng1209@gmail.com (J.Z.); thomasxdeng@gmail.com (T.D.); jamespeng624@gmail.com (J.P.); dagmardaniel8@gmail.com (D.D.)

<sup>2</sup> Department of Health, Nutrition & Exercise Sciences, Immaculata University, Immaculata, PA 19345, USA; qjia@immaculata.edu

\* Correspondence: zuyi.huang@villanova.edu; Tel.: +1-610-519-4848

Received: 2 October 2019; Accepted: 2 December 2019; Published: 17 December 2019



**Abstract:** The spread of antimicrobial resistance pathogens in humans has increasingly become an issue that threatens public health. While the NCBI Pathogen Detection Isolates Browser (NPDIB) database has been collecting clinical isolate samples over time for various countries, few studies have been done to identify genes and pathogens responsible for the antimicrobial resistance in clinical settings. This study conducted the first multivariate statistical analysis of the high-dimensional historical data from the NPDIB database for six different countries from majorly inhabited landmasses, including Australia, Brazil, China, South Africa, the UK, and the US. The similarities among different countries in terms of genes and pathogens were investigated to understand the potential avenues for antimicrobial-resistance gene spreading. The genes and pathogens that were closely involved in antimicrobial resistance were further studied temporally by plotting time profiles of their frequency to evaluate the trend of antimicrobial resistance. It was found that several of these significant genes (i.e., *aph(3'')-Ib*, *aph(6)-Id*, *blaTEM-1*, and *qacEdelta1*) are shared among all six countries studied. Based on the time profiles, a large number of genes and pathogens showed an increasing occurrence. The most shared pathogens responsible for carrying the most important genes in the six countries in the clinical setting were *Acinetobacter baumannii*, *E. coli* and *Shigella*, *Klebsiella pneumoniae* and *Salmonella enterica*. South Africa carried the least similar antimicrobial genes to the other countries in clinical isolates.

**Keywords:** antimicrobial resistance; clinical pathogens; principal component analysis; hierarchical clustering; data analysis

## 1. Introduction

In a post-antibiotic era, an increasing prevalence of antimicrobial resistance, often abbreviated as “AMR”, has been identified as a critical health hazard on a global scale. An evaluation of 1606 inpatient and outpatient samples from an African country in 2014 revealed that pathogenic resistance to older established antibiotics is rising; furthermore, if unchecked, the magnitude of mortality attributable to AMR may climb to 10 million lives by 2050 [1]. The inappropriate regulation and use of basic medications, including antibiotics, are frequently cited reasons for the spread of this phenomenon. Global human antibiotic consumption in 2000 was at 54,083,964,813 single dose units (also considered standard units, or SU’s) and had spiked by 36% within a decade [2], with the World Health Organization estimating that only half of prescribed antibiotics are currently used correctly [3]. The common cold, for instance, is incorrectly believed by a third of people to be treatable with antibiotics and is also the

most common condition for prescription [4]. Overexposure to antimicrobials enables pathogens to gain adaptive survivability via mutation or by acquiring resistance genes through horizontal transfer, in which plasmids containing genetic material disperse and pass between different species of bacteria.

Distinct regional variations in AMR occurrence have been observed to exist. For example, differences in methicillin-resistant *Staphylococcus aureus* (MRSA) rates were found among the United States, Canada, Latin America, and Europe [5]. Opportunely, the NCBI Pathogen Detection Isolates Browser (NPDIB), designed for studying genetic similarity and AMR repertoires in isolate profiles, has partnered with organizations like the CDC and WHO, allowing large volumes of data to be collected from tens of thousands of strains in the search for bacterial genomes with AMR genes [6]. Zhang et al. 2019 [7] presented the first study on the information from the NPDIB database. In particular, Zhang et al. analyzed the data for foodborne pathogen samples from six states in the US and studied the correlation between the location and resistance genes in foodborne pathogens. However, they did not incorporate a time-based analysis, but they did address gene transport in relation to the agricultural and livestock industries on a small scope.

The objective of this work was to investigate the status and trend of pathogens in the clinical setting and AMR genes relative to geographic location and time. Six different countries, each representing population-heavy regions of majorly inhabited landmasses, were selected and sampled: Brazil, China, Australia, South Africa, the United Kingdom, and the United States. Multivariate statistical analysis methods, i.e., principal component analysis (PCA) [8,9] and hierarchical clustering [10–12], were utilized to analyze high-dimensional clinical samples from the NPDIP database. The genes and pathogens that are heavily involved in antimicrobial resistance were identified for individual countries. From this, we aimed to study the status and trend of antimicrobial resistance for individual countries and then investigate the similarities and differences in time-based gene profiles between countries.

## 2. Materials and Methods

This work collected the data from the NCBI database, which was adapted into spreadsheets for multivariate statistical analysis. PCA was then performed on this data set on both pathogens and genes to represent the multidimensional data into a two-dimensional plot. This allows us to identify which gene or pathogen trends differed from the rest of the group, and could be deemed as ‘important’. The occurrence profiles over years for those important genes and pathogens were further used to study the trend of antimicrobial resistance. PCA and hierarchical clustering were also used to study the similarity of antimicrobial resistance in the six countries on the basis of the important genes and pathogens identified for those countries. The following hypotheses were tested in this study: (1) the microbial resistance in the clinical setting had an increasing trend over time; and (2) there were a certain amount of AMR genes and pathogens shared by different countries, which evidenced the occurrence of gene transfer between those selected countries.

### 2.1. Data Extracted from the NCBI Pathogen Detection Isolates Browser

The NCBI database provides the following information for each sample: the location of the sampling, time of the sampling, the name of the pathogen, antimicrobial resistant genotypes, isolation source, isolation type (i.e., clinical versus environmental samples), and others. The data utilized during this project were collected between 2010 and 2019, as most countries other than the US did not have data for other years in the NPDIP database. Clinical samples, which were sampled from human stool, blood, urine, sputum and other sources, were separately downloaded and pre-processed in MATLAB for each of the following six countries: Australia, Brazil, China, South Africa, the UK, and the US. Important pathogens and AMR genes were then identified from these samples. The pre-processed datasets were saved into Excel files for statistical analysis in the R language, which is one of the most commonly used tools in analyzing high-dimensional datasets. The data matrix imported into R for each individual country was organized in the following format: each row represents one clinical sample while each column provides the information for that sample. As seen in Figure 1, the information contained in the

columns included the scientific name of the pathogen detected in the sample, the sampling year, the isolation sources, the number of AMR genes detected in the sample (i.e., `number_amr_genes`), and whether or not genes were detected in the sample (with a value of 1 indicating the gene shown in the column was found in the sample). Typically, the data matrix for a country contains thousands of rows (i.e., samples) and one to two hundred columns (i.e., most columns used for representing genes). Therefore, the data matrix is typically of one to two hundred dimensions.

scientific_name	collection_date	number_amr_genes	aac(3)-I	aac(3)-II	aac(3)-IIa	aac(3)-IIb	aac(3)-IV	aac(3)-VIa	aac(6')-Ib	aac(6')-Ib'
1	2016	12	1	0	0	0	0	0	0	0
1	2016	15	1	0	0	0	0	0	0	0
1	2016	12	1	0	0	0	0	0	0	0
1	2016	12	1	0	0	0	0	0	0	0
1	2016	12	1	0	0	0	0	0	0	0
1	2016	12	1	0	0	0	0	0	0	0
1	2016	16	1	0	0	0	0	0	0	0
1	2016	16	1	0	0	0	0	0	0	0
1	2016	16	1	0	0	0	0	0	0	0
1	2016	16	1	0	0	0	0	0	0	0
1	2016	11	0	0	0	0	0	0	0	0
1	2016	16	0	0	0	0	0	0	0	0
1	2016	7	0	0	0	0	0	0	0	0
1	2016	11	0	0	0	0	0	0	0	0
1	2016	7	0	0	0	0	0	0	0	0
1	2016	7	0	0	0	0	0	0	0	0
1	2016	9	0	0	0	0	0	0	0	0
1	2016	8	0	0	0	0	0	0	0	0
1	2016	8	0	0	0	0	0	0	0	0
1	2016	8	0	0	0	0	0	0	0	0
1	2016	8	0	0	0	0	0	0	0	0
1	2016	8	0	0	0	0	0	0	0	0
1	2016	8	0	0	0	0	0	0	0	0

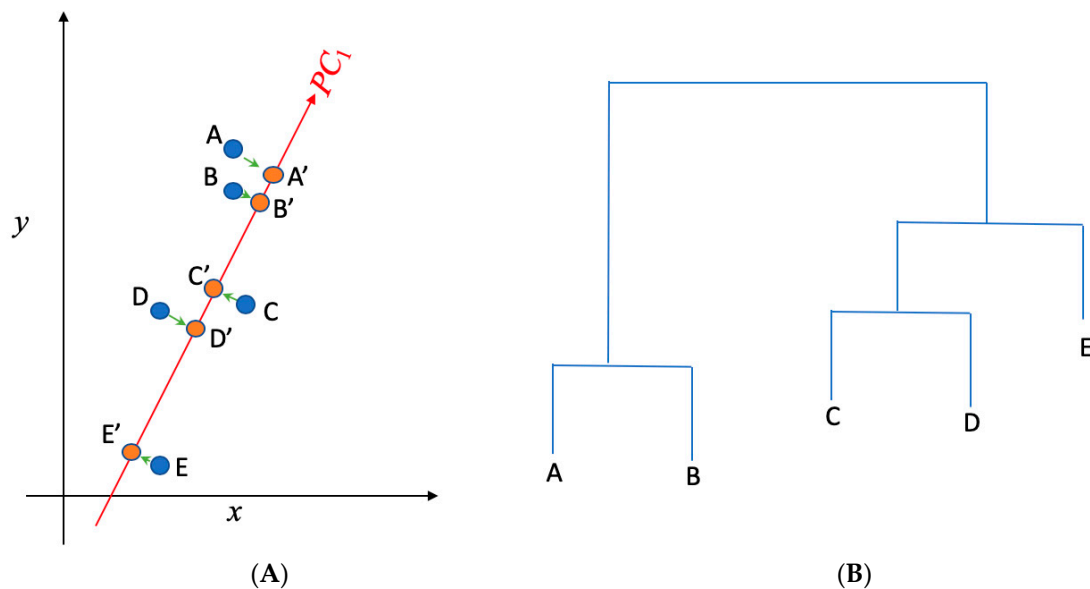
**Figure 1.** Snapshot of a section of the extracted data from the NCBI Pathogen Detection Isolates Browser for the US.

## 2.2. Principal Component Analysis (PCA) and Hierarchical Clustering

PCA is one of the most commonly used approaches to visualize a multidimensional data set in a two-dimensional space. The new coordinate dimensions (e.g., principle component 1, i.e., PC1 in Figure 2A) are the linear combination of the original coordinate dimensions (e.g.,  $x$  and  $y$  in Figure 2A). A lower principal component number, with 1 being the lowest, indicates a larger variance in the projections into that coordinate direction. In other words, more information about each of the points' distinguishability is contained in that projected coordinate direction. This allows for more efficient discernment of individual data points, as it is impossible to visualize the data in more than three dimensions. By projecting the data points onto a plot characterized by principle components 1 and 2, it is possible to visualize the variation and identify possible outliers; in the case of this experiment, that means the important pathogens or important genes. The function `prcomp()` from R was used to perform principal component analysis on the data matrix for each of the six countries. Before implementing PCA to identify important genes, a data matrix was created for PCA in such a way that each gene was represented in a single row and each pathogen was listed in a column to record the occurrences of each gene in each detected pathogen in the samples for each country. This data matrix was transposed to identify the pathogens that carried AMR genes.

Once the high-dimensional data was projected onto the PC1~PC2 space, hierarchical clustering was further used to determine the important genes and pathogens by finding those that are the least similar to the rest of the genes and pathogens. Depicted in Figure 2B is a graph of genes along with a hierarchical clustering of them. Hierarchical clustering is a tool that groups similar items (e.g., genes and pathogens) into groups called clusters. It does so by finding the Euclidean distance between all pairs of points. The program then finds the pair closest to each other and graphs it on the hierarchical clustering graph as a tree. The program then replaces the two points with a single point at the midpoint

of the two points mentioned and repeats this process slowly moving up the tree and replacing points until it comes down to a single point. From the hierarchy, it can be noted that branches that are higher up on the tree are the farthest from certain central branches. Data points associated with these branches are typically the outliers shown in the PC1~PC2 space, which correspond to the genes and pathogens that are mostly involved in antimicrobial resistance. The hierarchical clustering tree along with the outliers in the PC1~PC2 space was thus used in this work to identify both the important genes and important pathogens in the data.



**Figure 2.** (A) Principal component analysis is used to present the two dimensional (x,y) gene data into one dimensional space (PC1); (B) hierarchical clustering is implemented to cluster genes in (A) into separate groups. A, B, C, D, and E present the data in the x~y space in (A), while A', B', C', D', and E' present the projections of those data points in the PC1 space. While A, B, and C are not distinguishable in the x direction (as they have the similar projections onto x), they can be separated in the PC1 direction. PC1 represents the direction that returns the largest variance in the data projections onto it. Steps shown in (A) and (B) were used to separate and identify important AMR genes and clinical pathogens from the data.

On the basis of those genes identified as important for antimicrobial resistance in individual countries, the genes that were common in all six countries were further used to construct a data matrix in which each row corresponded to one country and each column represented one gene. The matrix recorded the occurrences of each gene in each country during 2010 to 2019. PCA and hierarchical clustering were then implemented to the data matrix to cluster the six countries and study the similarity of those countries in carrying those AMR genes. Similarly, a data matrix was built for the occurrences of those common pathogens that were found in the six countries. In other words, each column in the data matrix represented one clinical pathogen, while each row was associated with one of the six countries. The similarity of the six countries in carrying those important clinical pathogens was then studied from the results returned by PCA and hierarchical clustering.

### 2.3. Investigation of the Trend of Antimicrobial Resistance

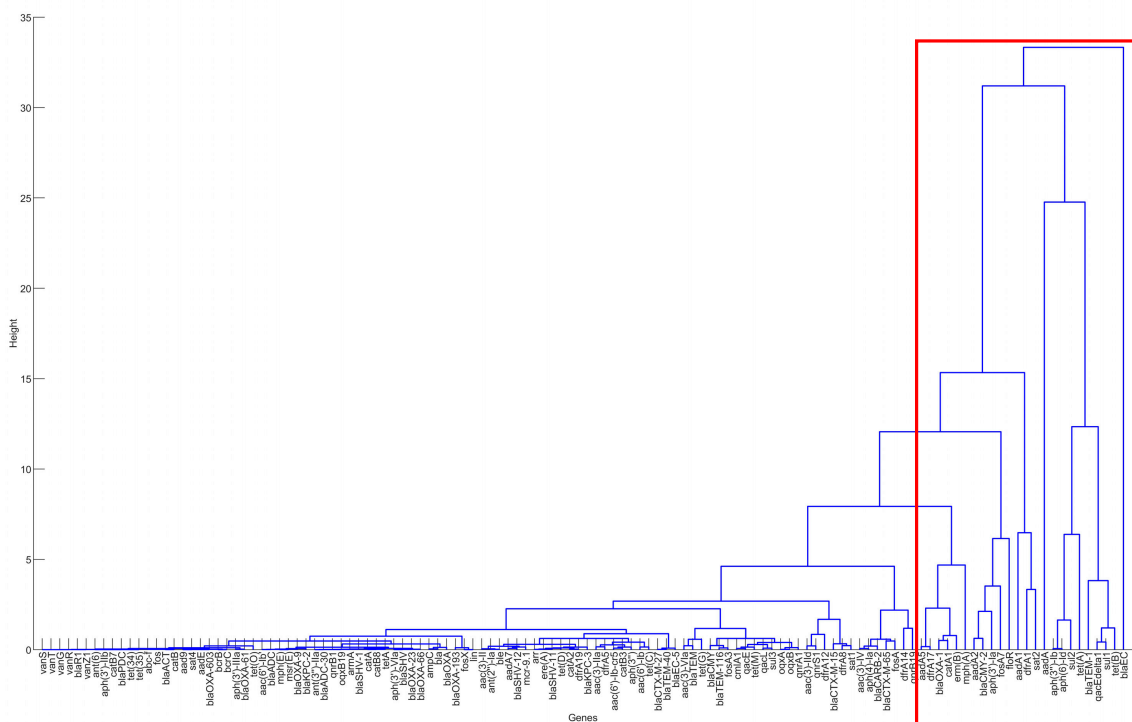
The historical profiles of the genes and pathogens mostly involved in antimicrobial resistance were plotted to study the trend of antimicrobial resistance. In particular, the historical profiles of those important genes and pathogens were plotted over years. In order to account for differences in the number of samples collected each year, we normalized these historical profiles by dividing the total

occurrences of genes/pathogens that were detected in that year with the number of samples obtained in the same year. This can resolve the bias due to the fact that some years have more samples than others.

### 3. Results

#### 3.1. Identification of Important Genes in Clinical Samples from Australia, Brazil, China, South Africa, the UK, and the US

The PCA and clustering approaches described in the Materials and Methods section were used to identify significant resistance genes from the six selected countries. The projection and detailed visualization of data, made possible by hierarchical clustering, enabled the schematic comparison of those AMR genes. The result of hierarchical clustering for the US is shown in Figure 3 as an example for illustration. As there are more than 100 AMR genes in the dataset for the United States, those genes, especially those with few occurrences, are lumped together in the PC1~PC2 space. It is thus challenging to show all these genes clearly in the PC1~PC2 space. That is why only the result of hierarchical clustering is shown below. The genes that were heavily tested in the antimicrobial-resistance pathogens are typically those outliers in the PC1~PC2 space and those in the top few levels of branches in the hierarchical clustering tree. For example, the genes in the branches excluded by the red rectangle were identified as the genes important for antimicrobial resistance in the pathogens in the clinical samples. A similar approach was applied to identify important AMR genes for the other five countries. Table 1 summarizes the important AMR genes for the six countries. It can be seen that Brazil has the largest amount of important AMR genes, while South Africa harbors a significantly lower quantity. The genes that are shared by five and six countries are marked in the colors of blue and red, respectively. They can be generally categorized into: aminoglycoside phosphotransferase (*aph(3'')-Ib* and *aph(6)-Id*), beta-lactamase (*blaEC* and *blaTEM-1*), macrolide 2'-phosphotransferase (*mph(A)*), quaternary ammonium compound-resistance protein (*qacEdelta1*), and sulfonamide resistance genes (*sul1* and *sul2*).

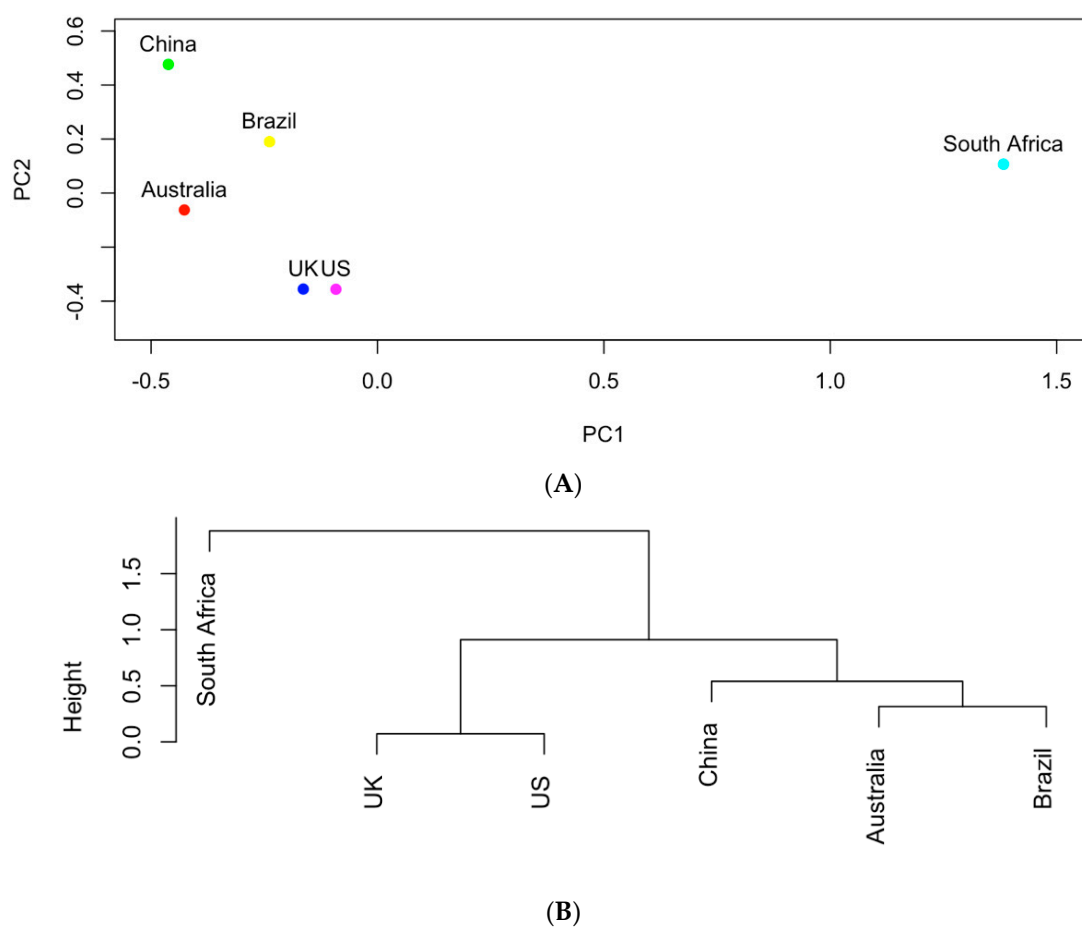


**Figure 3.** The result of hierarchical clustering for the genes in the clinical samples for the United States, 2010–2019.

**Table 1.** The genes mostly involved in antimicrobial resistance in the six countries. Red text indicates the genes shared by all six countries, while blue denotes genes common to five countries.

Australia	Brazil	China	South Africa	UK	US
<i>aac(3)-I</i>	<i>aac(3)-IIa</i>	<i>aadA2</i>	<i>aac(2′)-Ic</i>	<i>aadA</i>	<i>aadA</i>
<i>aadA1</i>	<i>aac(6′)-Ib</i>	<i>ant(3′′)-IIa</i>	<i>aph(3′′)-Ib</i>	<i>aadA1</i>	<i>aadA1</i>
<i>aadA5</i>	<i>aadA2</i>	<i>aph(3′′)-Ib</i>	<i>aph(6)-Id</i>	<i>aph(3′′)-Ib</i>	<i>aadA2</i>
<i>ant(3′′)-IIa</i>	<i>aadA7</i>	<i>aph(3′)-Ia</i>	<i>blaA</i>	<i>aph(6)-Id</i>	<i>aadA5</i>
<i>aph(3′′)-Ib</i>	<i>aph(3′′)-Ib</i>	<i>aph(6)-Id</i>	<i>blaCTX-M-15</i>	<i>blaEC</i>	<i>aph(3′′)-Ib</i>
<i>aph(3′)-Ia</i>	<i>aph(3′)-IIb</i>	<i>armA</i>	<i>blaTEM-1</i>	<i>blaOXA-1</i>	<i>aph(3′)-Ia</i>
<i>aph(6)-Id</i>	<i>aph(3′)-Ia</i>	<i>blaEC</i>	<i>erm(37)</i>	<i>blaTEM-1</i>	<i>aph(6)-Id</i>
<i>blaADC-30</i>	<i>aph(6)-Id</i>	<i>blaKPC-2</i>	<i>fosX</i>	<i>catA1</i>	<i>blaCMY-2</i>
<i>blaEC</i>	<i>blaCTX-M-15</i>	<i>blaOXA-23</i>	<i>lin</i>	<i>dfrA1</i>	<i>blaEC</i>
<i>blaOXA-23</i>	<i>blaEC</i>	<i>blaOXA-66</i>	<i>qacEdelta1</i>	<i>mph(A)</i>	<i>blaOXA-1</i>
<i>blaOXA-66</i>	<i>blaKPC-2</i>	<i>blaSHV-11</i>	<i>sul1</i>	<i>qacEdelta1</i>	<i>blaTEM-1</i>
<i>blaTEM-1</i>	<i>blaOXA</i>	<i>blaTEM-1</i>	<i>sul2</i>	<i>sat2</i>	<i>catA1</i>
<i>dfrA1</i>	<i>blaOXA-494</i>	<i>dfrA12</i>		<i>sul1</i>	<i>dfrA1</i>
<i>dfrA17</i>	<i>blaOXA-56</i>	<i>floR</i>		<i>sul2</i>	<i>dfrA17</i>
<i>erm(B)</i>	<i>blaPDC-5</i>	<i>fosA</i>		<i>tet(A)</i>	<i>erm(B)</i>
<i>mph(A)</i>	<i>blaSHV-11</i>	<i>mcr-1.1</i>		<i>tet(B)</i>	<i>floR</i>
<i>qacEdelta1</i>	<i>blaSPM-1</i>	<i>mph(A)</i>			<i>fosA7</i>
<i>sat2</i>	<i>blaTEM-1</i>	<i>mph(E)</i>			<i>mph(A)</i>
<i>sul1</i>	<i>catA1</i>	<i>msr(E)</i>			<i>qacEdelta1</i>
<i>sul2</i>	<i>catB7</i>	<i>oqxA</i>			<i>sat2</i>
<i>tet(A)</i>	<i>cmx</i>	<i>oqxB</i>			<i>sul1</i>
<i>tet(B)</i>	<i>crpP</i>	<i>qacEdelta1</i>			<i>sul2</i>
	<i>dfrA12</i>				<i>tet(A)</i>
	<i>fosA</i>				<i>tet(B)</i>
	<i>mph(A)</i>				
	<i>oqxA</i>				
	<i>oqxB</i>				
	<i>qacEdelta1</i>				
	<i>rmtD1</i>				
	<i>sul1</i>				
	<i>sul2</i>				
	<i>tet(A)</i>				

The genes shown in Table 1 were further used to study the similarities of the six countries in carrying those AMR genes in the clinical samples. In particular, the occurrences of these genes were recorded in a data matrix in which each row represented one country and each column is associated with one gene. Figure 4 shows the results of principal component analysis and hierarchical clustering on the data matrix. The UK and the US are the most similar in terms of genes with a total of 16 genes in common (as shown in Table 1). The *aadA* genes produce resistance to antibiotics streptomycin and spectomycin [13]. The *sul* genes produce resistance to sulfonamides [14] and *tet* genes produce resistance to tetracyclines [15]. All of the genes found in the UK were also found in the US. Australia and Brazil also are similar to each other based on the hierarchical clusters. South Africa appears to be an outlier among the countries in terms of genes. It tends to share the least number of genes in common with other countries. South Africa and China share merely four genes (*aph(3′′)-Ib*, *aph(6)-Id*, *blaTEM-1*, *qacEdelta1*), which are also the same genes shared by all countries.



**Figure 4.** (A) The projection of the six countries onto the PC1~PC2 space based on AMR genes; (B) hierarchical clustering analysis of the similarities of the six countries in carrying AMR genes.

### 3.2. Identification of Important Clinical Pathogens That Carried Antimicrobial Resistance Genes in Australia, Brazil, China, South Africa, the UK, and the US

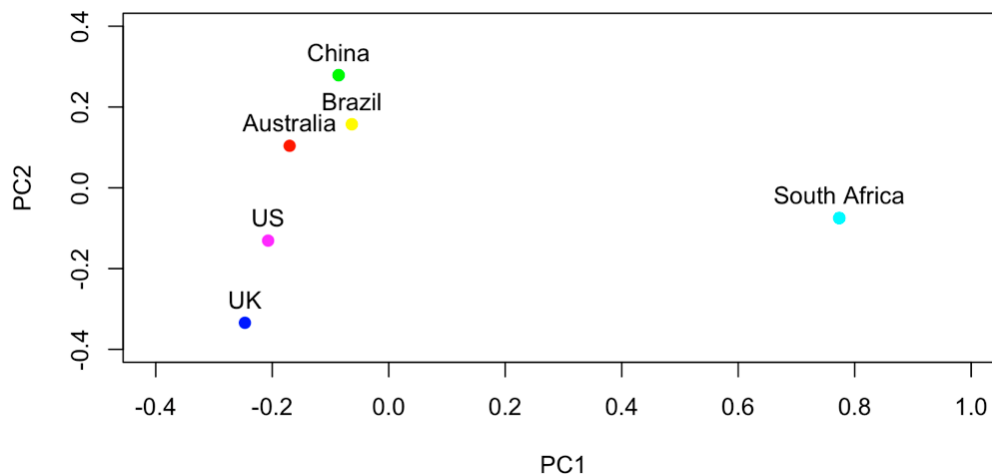
In addition to studying the genes mostly involved in antimicrobial resistance, PCA and hierarchical clustering were performed on the datasets to identify the clinical pathogens mostly involved in carrying AMR genes within each country. Table 2 below lists the important pathogens from the six countries. Seven pathogens, including *Acinetobacter baumannii*, *E. coli* and *Shigella*, *Enterobacter*, *Klebsiella pneumoniae*, *Listeria monocytogenes*, *Mycobacterium tuberculosis*, and *Salmonella enterica*, were found to be important to more than one of the analyzed countries, with the exception of *Vibrio cholerae* and *Pseudomonas aeruginosa* (the former only found important in China and the latter in Brazil). *E. coli* and *Shigella* and *Klebsiella pneumoniae* are the two pathogens that are common in all six countries. *Acinetobacter baumannii* and *Salmonella enterica* are shared by all countries other than South Africa. *Mycobacterium Tuberculosis* is important only in South Africa and Australia.

The occurrence of the important pathogens listed in Table 2 were further used to study the similarities of the six selected countries in carrying those clinical pathogens. It can be seen in Figure 5 that South Africa shows the least similarity in its clinical pathogens to the other five countries. Following South Africa, China contains the least similar clinical pathogens to the other four countries. Compared to Austria and Brazil, the UK is clustered with the US in the same group in the hierarchical clustering tree in Figure 5B. This indicates that the UK and the US share more similar clinical pathogens than the other selected countries.

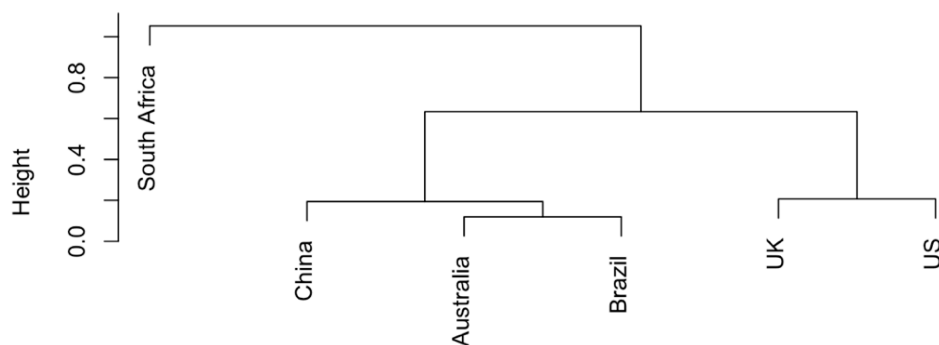


**Table 2.** The important pathogens for the six countries. Pathogens were identified through hierarchical clustering.

Australia	Brazil	China	South Africa	UK	US
<i>Acinetobacter baumannii</i>	<i>Acinetobacter baumannii</i>	<i>Acinetobacter baumannii</i>	<i>E. coli</i> and <i>Shigella</i>	<i>E. coli</i> and <i>Shigella</i>	<i>Acinetobacter baumannii</i>
<i>E. coli</i> and <i>Shigella</i>	<i>E. coli</i> and <i>Shigella</i>	<i>E. coli</i> and <i>Shigella</i>	<i>Enterobacter</i>	<i>Enterobacter</i>	<i>E. coli</i> and <i>Shigella</i>
<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>	<i>Enterobacter</i>	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>	<i>Enterobacter</i>
<i>Mycobacterium tuberculosis</i>	<i>Pseudomonas aeruginosa</i>	<i>Klebsiella pneumoniae</i>	<i>Listeria monocytogenes</i>	<i>Salmonella enterica</i>	<i>Klebsiella pneumoniae</i>
<i>Salmonella enterica</i>	<i>Salmonella enterica</i>	<i>Salmonella enterica</i>	<i>Mycobacterium Tuberculosis</i>		<i>Salmonella enterica</i>
<i>Vibrio cholerae</i>					



(A)



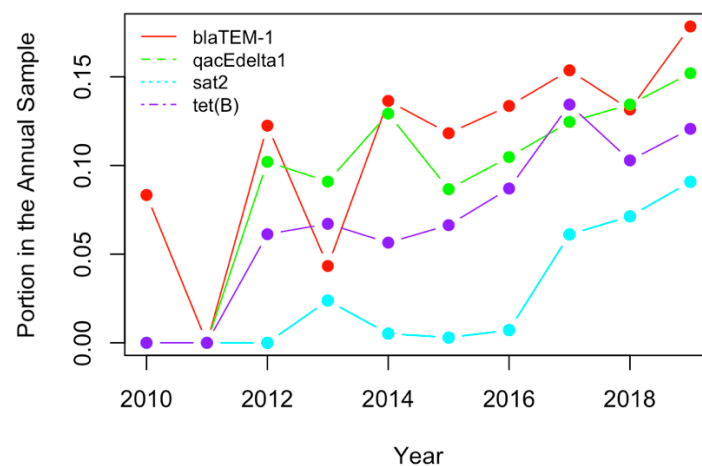
(B)

**Figure 5.** (A) PCA analysis of countries based on the occurrence data of clinical pathogens listed in Table 2; (B) hierarchical clusters of countries based on the occurrence data of clinical pathogens listed in Table 2.

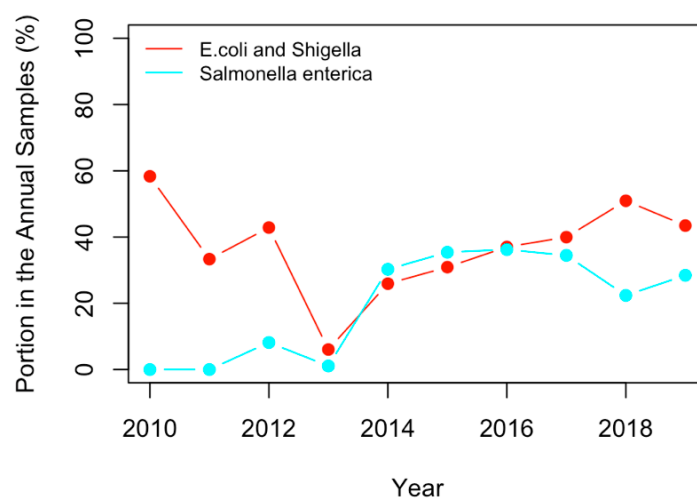
### 3.3. Investigation of the Trend in Antimicrobial Resistance from the Historical Profiles of Genes and Pathogens that Were Mostly Involved in Antimicrobial Resistance in Australia, Brazil, China, South Africa, the UK, and the US

The historical occurrence profiles of the important genes and pathogens identified in Sections 3.1 and 3.2 were plotted in this section to investigate the trend of antimicrobial resistance in the six selected

countries. The genes and pathogens with an increasing trend were then identified from their profiles. For example, Figure 6A shows the profiles of the four antimicrobial genes with an increasing trend in the US dataset, while Figure 6B illustrates the profiles of the important clinical pathogens with a generally increasing trend in the US in recent years. In particular, the portion of *E. coli* and *Shigella* in the annual clinical samples went down from 2010 to 2013 and then generally went up after 2013. This pathogen was regarded as a pathogen with an increasing occurrence recently. The occurrence of *Salmonella enterica* generally increased from 2013, although fluctuation was observed over time. Similar profiles were obtained for the important pathogens from the other five countries, and pathogens showing a similar historical trend were regarded as pathogens with increasing occurrences. Tables 3 and 4 list the genes and pathogens from Tables 1 and 2 that show an increasing trend for each country.



(A)



(B)

**Figure 6.** (A) The historical profiles of the AMR genes with a generally increasing trend in their occurrence in the clinical samples collected in the US; (B) the historical profiles of the clinical pathogens with a generally increasing trend in their occurrence in the clinical samples collected in the US.

Table 3 below displays the important genes listed in Table 1 that show generally increasing occurrences in the six countries. Among those genes shown in Table 1, the portions of genes with increasing occurrence over time for individual countries are: 54.5% for Australia, 56.3% for Brazil, 50% for China, 16.7% for South Africa, 75% for UK and 12.5% for US. While each country has certain

genes with increasing occurrence, South Africa and the US have the lowest portions of genes with increasing trends.

**Table 3.** The important genes from Table 1 with a generally increasing occurrence.

China	South Africa	Brazil	Australia	UK	US
<i>aadA2</i>	<i>fosX</i>	<i>aadA7</i>	<i>aac(3)-I</i>	<i>aadA1</i>	<i>blaTEM-1</i>
<i>blaEC</i>	<i>lim</i>	<i>aph(3'')-Ib</i>	<i>blaADC-30</i>	<i>aph(3'')-Ib</i>	<i>qacEdelta1</i>
<i>blaKPC-2</i>		<i>aph(3')-IIb</i>	<i>blaOXA-66</i>	<i>aph(6)-Id</i>	<i>sat2</i>
<i>blaSHV-11</i>		<i>aph(3')-Ia</i>	<i>aadA1</i>	<i>blaTEM-1</i>	
<i>blaTEM-1</i>		<i>aph(6)-Id</i>	<i>aph(3'')-Ib</i>	<i>dfrA1</i>	
<i>dfrA12</i>		<i>blaCTX-M-15</i>	<i>aph(6)-Id</i>	<i>mph(A)</i>	
<i>floR</i>		<i>blaOXA-49</i>	<i>sul2</i>	<i>qacEdelta1</i>	
<i>fosA</i>		<i>blaPDC-5</i>	<i>aadA5</i>	<i>sat2</i>	
<i>oqxA</i>		<i>blaSHV-11</i>	<i>dfrA17</i>	<i>sul1</i>	
<i>oqxB</i>		<i>blaSPM-1</i>	<i>dfrA1</i>	<i>sul2</i>	
<i>qacEdelta1</i>		<i>catB7</i>	<i>erm(B)</i>	<i>tet(A)</i>	
		<i>crpP</i>	<i>sat2</i>	<i>tet(B)</i>	
		<i>dfrA1</i>			
		<i>fosA</i>			
		<i>mph(A)</i>			
		<i>oqxB</i>			
		<i>rmtD1</i>			
		<i>sul1</i>			

Table 4 lists the clinical pathogens listed in Table 2 with increasing occurrence over time. Among those important pathogens shown in Table 2, the portions of clinical pathogens with increasing occurrence are: 40% for Australia, 40% for Brazil, 33.3% for China, 40% for South Africa, 0% for the UK and 40% for the US. This implies that 30 to 40% of the important clinical pathogens in most countries (with the exception of the UK) generally had increasing occurrences in the last 9 years.

**Table 4.** The important pathogens from Table 2 with an increasing occurrence.

Australia	Brazil	China	South Africa	UK	US
<i>Acinetobacter baumannii</i>	<i>Klebsiella pneumoniae</i>	<i>E. coli</i> and <i>Shigella</i>	<i>Listeria monocytogenes</i>	None	<i>E. coli</i> and <i>Shigella</i>
<i>E. coli</i> and <i>Shigella</i>	<i>Pseudomonas aeruginosa</i>	<i>Klebsiella pneumoniae</i>	<i>Mycobacterium Tuberculosis</i>	None	<i>Salmonella enterica</i>

## 4. Discussions and Conclusions

### 4.1. Common Genes and Pathogens in the Six Studied Countries

The important genes shown in Table 1 represent the AMR genes most often carried by pathogens in a particular country. It was further analyzed to identify the AMR genes unique to each studied country (Table 5). Genes *aac(3)-I* and *blaADC-30* are both unique to Australia. The genes unique to Brazil include *aac(3)-IIa*, *aac(6')-Ib*, *aadA7*, *aph(3')-IIb*, *blaCTX-M-15*, *blaOXA*, *blaOXA-494*, *blaOXA-56*, *blaPDC-5*, *blaSPM-1*, *catB7*, *cmx*, *crpP*, and *rmtD1*. For China, there are the unique genes *armA*, *blaSHV-11*, *mcr-1.1*, *mph(E)*, and *msr(E)*. South Africa contains the unique genes *aac(2')-Ic*, *blaA*, *erm37*, *fosX*, and *lin*. The UK does not have any unique genes. Lastly, the US has the unique genes *blaCMY-2* and *fosA7*.

**Table 5.** Genes unique to each country.

Australia	Brazil	China	South Africa	United Kingdom	United States
<i>aac(3)-I</i>	<i>aac(3)-IIa</i>	<i>armA</i>	<i>aac(2')-Ic</i>		<i>blaCMY-2</i>
<i>blaADC-30</i>	<i>aac(6')-Ib</i>	<i>blaSHV-11</i>	<i>blaA</i>		<i>fosA7</i>
	<i>aadA7</i>	<i>mcr-1.1</i>	<i>erm37</i>		
	<i>aph(3')-IIb</i>	<i>mph(E)</i>	<i>fosX</i>		
	<i>blaCTX-M-15</i>	<i>msr(E)</i>	<i>lin</i>		
	<i>blaOXA</i>				
	<i>blaOXA-494</i>				
	<i>blaOXA-56</i>				
	<i>blaPDC-5</i>				
	<i>blaSPM-1</i>				
	<i>catB7</i>				
	<i>cmx</i>				
	<i>crpP</i>				
	<i>rmtD1</i>				

The important genes shared by three or more countries in Table 1 were further identified and are listed in Table 6. They can be categorized by function: antibiotic-degrading enzymes, efflux pump, and modified antibiotic binding target. The majority of the genes produce antibiotic-degrading enzymes (*aadA1*, *aadA2*, *aph(3'')-Ib*, *aph(3')-Ia*, *aph(6)-Id*, *blaEC*, *blaTEM-1*, *catA1*, *mph(A)*, *sat2*). Three of these common genes produce efflux pumps (*qacEdelta1*, *tet(A)*, *tet(B)*), and one causes a modified antibiotic binding target (*dfrA1*, *sul1*, *sul2*). All countries share the genes *aph(3'')-Ib*, *aph(6)-Id*, *blaTEM-1*, and *qacEdelta1*. In particular, *aph* genes produce aminoglycoside modifying enzymes that catalyze ATP-dependent phosphorylation of a hydroxyl group [16]. *bla* genes produce beta-lactamases, which hydrolyze the beta-lactam bond in susceptible beta-lactam antibiotics [17]. The gene *qacEdelta1* produces resistance to quaternary ammonium compounds [18].

**Table 6.** Genes carried by different countries sorted by frequency. Only genes with a frequency of 3 or higher are shown.

	Australia	Brazil	China	South Africa	UK	US	Total
<i>aph(3'')-Ib</i>	1	1	1	1	1	1	6
<i>aph(6)-Id</i>	1	1	1	1	1	1	6
<i>blaTEM-1</i>	1	1	1	1	1	1	6
<i>qacEdelta1</i>	1	1	1	1	1	1	6
<i>blaEC</i>	1	1	1	0	1	1	5
<i>mph(A)</i>	1	1	1	0	1	1	5
<i>sul1</i>	1	1	0	1	1	1	5
<i>sul2</i>	1	1	0	1	1	1	5
<i>aph(3')-Ia</i>	1	1	1	0	0	1	4
<i>tet(A)</i>	1	1	0	0	1	1	4
<i>aadA1</i>	1	0	0	0	1	1	3
<i>aadA2</i>	0	1	1	0	0	1	3
<i>catA1</i>	0	1	0	0	1	1	3
<i>dfrA1</i>	1	0	0	0	1	1	3
<i>sat2</i>	1	0	0	0	1	1	3
<i>tet(B)</i>	1	0	0	0	1	1	3

In addition to the AMR genes shared by the six selected countries, all countries also share the pathogens *E. coli* and *Shigella*, and *Klebsiella pneumoniae*, as seen in Table 7. *E. coli* and *Shigella* were found to be responsible for carrying the genes *aph(3'')-Ib*, *aph(6)-Id*, *blaEC*, *blaTEM-1*, *mph(A)*, *qacEdelta1*, *sul1*, and *sul2* across the six countries. *Klebsiella pneumoniae* is responsible for carrying the same eight genes as *E. coli* and *Shigella*, excluding *blaEC*. In addition, *Salmonella enterica*, which is shared by five countries, is largely responsible for carrying the same eight genes as *E. coli* and *Shigella*, excluding *blaEC*.

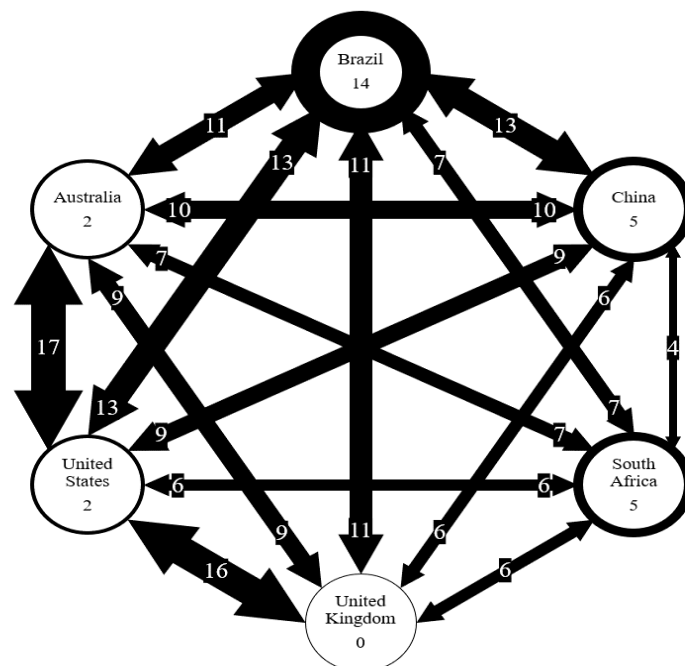
and *mph(A)*. The pathogens *Listeria monocytogenes*, *Mycobacterium tuberculosis*, *Pseudomonas aeruginosa*, and *Vibrio cholerae* are unique to each country. These unique pathogens were not found to be the ones mostly carrying AMR genes. The pathogens shared by the most countries were also found to carry the greatest number of AMR genes.

**Table 7.** Important pathogens carrying AMR genes from different countries sorted by frequency (Red: shared by all six countries; Orange - shared by five countries; Green: shared by four countries; and Blue: unique to one country).

	Australia	Brazil	China	South Africa	UK	US	Total
<i>E. coli</i> and <i>Shigella</i>	1	1	1	1	1	1	6
<i>Klebsiella pneumoniae</i>	1	1	1	1	1	1	6
<i>Salmonella enterica</i>	1	1	1	0	1	1	5
<i>Acinetobacter baumannii</i>	1	1	1	0	0	1	4
<i>Enterobacter</i>	0	0	1	1	1	1	4
<i>Listeria monocytogenes</i>	0	0	0	1	0	0	1
<i>Mycobacterium tuberculosis</i>	1	0	0	0	0	0	1
<i>Mycobacterium Tuberculosis</i>	0	0	0	1	0	0	1
<i>Pseudomonas aeruginosa</i>	0	1	0	0	0	0	1
<i>Vibrio cholerae</i>	0	0	1	0	0	0	1

#### 4.2. Similarities of the Six Selected Countries in Carrying AMR Genes and Clinical Pathogens

Figures 4 and 5 show that the six countries are clustered into three groups according to the AMR genes and pathogens carried in them: the US and the UK in one group, Australia and Brazil and China in one group, and South Africa in one group. The number of AMR genes that are shared between every two countries is listed in Figure 7. In contrast to Figure 4, the occurrence frequencies of those shared genes are not reflected in Figure 7. In spite of this, Figure 7 shows that the US and the UK contain a large number of AMR genes and that South Africa has the lowest number of AMR genes in common with other countries.



**Figure 7.** Number of AMR genes shared between six countries. Number on arrows represent the number of genes shared. Thickness of arrow correlated to the number of genes shared. Number in circles represents the number of genes unique to each country.

Some countries may be more similar than others due to the number of interactions that occur in between the countries. For example, the US and UK are the most similar based on AMR genes. The US imported \$60,783 million goods from the UK and exported \$66,312 million goods to the UK in 2018 [19]. On the other hand, South Africa is the least similar to the US, and the US trade in goods with South Africa in 2018 included \$5,517 million in imports and \$8,467 million in exports [19]. Interactions, like trade, may allow for the transfer of genes between countries as they can be carried along with the traded goods. For example, genes carried by microorganisms in livestock products (e.g., meats) can be transferred between countries. These microorganisms then spread genes to other microorganisms in human bodies or in the environment.

Climate factors may be a contributing factor to the countries' similarities. The US and UK both share similar conditions with a mostly temperature climate. The similarity in conditions would allow for similar species of pathogens to be present in the different countries. In addition, Australia and Brazil also share similar annual average temperatures, which could be a contributing factor to their similarity based on pathogens. Another factor contributing to similarity could be the amount of tourism or travel to the countries. High levels of travel to certain countries could cause pathogen or gene transfer among tourists and residents of the country, which would lead to similarities in pathogens and genes to other countries. For example, the US and the UK are both contained in a cluster that does not include South Africa. Both the UK and the US have high levels of tourism (the UK with 36 million international tourist arrivals in 2016 and the US with 76 million international tourist arrivals in 2016 [20]), while South Africa has a lower level (10 million international tourist arrivals in 2016 [20]). The differences in the levels of travel could have contributed to the differences in cluster placement.

#### 4.3. The Implication of AMR Gene Transfer in the Studied Countries

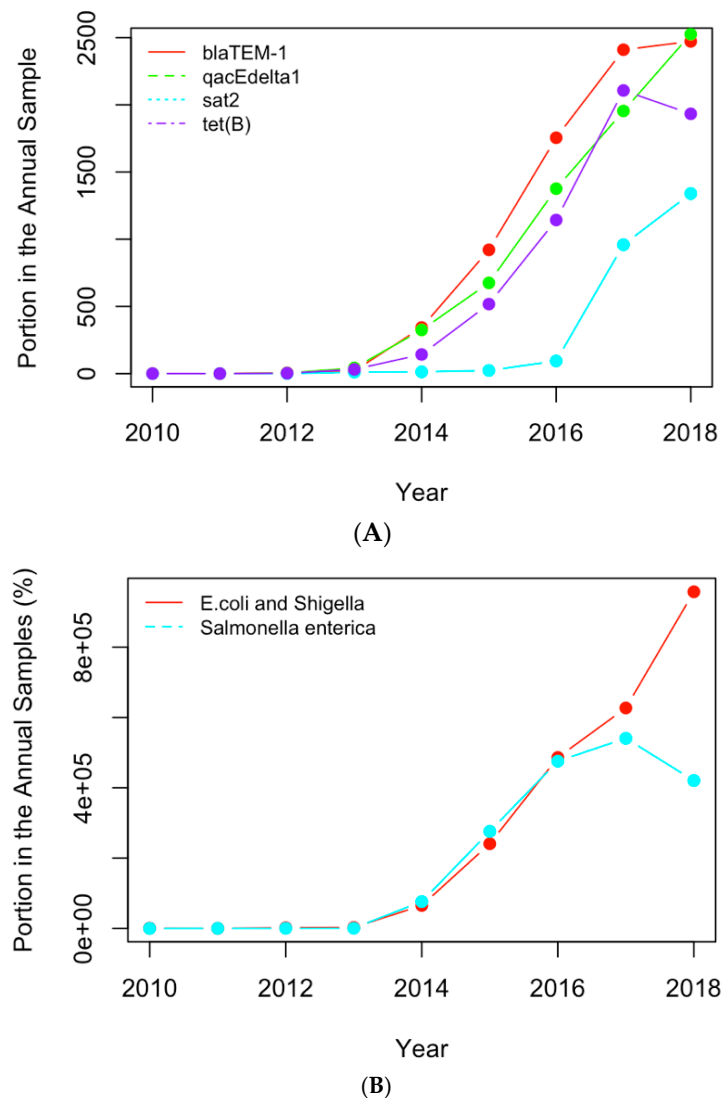
Due to the number of shared AMR genes and pathogens found across the studied countries and the comparatively small number of unique ones, it can be inferred that pathogenic similarities are occurring between continents. This gene transfer among pathogens of different countries would result in increased occurrences of AMR genes and increased similarity among the countries. This validated our hypothesis that gene transfer existed between the countries studied. The important genes shared by at least five countries were studied to determine which pathogens carry and transfer those genes. This was determined by identifying which pathogens had the highest frequency of those particular genes. The genes were found to be carried by four pathogens: *Acinetobacter baumannii*, *E. coli* and *Shigella*, *Klebsiella pneumoniae* and *Salmonella enterica*. *Acinetobacter baumannii* carries the genes *aph(3'')-Ib* and *aph(6)-Id*. *E. coli* and *Shigella* carries the genes *aph(3'')-Ib*, *aph(6)-Id*, *aph(3'')-Ib*, *blaEC*, *blaTEM-1*, *mph(A)*, *qacEdelta1*, *sul1*, and *sul2*. *Klebsiella pneumoniae* carries the genes *aph(3'')-Ib*, *aph(6)-Id*, *aph(3'')-Ib*, *blaTEM-1*, *mph(A)*, *qacEdelta1*, *sul1*, and *sul2*. *Salmonella enterica* carries the genes *aph(3'')-Ib*, *aph(6)-Id*, *aph(3'')-Ib*, *blaTEM-1*, *qacEdelta1*, *sul1*, and *sul2*.

Gene transfer could be due to increased travel and migration between countries, as passenger numbers are predicted to increase at more than 5% per year in the next 10 years [5]. With the continued rise in globalization and travel, it is becoming easier for pathogens to travel across countries and interact with other pathogens. These interactions include gene transfer, which can result in new resistance and diseases. A study in 2016 found that new antimicrobial resistance in colonizing *E. coli* was associated with international travel [7].

#### 4.4. The Trend of Antimicrobial Resistance Indicated from the Historical Profiles of AMR Genes and Clinical Pathogens

The normalized historical profiles of the AMR genes and clinical pathogens shown in Figure 6 are replotted with the actual numbers of samples containing those genes and pathogens in Figure 8. The year 2019 was not considered in the analysis of the historical profiles because data still has yet to be collected for the remainder of the year 2019. It can be seen that the numbers of samples are low in early years when compared to those for recent years. This may be due to the following reasons: (1) people

did not collect samples as often in the beginning of this decade; (2) limited resources were available to collect and input the samples into the NPDIB database; (3) the antimicrobial-resistance became more severe and thus caused more resistance cases. Therefore, the increase in the actual numbers of samples over the years may not be mainly due to the increase in antimicrobial resistance. That is why the normalized profiles (i.e., Figure 6) were used in this work to evaluate the trend of antimicrobial resistance from the data.



**Figure 8.** Replotting Figure 6 with the actual numbers of samples containing the AMR genes (A) and clinical pathogens (B) in USA.

Based on the normalized values of the gene frequency over time, certain portions of important AMR genes listed in Table 1 showed an increasing occurrence. In particular, 54.5% AMR genes for Australia, 56.3% AMR genes for Brazil, 50% AMR genes for China, 16.7% AMR genes for South Africa, 75% AMR genes for the UK and 12.5% AMR genes for the US showed an increasing occurrence, especially in the recent years from 2014 to 2018. When the absolute values are used, the increasing trend is even more apparent. The trend is contributing to the worsening AMR epidemic that is posing a critical health hazard on a global scale. It can be correlated to the global increase in antibiotic consumption in recent years. One study involving 76 countries found that global consumption increased 65% and the antibiotic consumption rate increased 39% between 2000 and 2015 [21]. The increase in AMR genes could be linked to increased travel between countries. International tourist arrivals increased by 6%

from 2017 to 2018 [22]. One study found a correlation between new AMR and international travel, with 9% of participants acquiring new strains of ESBL-producing *E. coli* [23].

**Author Contributions:** Conceptualization, Q.J. and Z.H.; methodology, K.L., J.Z., T.D., J.P., D.D., Q.J., and Z.H.; formal analysis, K.L., J.Z., T.D., J.P., D.D., Q.J. and Z.H.; resources, Q.J. and Z.H.; data curation, Q.J. and Z.H.; writing—original draft preparation, K.L., J.Z., T.D., J.P., and D.D.; writing—review and editing, Q.J., and Z.H.; supervision, Q.J. and Z.H.; project administration, Q.J., and Z.H.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- O'Neill, J. *Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations*; The Review on Antimicrobial Resistance: London, UK, 2014.
- Van Boeckel, T.P.; Gandra, S.; Ashok, A.; Caudron, Q.; Grenfell, B.T.; Levin, S.A.; Laxminarayan, R. Global Antibiotic Consumption 2000 to 2010: An Analysis of Cross Mark 742 National Pharmaceutical Sales Data. *Lancet Infect. Dis.* **2014**, *14*, 742–750. [CrossRef]
- Bank, W. *Drug-Resistant Infections: A Threat to Our Economic Future*; World Bank: Washington, DC, USA, March 2017.
- McNulty, C.A.; Boyle, P.; Nichols, T.; Clappison, P.; Davey, P. The public's attitudes to and compliance with antibiotics. *J. Antimicrob. Chemother.* **2007**, *60*, 163–168. [CrossRef] [PubMed]
- Diekema, D.J.; Pfaller, M.; Shortridge, D.; Zervos, M.; Jones, R.N. Twenty-year trends in antimicrobial susceptibilities among *Staphylococcus aureus* from the SENTRY antimicrobial surveillance program. *Open Forum Infect. Dis.* **2019**, *6*, S47–S53. [CrossRef]
- Available online: <https://www.ncbi.nlm.nih.gov/pathogens/about/> (accessed on 15 August 2019).
- Zhang, N.N.; Liu, E.; Tang, A.; Ye, M.C.; Wang, K.; Jia, Q.; Huang, Z.Y. Data-Driven Analysis of Antimicrobial Resistance in Foodborne Pathogens from Six States within the US. *Int. J. Environ. Res. Public Health* **2019**, *16*, 4835. [CrossRef]
- Qin, S.J. Statistical process monitoring: Basics and beyond. *J. Chemom.* **2003**, *17*, 480–502. [CrossRef]
- Kourti, T. Application of latent variable methods to process control and multivariate statistical process control in industry. *Int. J. Adapt. Control. Signal. Process.* **2005**, *19*, 213–246. [CrossRef]
- Arnau, V.; Mars, S.; Marin, I. Iterative cluster analysis of protein interaction data. *Bioinformatics* **2005**, *21*, 364–378. [CrossRef] [PubMed]
- Wang, X.Z.; Smith, K.; Hyndman, R. Characteristic-based clustering for time series data. *Data Min. Knowl. Discov.* **2006**, *13*, 335–364. [CrossRef]
- Bar-Joseph, Z.; Demaine, E.D.; Gifford, D.K.; Srebro, N.; Hamel, A.M.; Jaakkola, T.S. K-ary clustering with optimal leaf ordering for gene expression data. *Bioinformatics* **2003**, *19*, 1070–1078. [CrossRef]
- Clark, N.C.; Olsvik, O.; Swenson, J.M.; Spiegel, C.A.; Tenover, F.C. Detection of a streptomycin/spectinomycin adenyltransferase gene (*aadA*) in *Enterococcus faecalis*. *Antimicrob. Agents Chemother.* **1999**, *43*, 157–160. [CrossRef]
- Pei, R.T.; Kim, S.C.; Carlson, K.H.; Pruden, A. Effect of River Landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes (ARG). *Water Res.* **2006**, *40*, 2427–2435. [CrossRef]
- Liu, M.M.; Zhang, Y.; Yang, M.; Tian, Z.; Ren, L.R.; Zhang, S.J. Abundance and Distribution of Tetracycline Resistance Genes and Mobile Elements in an Oxytetracycline Production Wastewater Treatment System. *Environ. Sci. Technol.* **2012**, *46*, 7551–7557. [CrossRef] [PubMed]
- Janssen, G.R.; Ward, J.M.; Bibb, M.J. Unusual Transcriptional and Translational Features of the Aminoglycoside Phosphotransferase Gene (*Aph*) from *Streptomyces*-Fradiidae. *Genes Dev.* **1989**, *3*, 415–429. [CrossRef] [PubMed]
- Literak, I.; Dolejska, M.; Radimersky, T.; Klimes, J.; Friedman, M.; Aarestrup, F.M.; Hasman, H.; Cizek, A. Antimicrobial-resistant faecal *Escherichia coli* in wild mammals in central Europe: Multiresistant *Escherichia coli* producing extended-spectrum beta-lactamases in wild boars. *J. Appl. Microbiol.* **2010**, *108*, 1702–1711. [CrossRef] [PubMed]
- Sidhu, M.S.; Sorum, H.; Holck, A. Resistance to quaternary ammonium compounds in food-related bacteria. *Microb. Drug Resist.-Mech. Epidemiol. Dis.* **2002**, *8*, 393–399. [CrossRef] [PubMed]



19. Trade, F.U.S. *Trade in Goods by Country*. Available online: <https://www.census.gov/foreign-trade/balance/index.html> (accessed on 20 November 2019).
20. Antunes, P.; Machado, J.; Sousa, J.C.; Peixe, L. Dissemination of sulfonamide resistance genes (sul1, sul2, and sul3) in Portuguese Salmonella enterica strains and relation with integrons. *Antimicrob. Agents Chemother.* **2005**, *49*, 836–839. [[CrossRef](#)] [[PubMed](#)]
21. Blyth, D.M.; Mende, K.; Maranich, A.M.; Beckius, M.L.; Harnisch, K.A. Antimicrobial resistance acquisition after international travel in U.S. travelers. *Trop. Dis. Travel Med. Vaccines* **2016**, *2*. [[CrossRef](#)] [[PubMed](#)]
22. CDC. What Is Salmonella. 2019. Available online: <https://www.cdc.gov/salmonella/general/index.html> (accessed on 15 August 2019).
23. Murray, C.K.; Blyth, D.M. Acquisition of Multidrug-Resistant Gram-Negative Organisms during Travel. *Mil. Med.* **2017**, *182*, 26–33. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).