Biodiversity Measures to Summarize Antibiotic Resistance

By

© Sovit Raj Chalise

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biology, Faculty of Science

Memorial University of Newfoundland

St. John's, Newfoundland & Labrador

June 2016

Abstract

Studies of antibiotic resistance frequently focus on resistance to specific antibiotics, or classes of antibiotics, without summarizing the overall resistance for all antibiotics. Biodiversity indices, such as the Shannon-Weiner (SW) index and Bray-Curtis distance, are widely used in ecology to measure overall species abundance and variation within and between regions. We explore the use of biodiversity indices to summarize antibiotic resistance. We illustrate this approach by analyzing data on antibiotic resistance in clinical *Salmonella enterica* isolates. To understand changes in resistance patterns within and between the provinces of Canada, *S. enterica* serotypes, and over time, we measure the SW diversity and Bray-Curtis distance and visualize the differences between provinces, serotypes and over time using ordinations. We describe the different types of antibiotic resistance data that are required for measuring alpha and beta diversity and explain the interpretation of biodiversity measures when applied to antibiotic resistance data.

Acknowledgements

I would like to sincerely thank my supervisor Dr. Amy Hurford for this opportunity, and her continuous support and patient guidance throughout the entire project. I really appreciate your valuable time and hard work in making my thesis as clear and concise as possible. I would also like to thank my committee members Dr. Shawn Leroux and Dr. Kapil Tahlan for their insights, guidance and feedback on my research and review of this thesis. I would also like to thank my lab members, Matt Rittenhouse, Fabio Frazao, and Abdou Fofana for their assistance and advice throughout my entire project. At last, I would like to thank my parents, Akendra Raj Chalise and Sabitri Chalise, along with all my friends and family for their continuous support in most stressful times. I will always remember the help, support and guidance I received throughout my entire project without which I could not have accomplished this. Thank you so much.

Table of Content

List of Figures

APPENDIX

List of Tables

APPENDIX

Chapter 1

Introduction

Antibiotics have substantially reduced the duration and number of deaths due to bacterial infections (Davies et al. 2010). However, the recent increase in antibiotic resistance, in particular, resistance to multiple antibiotics, threatens the effective treatment of bacterial infections (Yeh et al., 2009, Davies et al. 2010, MacLean et al. 2010, Magiorakos et al. 2011, Campbell et al. 2014, WHO, 2015). Some bacteria, called "superbugs" are now resistant to almost all the antibiotic treatments available and many *Mycobacterium tuberculosis* and *Staphylococcus aureus* infections are resistant to several different antibiotics. Globally, more than 700,000 people die each year as a result of antibiotic resistance (O'Neill, 2016) and a recent study estimates 300 million premature deaths and economic losses up to $100 trillion dollars due to antibiotic resistance by the year 2050 (Phimister et al., 2015).

Biodiversity refers to the great variety of life forms present on earth. Measures to estimate biodiversity have been widely used in ecology to summarize species composition within and between communities. Alpha biodiversity measures species composition within a community by jointly considering richness (the number of different species present) and evenness (equality in relative species abundance; Magurran, 1988). Beta biodiversity

measures the pairwise difference in the relative abundance of species between communities and is used to compare similarities in the composition of communities. Recently these biodiversity metrics have seen a limited application to populations of pathogens with antimicrobial resistance (Pakyz et al. 2008, de la Pedrosa et al. 2009, Blaak et al. 2015, Zhang et al. 2014, Sigala and Unc 2013, Abay et al. 2014).

In this thesis, I explore the application of biodiversity measures to antibiotic resistance data to measure antibiotic resistance in a way that summarizes the level of resistance across many different antibiotics. This method is advantageous for two main reasons.

Firstly, in clinical settings, treatment can occur without knowing antibiotic sensitivities, for example, infections such as sepsis, meningitis, pneumonia, pyelonephritis, gastroenteritis, osteomyelitis, and cellulitis in children are diagnosed and treated without any information on antibiotic sensitivities (Bruel et al. 2007). In these instances, a summary measure of antibiotic resistance would provide information on how likely uninformed antibiotic choices are to succeed. I use biodiversity measures to quantify the uncertainty regarding which antibiotics an isolate is resistant to and to identify bacterial populations that are more or less similar in terms of resistance.

Secondly, a bacterial pathogen can be resistant to none, multiple or all antibiotics available, which makes measuring the overall level of antibiotic resistance at the isolate or population level difficult since many antibiotics are involved. If at the population level the prevalence of antibiotic resistance increases for some antibiotics but decreases for others, there are no existing methods that measure whether *overall* antibiotic resistance is increasing or decreasing. Our method has novel public health implications as it provides

summary information regarding antibiotic resistance across a wide number of antibiotics within a population and also compares similarities and differences in antibiotic resistance between different populations.

This summary information regarding overall antibiotic resistance within a region is useful for public health personnel to prioritize regions that would benefit from susceptibility testing. If the overall antibiotic resistance for a region is high, then it is unlikely that the same antibiotic choices will be effective for different patients in that region in the absence of susceptibility tests. When interventions such as antibiotic stewardship are implemented, measuring the overall antibiotic resistance before and after the intervention is an approach to quantifying the efficacy of the intervention. Additionally, a summary measure is useful to determine if antibiotic resistance is increasing over time, so as to determine if an intervention is needed or should be expected in the future. The summary measure provides information on how the level of antibiotic resistance for one hospital or region compares to others, which is helpful in determining which are in need of increased antibiotic stewardship or susceptibility testing.

In Chapter 2, I provide an overview of biodiversity measures and their application to ecology and microbial studies. In Chapter 3, I propose methods to apply biodiversity theory to antibiotic resistance data and illustrate these methods using clinical data from human *Salmonella enterica* isolates. The choice to analyze *S. enterica* data was made because the CIPARS *Salmonella* data set is one of the few that contains complete information on antibiotic resistance at the individual isolate level. The data set also consists of a relatively large number of isolates. Additionally, the epidemiology of *S. enterica* is similar enough to

other pathogens of potential interest that the analysis of the *S. enterica* data will be informative for the analysis of these other pathogens (see Section 3.2.2, Chapter 3 for further discussion on the choice of *Salmonella* dataset).

References

Abay, S., Kayman, T., Otlu, B., Hizlisoy, H., Aydin, F. & Ertas, N. (2014). Genetic diversity and antibiotic resistance profiles of Campylobacter jejuni isolates from poultry and humans in Turkey. International Journal of Food Microbiology, 178(16), 29-38

Blaak, H., Van Hoek, A., Hamidjaja, R., Van Der Plaats, R., Kerkhof-de Heer, L., De Roda Husman, A., & Schets, F. (2015). Distribution, Numbers, and Diversity of ESBL-Producing E. coli in the Poultry Farm Environment. *PloS One, 10*(8), E0135402.

Campbell, E. M., & Chao, L. (2014). A Population Model Evaluating the Consequences of the Evolution of Double-Resistance and Tradeoffs on the Benefits of Two-Drug Antibiotic Treatments. *PLoS ONE*, *9*(1), e86971. doi:101371/journal.pone.0086971

Davies, J., & Davies, D. (2010). Origins and Evolution of Antibiotic Resistance. *Microbiology and Molecular Biology Reviews*, *74*(3), 417–433. doi:10.1128/MMBR.00016-10

De la Pedrosa, E., Baquero, F., Loza, E., Nadal-Serrano, J., Fenoll, A., Del Campo, R., & Cantón, R. (2009). High clonal diversity in erythromycin-resistant Streptococcus pneumoniae invasive isolates in Madrid, Spain (2000–07). *Journal of Antimicrobial Chemotherapy, 64*(6), 1165-1169.

Holm, A. M., Cordoba, G. R., Siersma, V., Bjerrum, L., Sørensen, T., & Jessen, L. (2015). Point of care susceptibility testing in primary care - Does it lead to a more appropriate prescription of antibiotics in patients with uncomplicated urinary tract infections? Protocol for a randomized controlled trial. *BMC Family Practice, 16*(1), BMC Family Practice, 21 August 2015, Vol.16(1).

MacLean, R. C., Hall, A. R., Perron, G. G., & Buckling, A. (2010). The population genetics of antibiotic resistance: integrating molecular mechanisms and treatment contexts. *Nature Reviews Genetics*, *11*(6), 405–414. doi:10.1038/nrg2778

Magiorakos, A. P., Srinivasan, A., Carey, R.B., Carmeli, Y., Falagas, M.E., Giske, C.G., et al. (2012). Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: An international expert proposal for interim standard definitions for acquired resistance. *Clinical Microbiology and Infection, 18*(3), 268-281.

Magurran, A. E. (1988). Ecological diversity and its measurement. Princeton, N.J: Princeton University Press.

O'Neil, J. (2016). Tackling Drug-Resistant Infections Globally: Final Report and

Recommendations. The Review of Antimicrobial Resistance. Web. 2016. Retrieved from http://amr review.org/sites/default/files/160525_Final%20paper_with %20cover.pdf

Pakyz, A., Powell, J., Harpe, S., Johnson, C., Edmond, M., & Polk, R. (2008). Diversity of antimicrobial use and resistance in 42 hospitals in the United States. *Pharmacotherapy, 28*(7), 906-12.

Phimister, Arias, & Murray. (2015). A New Antibiotic and the Evolution of Resistance. *The New England Journal of Medicine, 372*(12), 1168-1170.

Sigala, J., & Unc, A. (2013). Pyrosequencing estimates of the diversity of antibiotic resistant bacteria in a wastewater system. *Water Science and Technology: A Journal of the International Association on Water Pollution Research, 67*(7), 1534-43.

Yeh, P. J., Hegreness, M. J., Aiden, A. P., & Kishony, R. (2009). Drug interactions and the evolution of antibiotic resistance. *Nature Reviews Microbiology*, *7*(6), 460–466.

Van Den Bruel, A., Aertgeerts, B., Bruyninckx, R., Aerts, M., & Buntinx, F. (2007). Signs and symptoms for diagnosis of serious infections in children: A prospective study in primary care. *The British Journal of General Practice : The Journal of the Royal College of General Practitioners, 57*(540), 538-46.

WHO. (2015). Antimicrobial resistance. World Health Organization Media Centre. Fact Sheet. Web. 2015. Retrieved from http://www.who.int/mediacentre/factsheets/fs194/en/

Zhang, Bo, Zhang, Zi-wei, Meng, Fan-Yu, Wu, Qiong, Xu, Shi-Wen, & Wang, Xiao-Long. (2014). Analysis of the structure of bacteria communities and detection of resistance genes of quinolones from

Chapter 2

Overview of Ecological Diversity Indices

2.1 Introduction

Diversity indices are widely used to summarize the species composition of communities with over 5,900 studies in ecology having calculated some measure of biological diversity (Web of Science, 2016). Ecologists make a distinction between diversity measured at the regional and landscape level, which are termed alpha and gamma diversity respectively. Under this distinction, the regions are subunits within the landscape and landscape level (gamma) diversity is calculated by pooling species abundances across the regions. Comparison of the alpha and gamma diversity values provides information on how diversity changes *between* regions. This is termed beta diversity and for many studies, beta diversity may be the quantity of most interest.

This chapter presents an overview of some of the widely used alpha and beta diversity measures in ecology, which may be useful in the study of antibiotic resistance as discussed in the next chapter (Chapter 3).

2.2 Measuring regional (alpha) diversity

Species richness is defined as the number of species present in an area. Species evenness measures the similarity in the proportional abundance of these species. Various different alpha diversity indices differ in terms of the relative contribution that species richness versus species evenness makes to the overall index value (Magurran, 1988). The necessary properties of an alpha diversity index are: (i) incorporates both species richness and species evenness; (ii) attains a maximum value when all species are present in equal proportions; and (iii) for two communities with equal evenness, should attain a higher value for the community with more species (Desrochers and Anand, 2004).

In 1948, Claude Shannon defined the concept of entropy in information theory as the amount of uncertainty associated with the identity of a character in a message (Shannon and Weaver, 1948). When applied to ecology, Shannon entropy quantifies the uncertainty of the species identity of an individual in the population: for a highly diverse population the species identity is highly uncertain because there are a large number of equally common species, whereas a population has low diversity if dominated by one or a few common species (Pielou, 1975). This idea forms the basis of the Shannon-Wiener (SW) diversity index.

The SW diversity index is calculated as,

$$H' = -\sum_{i=1}^{s} p_i \ln(p_i), \quad (1)$$

where p_i is the proportion of individuals in the ith species, $ln(p_i)$ is the natural logarithm of p_i, and s is the total number of species (i.e., the species richness, Magurran, 1988, Anderson et al. 2011). A simple interpretation of the SW diversity index value arises from exponentiating the SW diversity value ($e^{H'}$). This calculation gives the '*effective diversity*' (Anderson et al. 2011): the number of species, which if equally abundant, would have the same SW diversity value as the community of interest. Table 2.1 provides an example of calculating alpha diversity using the SW index. C1 and C2 have the greatest alpha diversity because these communities have the greatest number of species present and a more even frequency distribution.

Another widely used diversity index is the Simpson's index (Magurran, 2004),

$$H' = 1 - \sum_{i=1}^{s} p_i^2 \, . \quad (2)$$

The Simpson's index can be interpreted as the expected probability that two randomly selected individuals sampled from the community without replacement are of different species, i.e. $\sum_{i=1}^{s} p_i(1 - p_i)$ where equation (2) is recovered by noting that every individual can only belong to one species, $\sum_{i=1}^{s} p_i = 1$. Different alpha diversity measures are unified by the concept of Hill numbers (Hill, 1973, Jost 2006, Jost 2007, Chao et al. 2014). Here, different values of a parameter in a common formula can be used to recover SW, Simpson's and other diversity measures.

In addition to the SW and Simpson's index, other alpha diversity indices may be appropriate for particular studies (Table 2.2, Table 2 in Hill et al. 2003, Table 1 in Jost

et al. 2006, Box 5.1 in Magurran and McGill, 2011, pp. 56). For any study, the best choice of alpha diversity index depends on whether it is more appropriate to emphasize rare or common species. However, frequently this criterion does not help select an index and the SW or Simpson's are used simply because they are easy to interpret and frequently used in other studies (Pakyz et al. 2008, de la Pedrosa et al. 2009, Sigala and Unc, 2013, Morris et al. 2014). Alpha diversity indices that more heavily emphasize species richness are sensitive to the detection of rare species and alpha diversity indices, in general, are prone to undersampling bias as they depend on species richness, which is sensitive to sample size (Soetaert and Heip, 1990, Magurran, 2004, Chao et al. 2014). Several studies (Walker et al. 2008, Ricotta et al. 2012, Gotelli and Ellison, 2012, Chao and Jost, 2012, Colwell et al. 2012) have addressed the problem of undersampling bias for diversity indices. Recently, Chao et al. (2014) explored the use of interpolation and extrapolation to reduce the effects of undersampling bias on alpha diversity estimates; interpolation or rarefaction involves standardizing all samples to smaller equal sample sizes, and extrapolation involves standardizing to larger equal sample sizes.

Some studies (Stirling and Wilsey, 2001, Heino et al. 2008) recommend using several different alpha diversity measures to dilute the effect of biases that may arise from any one index. Leinster et al. (2012) propose an approach that replaces a flood of diversity indices with a single formula that measures diversity in terms of effective numbers. This approach of Leinster et al. (2012) enables comparisons of communities based on 'diversity profiles' and takes into account the similarities between species, not just their relative abundance.

Several previous studies have used ecological diversity indices to quantify the composition of a population of antibiotic-resistant isolates. Pakyz et al. (2008) compared the alpha diversity of prescribed antibiotics and antibiotic resistance data for 17 hospitals. To calculate the SW and Simpson's diversity in the antibiotics prescribed to patients, Pakyz et al. (2008) defined 14 different classes of antibiotics and p_i (as it appears in equations (1) and (2)) was calculated as the proportion of defined daily doses (DDD) due to each of the 14 classes. Using this approach, Pakyz et al. (2008) were able to summarize the level of uncertainty in which antibiotics were prescribed to patients across antibiotic classes. In the same study, Pakyz et al. (2008) calculated the diversity of antibiotic resistance for specific pathogen-antibiotic combinations (see Table 1 in Pakyz et al. 2008). For example, one of the pathogen-antibiotic combinations was to test *Staphylococcus* isolates for resistance to methicillin and here p_i is the proportion of Staphylococcus isolates resistant to methicillin. Using this approach diversity values were calculated for specific pathogen-antibiotic combinations (i.e. *Staphylococcus* isolates resistant to methicillin), but diversity summarized for all different types of resistance that can occur in *Staphylococcus* was not.

Other studies have used molecular approaches to distinguish between different types of resistant bacteria (de la Pedrosa et al. 2009, Sigala and Unc, 2013, Abay et al. 2014, Blaak et al. 2015). For example, de la Pedrosa et al. (2009) calculated the diversity of erythromycin-resistant *Streptococcus* by calculating the proportion of isolates belonging to different 'sequence types'. Each sequence type groups together ancestors of the same founding genotype and so groupings are based on shared ancestry, rather

than shared phenotypes, such as being resistant to a particular antibiotic. All of the *Streptococcus* isolates were resistant to erythromycin and so sequence types are a finer partitioning of isolates within a phenotypic class. Another approach is to calculate the species diversity of a population of antibiotic-resistant isolates where p_i is the fraction of antibiotic-resistant isolates belonging to a particular species (Zhang et al. 2014). As such, all of the past studies applying biodiversity measures to antibiotic resistance data focus on quantifying species or molecular diversity and not the antibiotic resistance diversity.

2.3 Comparing regional diversities (Beta diversity)

While alpha diversity indices are useful to quantify the biodiversity within a region, beta diversity indices focus on making comparisons of relative species abundance between regions. Robert H. Whittaker (1960, 1972) first defined the concept of beta diversity as the variation in community composition across regions where regions are subunits within a landscape. The classical approach defines beta diversity based on alpha and gamma diversity values,

$$\beta_{Shannon} = H_\gamma / H_\alpha. \quad (3)$$

(Jost 2007, Anderson et al. 2011) where, H_γ = exp(H'_{pooled}) is the exponential of SW index obtained by pooling abundances for each species across all alpha level units and H_α = exp(H') is the exponential of SW index for a particular alpha level sampling unit.

Approaches to measuring beta diversity that uses equation (3) or a related equation are referred to as the classical approach, however, the second view of beta diversity is to measure similarity, dissimilarity or distances between regions (referred to as the multivariate approach). Tuomisto (2010a), Anderson et al. (2011) and Bennett and Gilbert (2015) have reviewed some important differences between the classical and multivariate approaches in measuring beta diversity and several authors (Legendre and Legendre, 1998, Koleff et al. 2003, Tuomisto, 2010 a, b, Anderson et al. 2011, Barwell et al. 2015) have reviewed various important properties relating to most commonly used beta diversity measures in community data analysis.

The multivariate approach calculates a 'dissimilarity matrix', D, containing all pairwise dissimilarities in species abundance between regions. This dissimilarity matrix is calculated from a 'community matrix', C, where each row in the matrix represents each of n species within a given region and each column represents a different region (Anderson et al. 2011; see Table 2.1). Under the multivariate approach, two regions have low beta diversity when they have an identical composition and equal relative abundances for all species. If the two regions do not share any species, then they are compositionally very different and have high beta diversity. The measure of relative compositional differentiation ranges between 0 and

1, where 0 is assigned to identical regions and 1 means no species are shared between regions (Jost et al. 2011).

Jost et al. (2011) argue that at least three properties should be present in any ecologically useful measure of compositional similarity: i) the measure is a monotonically increasing function of similarity, ii) the measure is based on the relative abundance of species, not the raw abundances, and that iii) the measure is replication-invariant. Replication invariance means, for example, that if the landscape is comprised of k regions each with identical species composition and species abundance, then the similarity between each pair of regions is the same irrespective of how many regions there are (see Magurran and McGill, 2011, pp. 67-68 for a detailed discussion).

Multivariate measures of beta diversity indices are classified into two broad classes of measure (binary or quantitative, Table 2.3) based on the data type (Krebs, 1999). Binary similarity coefficients are used for presence/absence data, considering only whether a species is present or not and ignoring relative abundances (Krebs, 1999, Chao et al. 2006, Anderson et al. 2011, Barwell et al. 2015). When relative abundance data is available, quantitative measures of beta diversity are generally preferred (Barwell et al. 2015) and are less affected by incomplete sampling (Beck et al. 2013, Barwell et al. 2015, Bennett and Gilbert, 2015). There are several abundance-based similarity measures and different measures emphasize different aspects of community data. For example, the Bray-Curtis, Morisita-Horn index, modified Gower, Euclidean on proportions, chi-squared, and Hellinger indices exclude instances of

joint absences and as such, under these metrics, two regions are not considered more similar if they both lack certain species. Like alpha diversity indices, different dissimilarity measures emphasize rare or common species and various data transformations (like proportions, square roots, fourth roots, logarithm) can be applied to raw data before measuring dissimilarity, resulting in different emphasis being place on rare/common species and exclusion/inclusion of joint absence information (Table 2.3, Clarke, 1999, Legendre et al. 2005, Kindt and Coe, 2005, pp130, Table 1 and Figure 5 in Anderson et al. 2011 list several dissimilarity metrics and the circumstances where each is most appropriate). Often several dissimilarity measures are viewed as equally appropriate and Bray-Curtis is selected simply because it has been frequently applied in other studies (Clarke et al. 2006, Jost et al. 2011, Anderson et al. 2006, 2011) thus facilitating comparisons.

The Bray-Curtis distance, developed by J. Roger Bray and John T. Curtis during their research work on plant community ordination, is the most popular and frequently used abundance based dissimilarity measure (Bray and Curtis, 1957, Clarke et al. 2006, Jost et al. 2011, Anderson et al. 2006, 2011). It is widely used in ecology to create distance matrices for vegetation ordination analysis (Gotelli and Ellison, 2004, Chao et al. 2006). The desirable properties of the Bray-Curtis dissimilarity measure include excluding joint absences, insensitivity to the choice of units, robustness to the addition of a new region, and when relative abundance of species are same between regions it recognizes differences in total abundance. Bray-Curtis distance is calculated as,

$$B = \frac{\sum_{k=1}^{N} |X_{ak} - X_{bk}|}{\sum_{k=1}^{N} (X_{ak} + X_{bk})}, \quad (4)$$

where, X_{ak} is the abundance of species k in the ath region, X_{bk} the abundance of species k in the bth region, and N is the total number of species recorded across both units. Bray-Curtis is a measure of relative compositional differentiation and as such, the index gives a value of 0 if the communities are identical and 1 if the communities are completely different (Table 2.3; see Magurran and McGill, 2011, chapter 6 for a detailed review of Bray-Curtis distance along with other various similarity and dissimilarity measures).

Table 2.1 shows a sample calculation of Bray-Curtis dissimilarity. The dissimilarity matrix reveals that C1 is dissimilar to both C2 (B=0.64) and C3 (B=0.67) and that C2 is relatively similar to C3 (B=0.22). Referring to the community matrix (Table 2.1), we can see that the only difference between C2 and C3 is that C2 has 4 individuals of species D.

2.3.1 Ordination Analysis

Dissimilarity matrices are often difficult to interpret because they involve many pairwise comparisons. For example, a landscape consisting of 10 different regions would give rise to a dissimilarity matrix that involves 45 pairwise comparisons. For such a landscape, how can we simply represent the dissimilarity between the different regions? Ordination techniques show the relationship between communities and species in a low-dimensional space as accurately as possible (Gauch, 1982, Palmer, 2006), where frequently this low-dimensional space is a two-dimensional

graph with a horizontal and vertical axis (i.e., Figure 2.1). Regions that appear closer in ordination are interpreted as being similar in species composition and the ones that appear apart as containing different species.

Principal Component Analysis (PCA) is one of the oldest (Pearson, 1901) and most widely used unconstrained ordination methods (Kindt and Coe, 2005, pp 154, Ramette, 2007, Paliy and Shankar, 2016). The Principle Components (PCs) are a linear combination of the species abundance that account for as much variance in species abundance as possible (Ramette, 2007, Paliy and Shankar, 2016). The first principal component (PC1) is chosen to explain as much variance as possible and successive principal components are orthogonal to PC1 and explain progressively less of the remaining variance (Kindt and Coe, 2005).

PCA can be performed only to represent the Euclidean distances (see Table 2.1) between regions. Generally, the ability to represent only one dissimilarity measure limits the applicability of PCA for ecological studies. As a dissimilarity measure, some of the limitations of Euclidean distance are the inclusion of joint absences and sensitivity to the abundance of each species rather than shared species. To overcome these limitations, one option is to transform the community matrix prior to performing a PCA (Kindt and Coe, 2005, Legendre and Gallagher 2001). For example, Euclidean distance performed on proportions ensures that results are not influenced by differences in total abundance between sites (Kindt and Coe, 2005) and excludes joint absence information (Anderson et al. 2011).

Figure 2.1A shows a PCA ordination corresponding to the community data represented in Table 2.1. The data were transformed into proportional abundances and the PCA was performed on the variance-covariance matrix (Table 2.4; Scaling 1) formed by measuring the covariance in species abundance for the community matrix (Table 2.1). In Figure 2.1A, the distance between the communities C1, C2, and C3 is the Euclidean distance. Euclidean distance would not be represented accurately if the PCA was performed on a matrix that describes the correlation between species abundances (Scaling 2; Kindt and Coe, 2005, Ramette, 2007). Since they appear closer in the two-dimensional PCA, Figure 2.1A shows that communities C2 and C3 are most similar, which can be confirmed by referring to Table 2.1 (Dissimilarity matrix - Euclidean distance). The first principal component (PC1) explains 75.41% and the second principal component (PC2) explains 24.58% of the variation in species abundance. Arrows are drawn for species S2 and S4, which shows the direction from the origin for which communities have higher than average relative abundance for these species. For example, community C2 appears in the direction of S4 because C2 has higher than average relative abundance for S4 (Figure 2.1A). The length of each arrow projected onto the PC1 or PC2 axis quantifies the contribution of that species to the respective principal component (Green, 1990). For example, the abundance of species S2 makes the biggest contribution to PC1 with species S1, S3 and S4 making smaller contributions, while species S4 makes the biggest contribution to PC2 (Figure 2.1A). The perpendicular lines (dashed) from each community onto the species S2 arrow, as shown in Figure 2.1A, indicate the relative ranking of the abundance of S2 for each community. Since C3 is projected farthest from the origin in the direction of

S2, it has the highest relative abundance of S2. Community C1 is projected onto the reflection of the S2 vector, which means it has lower than an average relative abundance of species S2. Referring to Table 2.1 (Community matrix), C3 has a higher relative abundance of species S2 than C2 and examining the perpendicular lines (Figure 2.1A), C3 shows a higher relative abundance of species S2 than C2. However, the positions in PCA only approximate the exact relationships in the data since only a percentage of the total variation is explained by the graph (see Legendre and Legendre, 1998, chapter 9 and Ramette, 2007 for a detailed discussion on PCA method).

Another technique for unconstrained ordination is Principal coordinates Analysis (PCoA). PCoA is similar to PCA but can be performed using any distance/dissimilarity measure. Non-metric Multidimensional Scaling (NMDS) is an unconstrained ordination technique, related to PCoA, where NMDS ordination ranks the dissimilarity between sites of a community matrix and uses those ranks to map the sites onto the chosen N-dimensional ordination space (usually a two-dimensional graph). As NMDS preserves only the ranked dissimilarities, we cannot conclude that a pair of regions appearing twice as far apart as another pair of regions are twice as dissimilar, rather we can only conclude that the regions appearing more distant are more dissimilar. Because high dimensional data (i.e., many different pairwise dissimilarities) are represented in a lower dimensional space it may not be possible for the NMDS ordination to preserve the complete accuracy of the data. The accuracy of the NMDS in representing the relationships present in the dissimilarity matrix can also be visualized using a Shepherd diagram (Shepard, 1962), which plots the distances

between the points in the ordination space versus the values of the dissimilarity matrix (Legendre and Legendre, 1998). The mismatch between the rank order of the pairwise dissimilarities in the dissimilarity matrix and in the NMDS ordination is measured as 'stress' (Palmer, 2006) and generally, a stress value of ≤ 0.15 is considered acceptable (Clarke, 1993, Paliy and Shankar, 2016).

Figure 2.1B shows an NMDS ordination corresponding to the community data represented in Table 2.1. Figure 2.1B shows that communities C2 and C3 are most similar to each other and that both of these communities are relatively dissimilar to C1. Referring to Table 2.1 (Dissimilarity matrix – Bray-Curtis) we can verify that this figure is a reasonable representation (i.e., from most similar to least similar: C2 and C3 (0.22), C1 and C2 (0.64), and C1 and C3 (0.67)).

Several previous studies have used beta diversity measures and ordination techniques to compare species diversity and compositions of bacterial communities. Price et al. (2009) used molecular techniques, Bray-Curtis distance, and NMDS ordination methods to compare the bacterial community in the wounds of antibiotic-treated patients to that of untreated patients. Price et al. (2009) found that antibiotic treatments affect bacterial community structure by reducing the number of some bacteria while having no effect on others. Hong et al. (2010) compared the composition of microbiota in the feces of infants with and without eczema (a chronic disorder in children). The presence of different bacterial phyla was detected using molecular techniques and then PCA was performed on the abundance of individual taxonomic units. Hong et al. (2010) found a strong correlation between the

abundances of particular bacterial genera and the health status of infants during the early stage of infancy. A study by Caucci et al. (2016) used beta diversity measures and ordinations, finding that water treatment influenced the abundance and composition of the microbial community, but had no significant effect in the relative abundance of the antibiotic resistance genes. As for the application of alpha diversity measures to antibiotic resistance data, all the previous studies of beta diversity and ordinations focus on comparing species or molecular composition between microbial communities and not the composition at the phenotypic level, i.e. composition of antibiotic resistance.

2.4 Conclusions

Ecological diversity indices have been in use for several decades (Whittaker 1972, Magurran, 1988). Alpha diversity indices are useful for quantifying the uncertainty regarding the species present in a community and account for both species richness and species evenness. There are various alpha diversity measures that emphasize the different aspect of species abundance (i.e., rare versus common). Beta diversity measures are useful in comparing two or more communities based on the species composition as they quantify the similarity in the relative abundance of species between communities. Like alpha diversity, there are numerous beta diversity measures that emphasize different aspects of species composition, such as inclusion versus exclusion of joint absence information or emphasis on common versus rare

species. Additionally, ordination techniques can be used to effectively visualize the differences in community composition measured using multidimensional beta diversity indices when there are more than two communities. Past studies demonstrate the application of ecological diversity indices to measuring the species or molecular diversity of microbial communities but not the phenotypic diversity (i.e. the diversity of antibiotic resistance). The approaches explained in this chapter can also be applied to antibiotic resistance data to quantify the uncertainty surrounding which antibiotics a bacterial isolate is resistant to and to compare antibiotic resistance between regions as explained in Chapter 3.

2.5 Tables and Figures

Table 2.1: Example of Community and Dissimilarity Matrices. The community matrix describes the abundance and relative abundance (parentheses) of species in each of three regions (C1, C2, C3). The Shannon-Weiner (SW) index (H'; equation 1) and the effective diversity ($e^{H'}$) is calculated for the community matrix. The dissimilarity matrix quantifies the pairwise dissimilarity between the regions as measured by Euclidean and Bray-Curtis distance. This example is based on Tuomisto et al. (2006).

Community matrix			
	C1	**C2**	**C3**
Species S1	5 (0.45)	1 (0.09)	1 (0.14)
Species S2	1 (0.09)	5 (0.45)	5 (0.71)
Species S3	4 (0.36)	1 (0.09)	1 (0.14)
Species S4	1 (0.09)	4 (0.36)	0 (0.00)
SW index (H')	1.16	1.16	0.8
Effective diversity ($e^{H'}$)	3.19	3.19	2.23

Dissimilarity matrix – Euclidean		
	C1	**C2**
C2	7.07	
C3	6.48	4.00

Dissimilarity matrix – Bray-Curtis		
	C1	**C2**
C2	0.64	
C3	0.67	0.22

Table 2.2: Comparison of some common Alpha Diversity Indices

Indices	***What they do?***	***Formulae***
1. Shannon-Weiner index	➢ Measures the uncertainty of species identity in a community ➢ Combines species richness and evenness ➢ More moderate and broad weighting to rare and intermediate species	$H' = -\sum_{i=1}^{s} p_i \ln(p_i)$ p_i is the proportion of individuals belonging to the i^{th} species; s is the total number of species
2. Simpson's index	➢ Gives probability of two species drawn at random belonging to the same community ➢ Heavily weighs common species	$H' = 1 - \sum_{i=1}^{s} p_i^2$ p_i is the proportion of individuals belonging to the i^{th} species; s is the total number of species
3. Berger-Parker index	➢ Relative abundance of the most abundant species ➢ Rare species are completely ignored	$H' = \max(p_i)$ p_i is the proportion of individuals belonging to the i^{th} species
4. Renyi entropy	➢ Generalization of SW diversity index. ➢ Gives different diversity indices for values of q other than 1	$H' = \frac{1}{q-1}\left(-\ln \sum_{i=1}^{s} p_i^q\right)$ p_i is the proportion of individuals belonging to the i^{th} species; s is the total number of species; q represents the order of diversity other than 1

Table 2.3: Comparison of some common Beta Diversity Indices

Indices	*What they do?*	*Formulae*
A) Binary Similarity Coefficients		
1. Jaccard index	➢ Compares number of shared species to total number of species in combined communities ➢ Global view of communities ➢ Doesn't consider commonness and scarcity of species ➢ Weighs rare and common species equally	$\frac{S_{12}}{S_1 + S_2 - S_{12}}$ S_{12} – number of shared species in both communities S_1 – number of species in community 1 S_2 – number of species in community 2
2. Sorenson index	➢ Compares number of shared species to the mean number of species in a single community ➢ Local view of the community ➢ Doesn't consider commonness and scarcity of species ➢ Weighs rare and common species equally	$\frac{2 * S_{12}}{S_1 + S_2}$ S_{12} – number of shared species in both communities S_1 – number of species in community 1 S_2 – number of species in community 2
B) Quantitative Similarity Coefficients		
1. Morisita-Horn Index	• Highly sensitive to most abundant species • Rare species have little effect • Excludes joint absences	$S_{MH} = 1 - \frac{\sum_{k=1}^{N}(X_{ak} - X_{bk})^2}{\sum_{k=1}^{N} X_{\square k}^2 + \sum_{k=1}^{N} X_{bk}^2}$ S_{MH} = index of similarity N – total number of species in combined communities X_{ak} - number of individuals of species k in community a X_{bk}- number of individuals of species k in community b
2. Euclidean Distance	• Strongly depends on abundance of each species • Abundant species have more effect • Includes joint absences	$E_{ab} = \sqrt{\sum_{k=1}^{N}(X_{ak} - X_{bk})^2}$ N – total number of species X_{ak} - number of individuals of

		species k in community a X_{bk} – number of individuals of species k in community b
3. Manhattan Distance ▪ *The following two measures (i and ii) are based on Manhattan Distance*	• Standardized over individuals in a community • Rare and abundant species make equal contribution • Includes joint absences	$d_M(a,b) = \sum_{k=1}^{N} \lvert X_{ak} - X_{bk} \rvert$ N – total number of species X_{ak} - number of individuals of species k in community a X_{bk} – number of individuals of species k in community b
i) Bray-Curtis Distance	○ Recognizes differences in total abundance when relative abundance is same ○ Dominated by abundant species ○ Rare species have very little effect ○ Excludes joint absences	$B = \frac{\sum_{k=1}^{N} \lvert X_{ak} - X_{bk} \rvert}{\sum_{k=1}^{N} (X_{ak} + X_{bk})}$ B - measure of dissimilarity N - number of species in communities X_{ak} – number of individuals of species k in community a X_{bk} – number of individuals of species k in community b
ii) Canberra metric	○ Standardized over species instead of individuals ○ Rare species have more effect ○ Abundant species have relatively less effect ○ Includes joint absences	$C = \frac{1}{N} \sum_{k=1}^{N} \frac{\lvert X_{ak} - X_{bk} \rvert}{\lvert X_{ak} + X_{bk} \rvert}$ C - coefficients of dissimilarity N - number of species in communities X_{ak} – number of individuals of species k in community a X_{bk} – number of individuals of species k in community b
4. Gower's Distance	• Handles situations when missing values are present • Individual species can be weighted differently • Includes joint absences	$s_{ab} = \frac{\sum_k W_{abk} S_{abk}}{\sum_k W_{abk}}$ s_{ab} – Similarity coefficient comparing two communities a and b S_{abk} – contribution provided by k^{th} variable W_{abk} – usually 1 or 0, depends on whether comparison is valid for the k^{th} variable

Table 2.4: The covariance matrix corresponding the community matrix in Table 2.1. The covariance matrix describes the covariance between species (S1, S2, S3, S4) in each of three regions (C1, C2, C3) for the community matrix in Table 2.1.

Covariance Matrix				
Species	**S1**	**S2**	**S3**	**S4**
S1	0.04	-0.05	0.03	-0.01
S2	-0.05	0.10	-0.04	-0.01
S3	0.03	-0.04	0.02	-0.01
S4	-0.01	-0.01	-0.01	0.04

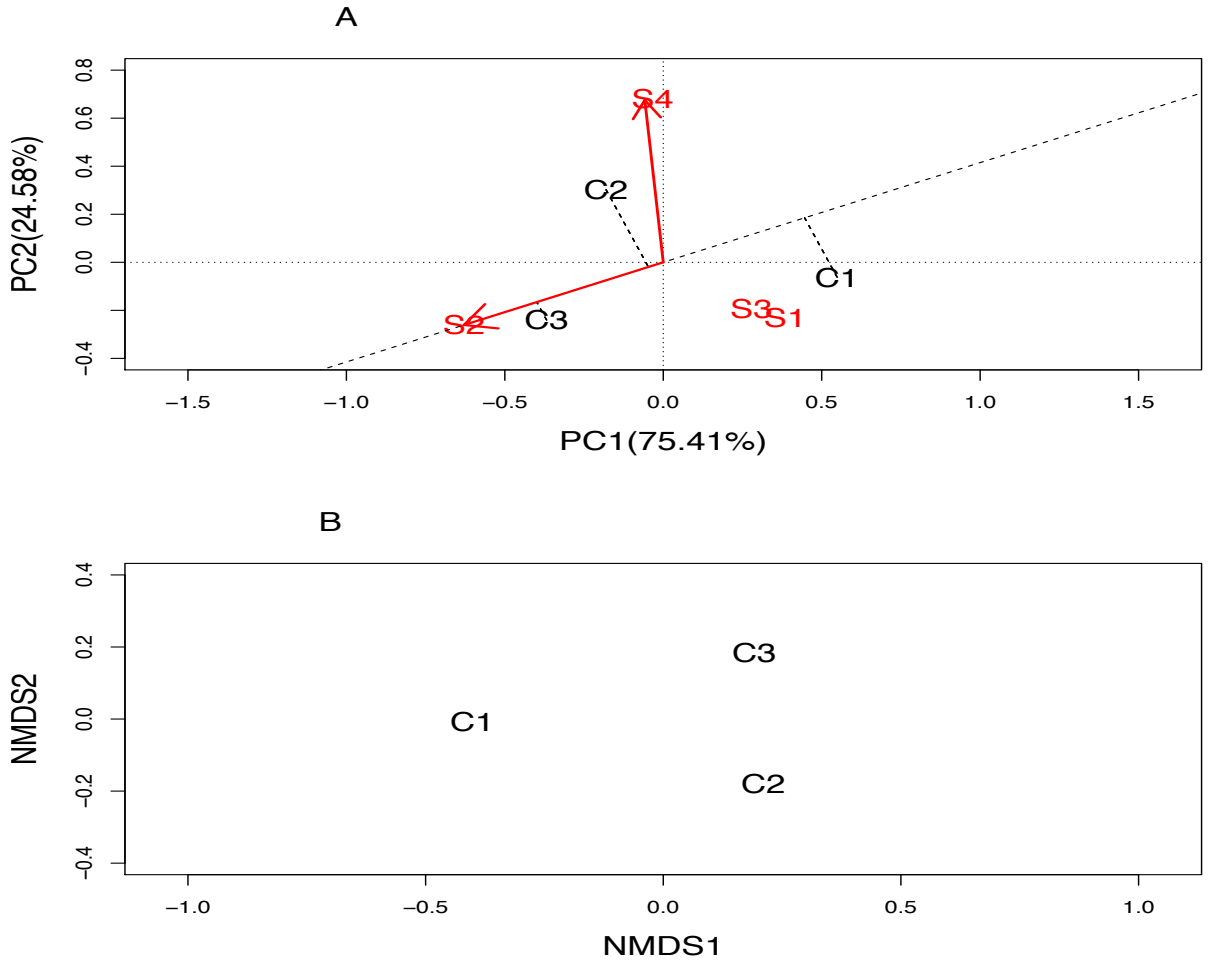


Figure 2.1: Ordinations for the Community Matrix shown in Table 2.1. A. PCA for the Community Matrix in Table 2.1, using the covariance matrix (Table 2.4), where PC1 explains 75.41% and PC2 explains 24.58% of the variance in the composition of species S1, S2, S3, and S4 with respect to communities C1, C2, and C3. For visual clarity, vector arrows are only drawn for species S2 and S4, but can be drawn for species S3 and S1 in the same way. The point of intersection of the perpendicular lines (dashed) from each community onto the species S2 arrow indicate ranking of communities from low (C1 lowest) to high (C3 highest) in relative abundance for species S2. Similarly, perpendicular lines can be drawn onto any species arrow for similar interpretation. B. NMDS for the Community Matrix in Table 2.1, using Bray-Curtis distance (Stress = 0). Stress describes the mismatch between the rank order of the pairwise dissimilarities between communities for Bray-Curtis distance and in the NMDS ordination. The NMDS was performed on raw abundance data. The PCA was performed on raw abundance data transformed into proportions so that the results are not influenced by differences in total abundance between communities and Euclidean distance on proportion excludes joint absences.

2.6 References

Abay, S., Kayman, T., Otlu, B., Hizlisoy, H., Aydin, F. & Ertas, N. (2014). Genetic diversity and antibiotic resistance profiles of Campylobacter jejuni isolates from poultry and humans in Turkey. International Journal of Food Microbiology, 178(16), 29-38

Anderson, M., Ellingsen, K., & McArdle, B. (2006). Multivariate dispersion as a measure of beta diversity. *Ecology Letters, 9*(6), 683-693.

Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L., et al. (2011). Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist: Roadmap for beta diversity. Ecology Letters, 14(1), 19–28. http://doi.org/10.1111/j.1461-0248.2010.01552.x

Barwell, L., Isaac, N., & Kunin, W. (2015). Measuring β - diversity with species abundance data. *Journal of Animal Ecology, 84*(4), 1112-1122.

Beck, J., Holloway, J.D. & Schwanghart, W. (2013) Undersampling and the measurement of beta diversity (ed D Orme). *Methods in Ecology and Evolution*, 4, 370–382.

Bennett, J. R., and Gilbert, B. (2016). Contrasting beta diversity among regions: how do classical and multivariate approaches compare?. *Global Ecology and Biogeography*.

Blaak, H., Van Hoek, A., Hamidjaja, R., Van Der Plaats, R., Kerkhof-de Heer, L., De Roda Husman, A., & Schets, F. (2015). Distribution, Numbers, and Diversity of ESBL-Producing E. coli in the Poultry Farm Environment. *PloS One, 10*(8), E0135402.

Bray, J., and J. Curtis. 1957. An ordination of the upland forest communities in southernWisconsin. Ecology Monographs 27: 325-349.

Caucci, S., Karkman, A., Cacace, D., Rybicki, M., Timpel, P., Voolaid, V., et al. (2016). Seasonality of antibiotic prescriptions for outpatients and resistance genes in sewers and wastewater treatment plant outflow. *FEMS Microbiology Ecology, 92*(5). doi:10.1093/femsec/fiw060

Chao, A., Chazdon, R. L., Colwell, R. K., and Shen, T.-J. (2005). A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecology Letters 8, 148–159.

Chao, A., Chazdon, R. L., Colwell, R. K., & Shen, T.-J. (2006). Abundance-Based

Similarity Indices and Their Estimation When There Are Unseen Species in Samples. Biometrics, 62(2), 361–371. http://doi.org/10.1111/j.1541-0420.2005.00489.x

Chao, A., and L. Jost. (2012). Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. Ecology 93:2533–2547.

Chao, A., Gotelli, N., Hsieh, T., Sander, E., Ma, K., Colwell, R., and Ellison, A. (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs, 84*(1), 45-67.

Clarke, K.R. (1993) Nonparametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, 18, 117–143.

Clarke, K.R. (1999). Nonmetric multivariate analysis in community-level ecotoxicology. Environmental Toxicology and Chemistry, 18(2), 118-127.

Clarke, K.R., Somerfield, P.J. & Chapman, M.G. (2006). On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray–Curtis coefficient for denuded assemblages. *J. Exp. Mar. Biol. Ecol.*, 330, 55–80.

Colwell, R. K., Chazdon, R. L., Chao, A. Y., Lin, S. J., Gotelli, N. X., Mao, C. T., & Longino, J. (2012). Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology, 5*(1), 3-21.

De la Pedrosa, E., Baquero, F., Loza, E., Nadal-Serrano, J., Fenoll, A., Del Campo, R., & Cantón, R. (2009). High clonal diversity in erythromycin-resistant Streptococcus pneumoniae invasive isolates in Madrid, Spain (2000–07). *Journal of Antimicrobial Chemotherapy, 64*(6), 1165-1169.

Desrochers, R., and Anand, M. (2004). From Traditional Diversity Indices to Taxonomic Diversity Indices, International Journal of Ecology and Environmental Science 30: 85-92

Gauch, H. (1982). *Multivariate analysis in community ecology* (Cambridge studies in ecology). Cambridge Cambridgeshire; New York: Cambridge University Press.

Gotelli, N.J., and Ellison, A.M. (2004). A Primer of ecological statistics. Sinauer Associates. Sunderland, Massachusetts.

Gotelli, N. J., and Ellison, A.M. (2012). A primer of ecological statistics. Second edition. Sinauer Associates, Sunderland, Massachusetts, USA.

Green, R. (1990). Data Analysis in Community and Landscape Ecology. (Book Review). *Biometrics, 46*(1), 287-288.

Heino, J., Mykrä, H., & Kotanen, J. (2008). Weak relationships between landscape characteristics and multiple facets of stream macroinvertebrate biodiversity in a boreal drainage basin. *Landscape Ecology, 23*(4), 417-426.

Hill, M. (1973). Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology, 54*(2), 427-432

Hill, T., Walsh, K., Harris, J., and Moffett, B. (2003). Using ecological diversity measures with bacterial communities. *FEMS Microbiology Ecology, 43*(1), 1-11.

Hong, P., Lee, B., Aw, M., Shek, L., Yap, G., Chua, K., et al. (2010). Comparative Analysis of Fecal Microbiota in Infants with and without Eczema (Fecal Microbiota of Infants). *PLoS ONE, 5*(4), E9964.

Jost, L. (2006). Entropy and diversity. *Oikos, 113*(2), 363-375.

Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology, 88*(10), 2427-2439.

Jost, L., Chao, A. and Chazdon, R. (2011) Compositional similarity and β (beta) diversity. Biological Diversity: Frontiers in Measurement and Assessment (eds A.E. Magurran & B.J. McGill), pp. 66–84. Oxford University Press, Oxford.

Kindt, R., and Coe, R. (2005). Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies. World Agroforestry Centre (ICRAF), Nairobi, Kenya. Web. 2015. Retrieved from http://www.worldagroforestry.org/downloads/Publications/PDFS/b13695.Pdf

Koleff, P., Gaston, K.J. & Lennon, J.J. (2003). Measuring beta diversity for presence-absence data. *J. Anim. Ecol.*, 72, 367–382.

Krebs, C. J. (1999). Ecological methodology (2nd ed). Menlo Park, Calif:Benjamin/Cummings.

Laliberte, E., and Legendre, P. (2010). A distance-based framework for measuring functional diversity from multiple traits. Ecology, 91, pp. 299–305

Legendre, P. and Gallagher, E.D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129, 271–280.

Legendre, P., and Legendre, Louis. (1998). *Numerical ecology* (2nd English ed.,

Developments in environmental modelling; 20). Amsterdam; New York: Elsevier.

Legendre, P., Borcard, D., & Peres-Neto, P. (2005). Analyzing Beta Diversity: Partitioning the Spatial Variation of Community Composition Data. *Ecological Monographs, 75*(4), 435-450.

Leinster, T., Cobbold, C., and Reeve, J. (2012). Measuring diversity: The importance of species similarity. *Ecology, 93*(3), 477-489.

Magurran, A. E. (1988). Ecological diversity and its measurement. Princeton, N.J: Princeton University Press.

Magurran, A.E. (2004). Measuring Biological Diversity. Blackwell, Oxford.

Magurran, A. E., and McGill, B. J. (Eds.). (2011). Biological diversity: frontiers in measurement and assessment. Oxford ; New York: Oxford University Press.

McLellan, S., Huse, S., Mueller-Spitz, S., Andreishcheva, E., & Sogin, M. (2010). Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environmental Microbiology, 12*(2), 378-392.

Morris, E., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T., et al. (2014). Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. *Ecology and Evolution, 4*(18), 3514-3524.

Mutlu, Ece A., Keshavarzian, Ali, Losurdo, John, Swanson, Garth, Siewe, Basile, Forsyth, Christopher, et al. (2014). A compositional look at the human gastrointestinal microbiome and immune activation parameters in HIV infected subjects. *PLoS Pathogens, 10*(2), PLoS Pathogens, Feb, 2014, Vol.10(2).

Pakyz, A., Powell, J., Harpe, S., Johnson, C., Edmond, M., & Polk, R. (2008). Diversity of antimicrobial use and resistance in 42 hospitals in the United States. *Pharmacotherapy, 28*(7), 906-12.

Paliy, O. and Shankar, V. (2016). Application of multivariate statistical techniques in microbial ecology. Mol Ecol, 25: 1032–1057. doi:10.1111/mec.13536

Palmer, M.W. (2006). Ordination Methods for Ecologists—an overview. Botany Department, Oklahoma State University, Stillwater, Oklahoma 74078 USA. Web. 2015. Retrieved from http://ordination.okstate.edu/overview.htm

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, Series 6, vol. 2, no. 11, pp. 559-572.

Pielou, E. C. (1975). Ecological diversity. New York: Wiley InterScience

Price, L., Liu, C., Melendez, J., Frankel, Y., Engelthaler, D., Aziz, M., et al. (2009). Community Analysis of Chronic Wound Bacteria Using 16S rRNA Gene-Based Pyrosequencing: Impact of Diabetes and Antibiotics on Chronic Wound Microbiota (Chronic Wound Microbiota). *PLoS ONE, 4*(7), E6462.

Ramette, A. (2007). Multivariate analyses in microbial ecology. FEMS Micorbiol Ecol 142-160. doi: http://dx.doi.org/10.1111/j.1574-6941.2007.00375.x

Ricotta, C., S. Pavoine, G. Bacaro, and A. T. R. Acosta. (2012). Functional rarefaction for species abundance data. Methods in Ecology and Evolution 3:519–525.

Shannon, C. E. and Weaver W. (1948) A mathematical theory of communication. The Bell System Technical Journal, 27, 379–423 and 623–656.

Shankar, V., Agans, R., Holmes, B., Raymer, M., & Paliy, O. (2013). Do gut microbial communities differ in pediatric IBS and health? *Gut Microbes, 4*(4), 347-52.

Shepard RN. 1962.Analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika* **27**:125–140.

Sigala, J., & Unc, A. (2013). Pyrosequencing estimates of the diversity of antibiotic resistant bacteria in a wastewater system. *Water Science and Technology: A Journal of the International Association on Water Pollution Research, 67*(7), 1534-43.

Soetaert, K., and Heip, C. (1990). Sample-size dependence of diversity indices and the determination of sufficient sample size in a high-diversity deep-sea environment. *Marine Ecology Progress Series, 59*, 305-307.

Stirling, G., & Wilsey, B. (2001). Empirical Relationships between Species Richness, Evenness, and Proportional Diversity. *The American Naturalist, 158*(3), 286-299.

Tuomisto, H., & Ruokolainen, K. (2006). Analyzing or Explaining Beta Diversity? Understanding the Targets of Different Methods of Analysis. *Ecology*, *87*(11), 2697–2708. Retrieved from http://www.jstor.org/stable/20069289

Tuomisto, H. (2010a). A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography*, 33, 2–22.

Tuomisto, H. (2010b). A diversity of beta diversities: straightening up a concept

gone awry. Part 2. Quantifying beta diversity and related phenomena. *Ecography*, 33, 23–45.

Walker, S. C., M. S. Poos, and D. A. Jackson. (2008). Functional rarefaction: estimating functional diversity from field data. Oikos 117:286–296.

Web of Science [v.5.19]. Biological Diversity. All years. Web. 2016. Retrieved from http://apps.webofknowledge.com.qe2aproxy.mun.ca/Search.do?product=WOS&SID=1Dv2eWWpPOmiJlLR1sT&search_mode=GeneralSearch&prID=1cece76f-7f41-4a43-918b-7058d658379f

Whittaker, R. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs, 30*(3), 279-338.

Whittaker, R. H. (1972). Evolution and Measurement of Species Diversity. *Taxon*, *21*(2/3), 213. http://doi.org/10.2307/1218190

Zhang, Bo, Zhang, Zi-wei, Meng, Fan-Yu, Wu, Qiong, Xu, Shi-Wen, & Wang, Xiao-Long. (2014). Analysis of the structure of bacteria communities and detection of resistance genes of quinolones from pharmaceutical wastewater. *Annals of Microbiology, 64*(1), 23-29.

Chapter 3

Application of Ecological Diversity Indices to Antibiotic Resistance

3.1 Introduction

Many studies of antibiotic resistance report the resistance of pathogens to specific antibiotics or classes of antibiotics (Di Giulio et al. 2015, Dyar et al. 2012, Abadi et al. 2010, Nys et al. 2004) with only limited evaluation of antibiotic resistance across a broad range of antibiotics. Recent studies support the need for indices summarizing antibiotic resistance across antibiotics, pathogen species, and disease syndromes (Laxminarayan and Klugman, 2011, Ciccolini et al. 2015). Given the success of biodiversity measures in ecology, there is great potential for this approach to be applied to summarize antibiotic resistance data.

The utility of applying alpha diversities to antibiotic resistance data is to quantify the uncertainty surrounding which antibiotics a bacterial isolate is resistant to. This

information is useful when infections are treated without prior susceptibility testing. For example, diseases such as urinary tract infections in adults (Holm et al. 2015), sepsis, meningitis, pneumonia, pyelonephritis, gastroenteritis, osteomyelitis, and cellulitis in children (Bruel et al. 2007) are diagnosed and treated without any knowledge of antibiotic susceptibility. If the population of bacteria associated with particular symptoms is highly diverse with respect to its antibiotic resistance then there is a high probability that different patients are infected with pathogens that have different resistance types. Therefore, it is less likely that an antibiotic that is successful for one patient will also be successful for the next patient. As such, for a population of bacteria that are highly diverse with respect to antibiotic resistance, when antibiotic susceptibility is unknown, the recommended strategies for prescribing antibiotics may be different than for populations with less antibiotic resistance diversity. Previous studies have used alpha diversity indices to quantify species or molecular diversity of resistant bacteria isolates (Pakyz et al. 2008, de la Pedrosa et al. 2009, Sigala and Unc, 2013, Abay et al. 2014, Blaak et al. 2015, Zhang et al. 2014), however no applications of alpha diversity measures to antibiotic-resistant populations have focused on the diversity of antibiotic resistance.

Beta diversity indices measure the similarity in the antibiotic resistance between regions. Regions that have a similar distribution of antibiotic resistance will have low beta diversity, appear close in an ordination, and the best antibiotic treatment strategies for patients in these regions will be similar. Similar to alpha diversity indices, all the previous applications of beta diversity measures and ordination techniques have focused on the species or molecular composition of bacterial communities (Price et al. 2009, Hong et al. 2010,

McLellan et al. 2011, Sigala and Unc, 2013, Shankar et al. 2013, Mutlu et al. 2014, Caucci et al. 2016) and not the composition with respect to antibiotic resistance.

In this chapter, we explore the utility of ecological diversity indices in summarizing patterns of antibiotic resistance. In the methods section, we discuss approaches for applying alpha and beta diversity indices to antibiotic resistance data. Then, we illustrate the application of our proposed methods for a case study that analyzes data from the Canadian Integrated Program for Antimicrobial Resistance Surveillance (CIPARS) on antibiotic resistance in clinical *Salmonella enterica* isolates collected from across Canada (CIPARS, 2013). In this case study, we measure SW diversity, an alpha diversity index, to quantify the uncertainty surrounding which antibiotics clinical *S. enterica* isolates are resistant to. Then, we test whether SW diversity is affected by different Canadian provinces, *S. enterica* serotypes, time periods or combinations of these factors.

We measure Bray-Curtis distance, a beta diversity index, to measure similarity in antibiotic resistance. We perform ordinations to visualize differences in the relative resistance to particular antibiotics between provinces, serotypes, time periods or combination between them. We conclude with recommendations for the future use of ecological diversity indices to analyze and understand antibiotic resistance data.

3.2 Material and Methods

3.2.1 Applying diversity indices to antibiotic resistance data

An isolate is a pathogen isolated from a host or surface. For a sample of isolates, when measuring alpha diversity to quantify the uncertainty regarding which antibiotics the population of isolates is resistant to, we need to define p_i (as it appears in equation 1 and 2 in Section 2.2) with regard to resistance. Suppose we let p_i be the fraction of pathogens resistant to the antibiotic *i*, and *s* be the total number of antibiotics considered. Under these definitions the sum $\sum_{i=1}^{s} p_i$ may exceed 1 and the properties (ii) and (iii) of a diversity index as described in Section 2.2 no longer hold or make sense. Furthermore, interpretations of the meanings of the values of the SW and Simpson's index as described in Section 2.2 no longer hold. We suggest two possible solutions to this problem:

1) The line-listed approach. This approach considers each resistance type (i.e., the complete combination of antibiotics that an isolate is resistant to for all isolates that are resistant to at least one antibiotic; see Table 3.1) such that p_i in equation 1 and 2 is the proportion of isolates belonging to a resistance type *i*, and s is the total number of resistance types; and

2) The aggregated approach. This approach considers resistance to each *antibiotic* (Table 3.1) such that p_i in equation 1 and 2 is the proportion of isolates resistant to a particular antibiotic, *i,* relative to the total frequency of resistance summed across all antibiotics, f_i, i.e.

$$p_i = f_i / \Sigma f_i, \quad (5)$$

and s in equation 1 and 2 is the total number of antibiotics.

When calculating beta diversity using the 'classical approach' (equation 3 in Section 2.3, Chapter 2) the same challenges arise as were described above for the application of alpha diversity indices to resistance data: specifically, p_i must be defined so that the $\sum_{i=1}^{S} p_i = 1$. However, when beta diversity is calculated using multivariate dissimilarity measures there is no such restriction that the approach need only be applied to data scaled to sum to one. Therefore, we define X_{ak} and X_{bk} (as it appears in equation 4 in Section 2.3, Chapter 2) as the number or frequency of isolates that are resistant to antibiotic k recovered from the population a and population b, respectively. This is referred to as an aggregated approach because 'k' indexes different antibiotics and not different resistance types, as a line-listed approach would. One isolate might be resistant to more than one antibiotic, however the application of multivariate beta diversity measures does not require a unique categorization of each isolate in the way that the alpha diversity measures do.

If beta diversity and the resulting ordinations are performed on the raw abundance data (i.e. frequency of resistance to different antibiotics), populations within a NMDS ordination that appear close to each other have similar magnitudes of resistance to the same antibiotics. If beta diversity analyses and ordinations are performed on abundance data transformed into proportions (i.e. the frequency of resistance to different antibiotics is normalized to sum to 1), then X_{ak} and X_{bk} (as they appear in equation 4 in Section 2.3) are the relative number of isolates that are resistant to the antibiotic k recovered from populations a and b, respectively, and under this transformation populations within a PCA ordination that appear close to each other have low dissimilarity as measured by Euclidean distance. For a PCA, the vectors represent different antibiotics and point in the direction of higher than an average relative abundance of particular antibiotics.

In the analyses that follow (Section 3.3), alpha diversity is calculated based on the line-listed approach and beta diversity is calculated using the aggregated approach (see Figure 3.1). The reasons for these choices will be discussed in more detail in section 3.4.1.

3.2.2 Data Summary

We chose to illustrate our method of applying biodiversity measures to antibiotic resistance data using *S. enterica* clinical isolates from CIPARS. To facilitate replication of our results, the data request form is provided in the Appendix. This dataset was selected due to the large number of isolates collected (over 35,000 isolates between January 1, 2003 and December 31, 2013) and because it is one of the few individual isolate level (line-listed) data sources available: more often antibiotic resistance data are available as population level frequencies of resistance to individual antibiotics (aggregated data). The human *S. enterica* isolates were tested for their susceptibility to a large number of antibiotics (17), which facilitated our goal of summarizing antibiotic resistance across a wide range of antibiotics. Even though gastroenteritis arising from *S. enterica* infections are not typically treated with antibiotics, the analysis would be similar for hospital-acquired (i.e., *Staphylococcus aureus*, *Enterococcus spp*., *Enterobacteriaceae*, *Pseudomonas aeruginos*a and *Acinetobacter spp*.) and community-acquired (i.e., *Mycobacterium tuberculosis*, *Streptococcus*, and pathogens causing Sexually Transmitted Infections) pathogens that are treated with antibiotics.

Provincial public health laboratories receiving high volumes of *S. enterica* isolates from hospital-based and clinical laboratories were responsible for providing a detailed set of

data to CIPARS for each *S. enterica* isolate they received (CIPARS, 2013). Eight serotypes: Enteritidis, Heidelberg, Newport, Paratyphi A, Paratyphi B, Typhi, Typhimurium, and I 4,[5],12:i:-, were routinely tested for susceptibility to 17 antibiotics (Amoxicillin-clavulanic acid, Amikacin, Ampicillin, Azithromycin, Ceftiofur, Ceftriaxone, Cefoxitin, Cephalothin, Chloramphenicol, Ciprofloxacin, Gentamicin, Kanamycin, Nalidixic acid, Sulfisoxazole, Streptomycin, Tetracycline, Trimethoprim-sulfamethoxazole), and the remaining isolates were stored for future susceptibility tests.

We excluded the serotypes Typhi and Paratyphi A and B from our analysis, as these serotypes are known to cause the typhoidal cases of *S. enterica* (CIPARS, 2013). As we were interested in studying the antibiotic resistance of *S. enterica* in different provinces of Canada, we only included the non-typhoidal cases, as the typhoidal cases are usually contracted during international travel. We also excluded the Canadian territories (Yukon, Northwestern Territories, and Nunavut) from our analysis because fewer than 3 isolates were submitted. The provinces of Prince Edward Island and Newfoundland and Labrador were also excluded from our alpha diversity analyses due to low sample sizes (see Section 3.2.3 for a more detailed explanation).

The line-listed data received from CIPARS contained designations of resistance types that included partial resistance (information on the breakpoints for minimum inhibitory concentrations for *S. enterica* serotypes for each antibiotic can be found in the CIPARS Annual Report; CIPARS, 2013). All occurrences of partial resistance were considered to be susceptible to that antibiotic.

The summary of the data on antibiotic resistance in clinical *S. enterica* isolates used for our analysis consists of 6737 resistant isolates each having been collected in one of the ten Canadian provinces and belonging to one of five different serotypes belonging to *S. enterica* (Table 3.2). Most isolates were resistant to multiple antibiotics (Table 3.2).

3.2.3 Alpha diversity

Alpha diversity is a population level measure and can only be calculated for a group of isolates, not an individual isolate. We defined our sampling unit as consisting of 30 isolates: the more isolates per sampling unit the more accurate the estimate of diversity within the sampling unit, but there are fewer sampling units reducing the sample size of the data set (see Section 3.4.4 for a discussion of the effect of sampling unit size). We chose to define a sampling unit according to a fixed number of isolates rather than a fixed length of time because alpha diversity are known to be sensitive to sample sizes (Soetaert & Heip, 1990, Chao et al. 2014) and because a fixed sample size approach is generally recommended (Magurran, 2004, pp. 133, Chao et al. 2014). While *S. enterica* exhibits a natural one-year cycle, discretizing the sampling period into yearly increments creates artificial break points: January 1, 2016 and December 31, 2016 are grouped into year 2016 despite being nearly a year apart, while December 31, 2015 is grouped into year 2015 despite being only 1 day prior to January 1, 2016. In addition, if we were to create sampling units based on a fixed amount of time, i.e. based on seasons or yearly increments, then the number of isolates would be different within each sampling unit and thus the alpha diversity estimates would not be accurate for comparing populations.

We will determine if province, serotype and time affect alpha diversity. We hypothesize that province may affect alpha diversity because agricultural practices, including antibiotic use, may differ between provinces. Also, provinces with higher population density are likely to differ from provinces with lower population density. Similarly, we hypothesize that the serotype a *Salmonella* isolate belongs to may affect alpha diversity because different serotypes have different growth rates and virulence (Su et al. 2004, Kim et al. 2005, Beceiro et al. 2013). A serotype that reproduces and/or mutates more rapidly could be expected to have higher alpha diversity. Likewise, we hypothesize that over time the alpha diversity may have changed because the antibiotic use may differ over time and also new types of resistance in *Salmonella* may have emerged over time (Glynn et al. 1998, Su et al. 2004).

We subsetted our data by selecting all isolates of a particular serotype and province, ordering by time and forming groups of 30 isolates. This was repeated for all province and serotype combinations. As such, each sampling unit of 30 isolates has an associated province, serotype and end time for when the 30 isolate group is completed. For each sampling unit, alpha diversities can be calculated using either the line-listed or aggregated approach. We were not able to include the provinces of Prince Edward Island (PEI) or Newfoundland and Labrador (NL) in our analysis because fewer than 30 isolates were recovered from PEI for different serotypes over the entire surveillance period and only one serotype (Heidelberg) had more than 30 isolates for NL (see Table A3.1 for distribution of isolates in PEI and NL for different serotypes).

In order to determine whether SW diversity varied between provinces, serotypes and over time we considered several different general linear models and used an information theoretic model selection approach (Table 3.3). The model fitting and selection was performed only for the line-listed approach (Section 3.2.1) since alpha diversity values generated using the aggregated approach are not meaningful (discussed in Section 3.4.1). Following the line-listed approach, SW diversity quantifies the uncertainty regarding the resistance type for a population of *Salmonella* isolates belonging to a particular serotype and collected from a particular province over a period of time. If the SW diversity for a population of isolates is high, then there is a high diversity of the resistant types. This corresponds to a high level of uncertainty regarding the resistant type of a population because in the absence of susceptibility testing it is more difficult to anticipate the antibiotics that an isolate is resistant to, as there are a wide variety of relatively abundant resistance types present in the population (see Section 2.2). For our study, we only considered resistant isolates because the interpretation regarding the uncertainty surrounding the antibiotics an isolate is resistant to would not be possible if susceptible isolates were included (further discussed in Section 3.4.4).

For the fitted models, our response variable was the SW diversity and we considered all the possible combinations of three predictor variables province, serotype and time. SW diversity and time are on a continuous scale, whereas provinces and serotypes are categorical variables. The time variable is the end time for each sampling unit of 30 isolates.

Model selection was performed using Akaike Information Criteria corrected for small sample size (AICc) using the package AICcmodavg (version 2.0-4) in R. The top-ranked model has the lowest AICc score. The deltaAICc is the difference between the AICc of a particular model and the top ranked model and we consider models with deltaAICc<7 to be no different (Bolker, 2008 pp.210). We fit coefficients for the general linear models using the lm() function in R (version 3.2.1). We evaluated the goodness of fit with multiple R^2 values from the model output (Anderson, 2008 pp.94-95). All the figures were made using either default R graphics package or ggplot2 (version 2.1.0) with cowplot (version 0.6.1.9999) and ggrepel (version 0.5). The CrossTable for testing independence between the categorical variables province and serotype was created using package gmodels (version 2.16.2) and the heat map for province-serotype distribution was created using plot3d (version 1.1).

Even though the model fitting and selection was performed only for the line-listed approach (Section 3.2.1), we did calculate SW diversity values for each province and serotype using the aggregated approach to determine how similar these results are to those generated with the line-listed approach.

3.2.4 Beta diversity

For beta diversity, it is not necessary to compare sampling units of equal size (as it was for the alpha diversity, see Section 3.2.3), because beta diversity is less affected by incomplete sampling (Beck et al. 2013, Barwell et al. 2015, Bennett and Gilbert, 2015). Hence, for beta diversity measurements with respect to time, we considered different surveillance periods

discretized by year (2003-2013). The significance of considering the year to year differences in beta diversity is that if the beta diversity between consecutive years is high, then it may be necessary to re-evaluate antibiotic prescription policies annually. We will determine the provinces, serotypes and years that have similar levels of antibiotic resistance to the same antibiotics. We hypothesize that some provinces may have similar antibiotic resistance because they are in geographic proximity, have similar agriculture practices or antibiotic use, a similar number of poultry farms, and/or similar population density. We hypothesize that some serotypes may have similar antibiotic resistance due to similar epidemiology and mutation rates. Time periods (years) may have similar antibiotic resistance due to similar antibiotic use in consecutive years.

For the beta diversity analysis we used the aggregated approach as described in Section 3.2.1 (see also Figure 3.1). We performed PCA and NMDS ordinations to identify provinces, serotypes, and years that are similar with respect to the frequency or relative frequency (normalized within a province, serotype or year) of resistance to particular antibiotics. We performed a PCA with Scaling 1 (variance-covariance matrix) on the raw abundance data transformed to proportions using the prcomp() function from the basic R package. We performed NMDS with Bray-Curtis distance using metaMDS() function from the vegan (2.3-5) package on the raw abundance data. The reasons for choosing transformed or raw abundance data will be discussed in Section 3.4.2.

3.2.5 Model/Data Analysis Assumptions

All models assume normally distributed errors, independence of the predictor variables, and variance homogeneity and all of these assumptions were satisfied except that a weak correlation was found between time and the other variables for some models (Table 3.3; see Table A3.2 and Figures A3.1, A3.2, A3.3, and A3.4 in Appendix). Rarefaction and extrapolation curves (Figure A3.5 in Appendix) constructed following Chao et al. (2014) using the iNext (version 2.0.8) package in R tended to reach an asymptote at about 100 isolates per sample as shown in Figure A3.5.

3.3 Results

Following the line-listed approach, we found that province Ontario has the highest species richness of the resistance types (193), whereas Nova Scotia has the lowest (49, Table 3.2; excluding Prince Edward Island and Newfoundland and Labrador from the alpha diversity analysis). *S. enterica* serotype Typhimurium has the highest species richness of the resistance types (210), whereas Newport has the lowest (61, Table 3.2). When fitting models for the alpha diversity values using the line-listed approach, we found that the top ranked model (model 1, AICc = 185.40, Table 3.3) consists of province, serotype, time, and an interaction term between serotype and time to predict the SW diversity (Table 3.4) and explains a large proportion of the variation in SW diversity (multiple $R^2 = 0.55$, Table 3.3). The model with all three predictor variables and a three-way interaction (model 10, Table 3.3) and the model with all three predictor variables and a two-way interaction between province and serotype (model 11, Table 3.3) explained the largest proportion of variation

in the SW diversity (Multiple R^2 =0.58). However, these models (models 10 and 11, Table 3.3) had a large number of parameters (K=60 and K=34 respectively, Table 3.3) compared to the top ranked model (model 1, K=17, Table 3.3).

We present our results for the alpha (line-listed approach) and beta (aggregated approach) diversity by splitting according to the province, serotype, and time.

3.3.1 Results – Province

Alpha diversity (Line-listed approach) – How is the uncertainty in resistance type affected by provinces?

Among the three predictor variables, the province only explains a small amount of variance in SW diversity (multiple R^2 = 0.07; model 15, Table 3.3) and there is substantial overlap in the 95% confidence intervals (CI). We found that the SW diversity of resistance types is highest for the province of Alberta (1.88 with 95% CI [1.66, 2.09]) and lowest for Nova Scotia (1.39 with 95% CI [0.70, 2.07]; Figure 3.2A; Table A3.3 in Appendix). In terms of the effective diversity (see section 2.2), Alberta has a mean diversity equivalent to 6.55 equally abundant resistance types, whereas Nova Scotia has a diversity equivalent to 4.01 equally abundance resistance types. As shown in Figure 3.2B, Nova Scotia, which has the lowest SW diversity, is predominated by just one resistance type, whereas Alberta having the highest SW diversity has the lowest relative abundance of the most abundant resistance type and the distribution of resistance types is more even.

When calculating SW diversity for provinces using the aggregated approach we found a strong correlation with the values of SW diversity as calculated using the line-listed approach (Figure 3.3A).

Beta diversity (Aggregated approach) - Which provinces have similar levels of resistance and relative resistance to the same antibiotics?

We found that Alberta and British Columbia were very close in the NMDS graph (Figure 3.2C) showing low beta diversity or high similarity in the frequency of resistance to the same antibiotics (see also Figure 3.4A). As shown in the PCA (Figure 3.2D), Alberta is least similar to all New Brunswick, Nova Scotia, and Newfoundland and Labrador. Consulting PC1 (72.4%), Alberta has higher than average relative resistance to Tetracycline and Streptomycin compared to New Brunswick, Nova Scotia, and Newfoundland and Labrador (Figure 3.2D; see also Figure 3.4A). Nova Scotia is dissimilar to New Brunswick, and other provinces except Prince Edward Island, because of very high relative resistance to Nalidixic acid (Figure 3.2D, PC2 (16.83%); see also Figure 3.4A). Consulting PC1 New Brunswick, Nova Scotia, and Newfoundland and Labrador appear similar because of similar relative resistance (higher than average) to Cefoxitin (Figure 3.2D; see also Figure 3.4A).

3.3.2 Results – Serotype

Alpha diversity (Line-Listed approach) - How is the uncertainty in resistance type affected by serotypes?

Amongst the three predictor variables, serotype explains most of the variance in SW diversity (multiple R^2 = 0.45; model 9, Table 3.3). We found that the *S. enterica* serotype Newport had the highest SW diversity of resistance types (2.44 with 95% CI [1.92, 2.97]) and *S. enterica* serotype Enteritidis had the lowest (1.17 with 95% CI [1.06, 1.28]; Figure 3.5A; Table A3.4 in Appendix). For Newport, the effective diversity is 11.47 equally abundant resistance types as compared to 3.22 for Enteritidis. As shown in Figure 3.5B, Enteritidis is predominated by just one resistance type, whereas Newport has the lowest relative abundance of the most abundant resistance type.

The 95% CI around the mean SW index values between serotypes are non-overlapping suggesting that the mean SW diversity for serotype Enteritidis is different from all other serotypes. In addition, the SW diversity of both Heidelberg and I 4,[5],12:i:- serotypes are different from Newport and Typhimurium, and also Newport is different from Typhimurium (Table A3.4).

Similar to provinces, the SW diversity values for serotypes using the aggregated approach were strongly correlated with the values of SW diversity as calculated using the line-listed approach (Figure 3.3B).

Beta diversity (Aggregated approach) - Which serotypes have similar levels of resistance and relative resistance to the same antibiotics?

The antibiotics that the serotype Enteritidis was resistant to were least similar to Heidelberg and Typhimurium, whereas the antibiotics serotype I 4,[5],12:i:- and Newport

were resistant to was more similar as shown in the NMDS (Figure 3.5C).The PCA (Figure 3.5D; see PC1 which explains 71.02% of the variance) shows that serotype Enteritidis has very high levels of relative resistance (higher than average) to Nalidixic acid (see also Figure 3.4B) compared to the other serotypes. Consulting PC2 (24.11%), Typhimurium and I 4,[5],12:i:- appear close because of similar level of relative resistance to Tetracycline and Sulfisoxazole (higher than average) and far from Heildeberg (Figure 3.5D) which has low relative resistance (lower than average) to Tetracycline and Sulfisoxazole and instead high levels of relative resistance to Ampicillin and Ceftriaxone (see also Figure 3.4B).

3.3.3 Results – Time

Alpha diversity (Line-listed approach) - How does the uncertainty in resistance types change over the surveillance period?

Among the three predictor variables, time explains the least amount of the variance in SW diversity (multiple $R^2 = 0.03$; model 13, Table 3.3). The coefficient associated with the time variable was -8.45×10^{-5} and the surveillance period that we considered lasted a total of 4016 days. As such, the best model implies a decrease in SW diversity from 1.84 to 1.50 over the surveillance period (Figure 3.6A). In terms of effective diversity, this is a decrease of 6.30 to 4.48 equally abundant resistance types.

Beta diversity (Aggregated approach) - Which years have similar levels of resistance and relative resistance to the same antibiotics?

As shown in the NMDS (Figure 3.6B), consecutive years are grouped together and, as such, the frequency of resistance for each antibiotic is similar between the consecutive years. Consulting the PCA (Figure 3.6C, PC1 (55.31%)), we see years 2011 and 2012 overlapping and 2013 appearing close to 2011 and 2012 because of a similar level of relative resistance (higher than average) to Amoxicillin-clavulanic acid and Ceftriaxone (see also Figure 3.4C). Consulting PC2 (28.70%), years 2007-2009, 2011, and 2012 appear close because of similar level (higher than average) of relative resistance to Nalidixic acid (see also Figure 3.4C).

3.3.4 Alpha Diversity Results – Serotype and Time

How does the uncertainty in resistance types for serotypes change over the surveillance period?

The top ranked model consists of an interaction between predictor variables serotype and time (model 1, Table 3.3). For serotype Enteritidis, there was virtually no change in SW diversity, whereas SW diversity for serotype Typhimurium increased over the surveillance period and decreased for all other serotypes (Figure 3.7A, Table 3.4). The coefficient associated with the time variable for serotype Typhimurium was 8.23×10^{-5} (model 6, Table 3.3) and as such SW diversity for Typhimurium increased from 1.92 to 2.25 (Figure

3.7A), which is an increase of 6.82 to 9.50 equally abundant resistant types. However, the greatest change in SW diversity was for serotype Heidelberg. The coefficient associated with time variable for Heidelberg was -1.75×10^{-4} (model6, Table 3.3) and hence over the surveillance period SW diversity for Heidelberg decreased from 1.90 to 1.19 (Figure 3.7A), which is a decrease of 6.69 to 3.29 equally abundant resistance types.

3.3.5 Beta Diversity Results – Combinations

Which Serotype and Year combinations have similar levels of resistance and relative resistance to the same antibiotics?

When performing the ordination on different serotype year combinations we found that the level of resistance to particular antibiotics was more similar for serotypes than for years as shown in the NMDS (Figure 3.7B). In addition, I 4,[5],12:i:- and Newport serotypes there are some similarities at the year level. Consulting PCA (Figure 3.7C, PC1 (67.18%)), serotype Enteritidis in all the years from 2003 to 2013 was different than other serotypes because of a high level of relative resistance to Nalidixic acid (see also Figure 3.4B).

Which Province and Serotype combinations have similar levels of resistance and relative resistance to the same antibiotics?

When performing the ordination on different province serotype combinations we found that the level of resistance to particular antibiotics was more similar for serotypes than provinces as shown in the NMDS (Figure A3.6A in Appendix). In addition, Typhimurium, I

4,[5],12:i:- and Newport serotypes there are some similarities at the province level. Consulting PCA (Figure A3.6B, PC1 (69.01%)), all the provinces that had serotype Enteritidis were different than other provinces because of a high level of relative resistance to Nalidixic acid (see also Figure 3.4B).

Which Province and Year combinations have similar levels of resistance and relative resistance to the same antibiotics?

The ordination for different province year combinations does not show a pattern, except that Prince Edward Island has a very different frequency of antibiotic resistance to particular antibiotics than all other provinces and the distribution for Prince Edward Island was different in every year (Figure A3.7A in Appendix). Consulting the PCA (Figure A3.7B, PC1 (46.01%)), Prince Edward Island in years 2009 and 2011 had very high levels of relative resistance to Nalidixic acid.

3.4 Discussion

The emergence of antibiotic-resistant organisms is a serious threat to public health as it limits the efficacy of antibiotic therapy and may increase the emergence rate of novel multidrug resistance (MDR) pathogens (Wright et al. 2005, Tanwar, 2014). Antibiotic-resistant populations are capable of causing severe, life-threatening infections that may be more difficult to manage when treatment options are limited (Martin et al. 2004, Helms et al. 2004, Varma et al. 2005, Eguale et al. 2016). We provide the first application of

biodiversity measures to summarize the antibiotic resistance across a broad range of antibiotics and illustrate our method on a case study on antibiotic resistance in *S. enterica* in Canadian jurisdictions. *S. enterica* infections are generally not treated with antibiotics in humans (unless a bloodstream infection occurs) and so our results do not have direct implications for how antibiotics are prescribed for *S. enterica*, however, our methods could be applied to other pathogens where there are clear implications for antibiotic treatment strategies.

3.4.1 Justification of analysis of line-listed versus aggregated data

The contribution of our work is principally a methodological one. We showed that alpha diversity measures, in particular SW diversity, can be used to quantify the uncertainty surrounding which antibiotics a bacterial isolate is resistant to. For antimicrobial resistance, we recommend using a line-listed approach when measuring alpha diversity, as the interpretations of the meanings of alpha diversity indices do not hold for the aggregated approach (see Table 3.1 for examples of line-listed and aggregated data). For example, consider the two scenarios as shown in Table 3.5. In scenario A, all the isolates are resistant to all the antibiotics, and in scenario B all the isolates are resistant to only one antibiotic. For the line-listed approach, Scenario A and Scenario B are judged as different because they have different SW diversity (equation 1; Table 3.5); however for the aggregated approach, both scenarios A and B, have the same SW diversity (Table 3.5). For scenario A there is little uncertainty regarding which antibiotics an isolate in the

population will be resistant to, as it will most certainly be resistant to all of them, while for scenario B there is a high degree of uncertainty and this difference is not reflected in the SW diversity calculations when using the aggregated approach. Therefore, the aggregated approach is not consistent with the interpretation of the SW diversity index as quantifying the uncertainty in the identity of an isolate in the population (as described in Section 2.2).

Practically speaking, aggregated data is more widely available; however line-listed resistance data is necessary to make a meaningful calculation of alpha diversity. That being said, we did find a strong correlation between the alpha diversity (SW diversity) results for the line-listed and aggregated approaches (Figure 3.3). Hence, the results obtained using aggregated approach for our case study is close to correct as obtained using the line-listed approach. To make use of aggregated data, future research might create models to predict the most likely line-listed data from the aggregated data and seek to understand under what assumptions aggregated data would be sufficient to accurately approximate the line-listed results.

Another relationship between line-listed and aggregated data (Table 3.1) arises due to the doubling property of Hill numbers (Hill, 1973, Chao et al. 2014). This doubling property applies to the SW diversity index when expressed as effective diversity (see Section 2.2). To illustrate the doubling property, consider a population that consists of one isolate that is resistant to both Amoxicillin and Amikacin. Under the aggregated approach, the normalized frequency of resistance to each of Amoxicillin and Amikacin is 0.5 (see equation 5 in Section 3.2.2), and for all other antibiotics is zero. The effective diversity ($e^{H'}$) is 2. Under the line-listed approach, the proportion of the isolates with resistance type 'AmAk'

(resistant to both Amoxicillin and Amikacin) is one and for all other resistance types is zero. Under the line-listed approach, the effective diversity is 1, and so the effective diversity is doubled under the aggregated approach. If we extend this example by considering a population consisting of one isolate that is resistant to three antibiotics, we find that the effective diversity is tripled under the aggregated approach. These calculations reveal that the SW diversity calculated under the aggregated approach is always greater than or equal to the SW diversity calculated under the line-listed approach. Hence, it is not possible to compare SW diversity values between populations unless the same approach has been used.

For a single population the SW diversity could be calculated using both approaches, and when doing so, the ratio of the effective diversities calculated under either approach is related to the prevalence of multiple resistance. As shown in the regression of SW diversity under the aggregated approach (vertical axis) against the line-listed approach (horizontal axis; Figure 3.3A**)**, provinces New Brunswick, Quebec, and Ontario are above the regression line and referring to Table 3.2, these provinces rank 1st, 2nd and 4th in terms of the most multiple resistance.

We also showed that measuring beta diversity where these results are visualized using ordinations provides information on the similarity in the frequency and relative frequency of antibiotic resistance to the same antibiotics between regions, serotypes or time periods. For antimicrobial resistance data, we recommended using the aggregated approach when measuring beta diversity because from a public heath perspective we are interested in knowing which antibiotics are likely to succeed and which are likely to fail due to

resistance. If beta diversity is measured using the line-listed approach, then the regions that appear close in the ordination have similar frequencies of the same resistance types. Under the list-listed approach, a region consisting only of the resistance type 'AmAkTeCf' (resistance to Ampicillin, Amikacin, Tetracycline, and Cefoxitin) is equally dissimilar to a region with only the resistance type 'AmAkTe' (resistance to Ampicillin, Amikacin, and Tetracycline) and to a region with only the resistance type 'Na' (resistance to Nalidixic acid), however the antibiotics that can be used to successfully treat infections in the 'AmAkTeCf' and the 'AmAkTe' regions are similar. Under the aggregated approach, regions that have similar abundances of resistance to the same antibiotics appear close to each other. In this instance, the regions that appear near to each other should avoid prescribing the same antibiotics.

3.4.2 Ordination and data transformation to proportions

We performed the PCA on raw abundance data transformed into proportions because the transformation normalized the frequencies of antibiotic resistance within the provinces, serotypes and years. This was necessary because it is only reasonable to suggest that a particular antibiotic be avoided if other more effective antibiotics are available. Eighty percent resistance to ciprofloxacin, in of itself, does not suggest that ciprofloxacin should be avoided. In fact, if the population is 90% resistant to all other antibiotics, then ciprofloxacin should be preferred. Hence, it is the relative frequency of resistance to each antibiotic that is important.

On the other hand, we performed the NMDS on the frequency of resistance data without any normalization so that provinces, serotypes and years that have similar frequencies of resistance to the same antibiotics would appear near to each other in the ordination. In this instance, when two populations both have 80% resistance to ciprofloxacin, all else equal, these two populations will appear nearer to each other than to populations with low frequencies of resistance to ciprofloxacin.

3.4.3 *Salmonella spp.* data analysis and public health implications

Our top ranked model (model 1, Table 3.3) was a combination of all three predictor variables and an interaction between serotype and time, and explained 55% of the variance in the SW diversity index. We recognize there are other factors that are contributing to the variance in the diversity of resistance types. Further, we found that that the diversity of resistance types for *S. enterica* serotype Enteritidis was different than all other serotypes and none of the provinces differed (Tables A3.3 and A3.4). This is in agreement with the PCA and NMDS results which showed that most of the variation in the resistance pattern was due to serotype Enteritidis, which had very high resistance to Nalidixic acid compared to other serotypes (Figure 3.7C and Figure A3.6B; see also Figure 3.4B). In the light of microbiology, particular serotypes have different growth rates and virulence factors that are correlated with the rates of resistance emergence (Su et al. 2004, Kim et al. 2005, Beceiro et al. 2013), and this may explain why serotype is the best predictor of differences in *Salmonella* resistance.

Additionally, the best model contained an interaction between serotype and time (Table 3.3). The SW diversity for serotype Typhimurium increased over the surveillance period, whereas the SW diversity for Enteritidis did not change and for all the other serotypes diversity decreased (Figure 3.7A). When averaged across all serotypes, SW diversity decreased over the surveillance period (Figure 3.6A). There are two reasons why diversity would decrease over time: either species richness is decreasing (i.e., there are fewer types of resistance in the population) or the distribution of resistance types become less even. Natural selection is a possible explanation for why a population might become less even over time, as we would expect that resistance types with higher fitness would become more abundant over time, while resistance types with lower fitness would become less abundant. Selective pressure may be closely linked with antibiotic use and continued use of the same antibiotic may result in strong directional selection for resistance to that antibiotic.

However, it is not obvious that SW diversity will necessarily decrease over time because new types of resistance may emerge, increasing species richness, and increasing SW diversity. In the past, new types of resistance in *S. enterica* have indeed emerged (Glynn et al. 1998, Su et al. 2004). Whether SW diversity increases or decreases over time depends on the relative rate that new resistance types appear in the data set versus the rate that evenness is lost, potentially due to natural selection. The relative strengths of these opposing forces may explain why SW diversity increased for some serotypes while decreased for others: in particular, serotype level differences in mutation rates and generation times may explain this finding.

When measuring beta diversity for time periods, we found that consecutive years had low dissimilarity and were grouped together (Figure 3.6B) because they had similar levels of antibiotic resistance to the same antibiotics (Figure 3.6C; Figure 3.4C). This low dissimilarity means that the effectiveness of an antibiotic prescribing strategy is not likely to change dramatically year to year.

To illustrate the types of public health implications that our analyses could have, we now suppose that rather than *Salmonella enterica* that the results of our analyses pertain to a pathogen that is treated with antibiotics in humans, for example, *Staphylococcus aureus*. While we did not find differences in alpha diversity between provinces (Section 3.3.1; Figure 3.2A), if we had we might have concluded that microbiological testing prior to the prescribing of antibiotics is most critical in the province with the highest alpha diversity.

The results shown in the ordinations give information regarding which populations have similar frequencies of resistance to the same antibiotics (Figures 3.2C, 3.5C, and 3.6B) and resistance to which antibiotics is most different between populations (Figures 3.2D, 3.5D, and 3.6C). Figure 3.2C suggests that, if these results corresponded to a pathogen that is treated with antibiotics in humans, antibiotics that are successful in Alberta are also likely to be successful in British Columbia. Figure 3.2D suggests that in Alberta, Tetracycline, and Streptomycin are best to be avoided due to high levels of resistance to these antibiotics relative to other provinces (Figure 3.4A). Nova Scotia and New Brunswick have higher than average levels of relative resistance for Cefoxitin than other provinces (Figure 3.2D; see also Figure 3.4A), hence they should avoid prescribing Cefoxitin.

We found that the alpha diversity was different between serotypes. Often, at the time of prescribing the infecting serotype is unknown or if testing is done to identify serotype, testing for resistance may also be completed. As such, it is less practical to prioritize susceptibility testing for particular serotypes that have higher than average SW diversity or to recommend the use of particular antibiotics for one serotype relative to the others.

Previous studies have revealed different levels of antibiotic resistance in *Salmonella* isolates found in different environmental sources (such as water) or agricultural regions (Johnson et al. 2003, Wilkes et al. 2011, Jokinen et al. 2015), and between (Michel et al. 2006, Laupland et al. 2010, Nesbitt et al. 2012) or within provinces (between urban versus rural places; Michel et al. 2006). Predictor variables such as whether the isolates were recovered from rural versus urban environments may explain differences in alpha diversity, however, we were unable to test these variables as the CIPARS data set lacked such level of information.

3.4.4 Additional consideration for using biodiversity indices

In our study, the alpha diversity measure contains no information regarding the percentage of susceptible isolates in a population; nonetheless, this information should be taken into consideration when interpreting alpha diversity values. For example, if two populations have the same alpha diversity, then the population that has the lower percentage of susceptible isolates has the greater need for effective antibiotic prescribing. Isolates that

are susceptible to all antibiotic are not considered to be a resistance type because doing so would mean that higher values of alpha diversity do not necessarily correspond to populations that are more difficult to treat with antibiotics. For example, consider a population that is 50% susceptible to all antibiotics and 50% resistant to Ampicillin only (Population A) and another population with 50% resistance to Ampicillin only and 50% resistance to Ciprofloxacin only (Population B). Both Populations A and B would have the same alpha diversity values if isolates susceptible to all antibiotics are considered a resistance type, but in the absence of susceptibility testing it is more straightforward to treat Population A because any antibiotic except for Ampicillin is effective for every isolate in the population, whereas for Population B there are two antibiotics that must be excluded to ensure complete susceptibility for all isolates.

As there are numerous diversity indices, the best choice of diversity index may depend on the research question. Diversity indices focus on two components: species richness and species evenness (Magurran, 1998), and how much weight should be placed on either of these components translate into how much emphasis is put on rare versus common species. For our study, we chose the SW diversity index because it is easy to interpret and frequently used in other studies (Hill et al. 2003, Pakyz et al. 2008, de la Pedrosa et al. 2009, Sigala and Unc, 2013, Morris et al. 2014). Additionally, SW index puts more emphasis on the rare and intermediate resistant types (Hill et al. 2003). For antimicrobial resistance, indices that emphasize rare resistance types may be appealing since these rare resistance types could be a new type of multidrug resistance that may only respond to a very narrow range of antibiotics or to different antibiotics than that had previously been effective. Therefore, while there may only be a few isolates of a new resistance type, these new

resistance types are a significant challenge and so choosing an index that places more weight on rare species may be most appropriate.

The choice of alpha diversity index will affect the results of an analysis. If a different alpha diversity index were used, some of our results would be different (Figure A3.8 in Appendix). In ecology, the Berger-Parker index is calculated as the proportional abundance of the most abundant species (Table 2.2 in Chapter 2) and this alpha diversity index for an antimicrobial resistance study would heavily weigh the most common resistance type. This contrasts with the SW index, which gives more weighting to the rare and intermediate species relative to the most abundant species (Hill et al. 2003). Nova Scotia is dominated by the most abundant resistance type, while Alberta has the lowest abundance of most common resistance type (Figure 3.2B). If instead of the SW index the Berger-Parker index were used to measure diversity, Nova Scotia would have the highest diversity rather than the lowest, while Alberta would have the lowest rather than the highest diversity (Figure A3.8).

Similar to alpha diversity, different beta dissimilarity measures may be appropriate for particular application of the theory. As discussed in section 2.3 (Chapter 2), beta diversity indices have different properties and various data transformations yield different results with different emphasis. In this chapter, we used Bray-Curtis distance (equation 4 in section 2.3; Table 2.3) since it is the most frequently used abundance based dissimilarity measure (Anderson et al. 2011). Additionally, Bray-Curtis distance emphasizes the antibiotics that have the highest frequencies of resistance for the two populations (Krebs, 1999). For an antimicrobial resistance study, dissimilarity measures that emphasize

commonly resistant antibiotics may be appealing because two populations appearing close in the NMDS will have similar frequencies of resistance to the antibiotics that resistance most frequently occurs for. When prescribing antibiotics for such populations, they can have a similar policy of antibiotic use (i.e. avoid prescribing same antibiotics) because the antibiotics that will fail for one population will also likely fail for the other population, and this is especially true for the antibiotics that resistance most frequently occurs.

For our study, if we had used a different beta diversity measure some of our results would change (Figure A3.9 in Appendix; Table 2.3). In ecology, Bray-Curtis distance is a measure that is dominated by abundant species and rare species have very little effect, whereas the Canberra metric is less affected by the more abundant species (Krebs, 1999; Table 2.3). Prince Edward Island and Newfoundland and Labrador have a similar distribution of resistance to the most commonly resistant antibiotics, but a different distribution of resistance to the less commonly resistant antibiotics (Figure 3.4A). Hence, Prince Edward Island and Newfoundland and Labrador appear closer in the NMDS under Bray-Curtis dissimilarity and further apart under Canberra dissimilarity (Figure A3.9). However, if we had used the Canberra metric for our analysis, then we would not be able to interpret populations appearing close in the NMDS as populations that can have similar policies of antibiotic use, because the antibiotics that isolates are most frequently resistant to make the biggest contribution to failure in treatments, while the Canberra metric emphasizes antibiotics that have low frequencies of resistance.

In addition to considering whether the emphasis on rare or common species is more appropriate, different dissimilarity measures may exclude joint absences (see Table 2.3).

We performed the PCA using a proportional transformation on raw abundance data (see Section 3.4.2) such that the distance between populations in the PCA was related to the Euclidean distance (Section 2.3.1). We performed the NMDS based on Bray-Curtis distance (equation 4 in section 2.3; Table 2.3) and both Euclidean (on proportions) and Bray-Curtis dissimilarity ignore instances of joint absences. Excluding joint absences means that if two populations have no resistance to the same antibiotic, this does not contribute to the measurement of the similarity between the two populations. While the choice of Bray-Curtis dissimilarity for the NMDS facilitates comparison with the PCA as they both ignore joint absences, other dissimilarity measures that include joint absences (Table 2.3) could be used and may be more appropriate because they acknowledge that two populations with 100% susceptibility to the same antibiotics are more similar than two populations where the susceptibility to this antibiotics was unknown. However, for our case study excluding joint absences does not make a difference because there are no instances of joint absences in the data.

We used biodiversity measures to summarize antibiotic resistance across a wide number of antibiotics. While physicians may know the prevalence of resistance to some antibiotics for their regions, our method goes beyond this by summarizing resistance and by making comparisons between populations across all antibiotics. For the *S. enterica* analysis there are 17 antibiotics to consider, and due to the large number of antibiotics a summary measure is helpful. An alternative to our method is the Drug Resistance Index (DRI; Laxminarayan et al. 2011). To summarize resistance at the level of pathogens, the DRI utilizes data describing both the frequency of resistance and antibiotic use. However, as noted by Laxmimarayan et al. (2011) data on susceptibility and antibiotic use is difficult to

obtain as many parts of the world have limited laboratory space and also lack well organized surveillance systems. An advantage of our method is that it requires less data since antibiotic use data, which may be difficult to obtain, is not required. Furthermore, the DRI and our method produce quantities with different meanings. While strength of the DRI is that antibiotic use data is incorporated, strength of our approach is that only resistance data is considered. Under our method, when comparisons are made between regions, serotypes or time periods, we know that the differences are due to changes in resistance, while for the DRI it cannot be determined if differences are due to changes in resistance or antibiotic use.

One of the important limitations of species diversity indices is that not all species are equal, but they are treated such in the measurement of diversity (Magurran, 2004, Chiarucci et al. 2011). Species can vary functionally, evolutionarily and ecologically. When measuring beta diversity, each antibiotic is treated as being equal, but different antibiotics have different toxicities and antibiotics can be broad or narrow spectrum. This means a resistance type 'Na' (resistance to Nalidixic acid only) is treated as being just as different to a resistance type 'Am' (Ampicillin) as it is to a resistance type 'AmNa' (resistance to both Ampicillin and Nalidixic acid), but arguably 'Na' is much more similar to 'AmNa' than it is to 'Am'.

A challenge in measuring species diversity is that it is impossible to account for all species and to accurately measure their relative abundances with a limited number of samples (Magurran, 2004). Samples might not be representative of the true population as there are 'unseen' species, which are present in the community but are missing from the sample data. Thus, the estimated species richness will underestimate the true species richness with

finite sampling and species richness is one component of all species diversity measures. For *Salmonella*, some infections go unreported because not all healthcare institutions submit isolates to the CIPARS surveillance program. In addition, some patients with *Salmonella* will not present at clinics, however, there is no reason to expect that the distribution of resistance in *Salmonella* for patients that do not present at clinics is any different from the distribution of resistance in patients that do.

Diversity indices are sensitive to sample size and to reduce bias when comparing diversity between communities several authors suggest using samples of equal size (Hill et al. 2003, Magurran, 2004, Chao et al. 2014). For our analysis, we followed a standard sample size of 30 isolates, but following Chao et al.'s (2014) method of rarefaction and extrapolation approximately 100 individuals per sample would be an ideal sample size to make accurate comparisons. We choose to use 30 samples due to the size of the *Salmonella* data set to illustrate the method of data analysis, however, the results of rarefaction and extrapolation suggest that a larger number of samples may yield more robust conclusions (Figure A3.5).

In this chapter, we discuss the application of alpha and beta diversity measures to summarize antibiotic resistance data. We also illustrate the application by analyzing data describing antibiotic resistance in *S. enterica*. Measuring the alpha diversity of antibiotic resistance provides information on whether a treatment would likely succeed or fail in the absence of any susceptibility testing. Measuring the beta diversity of antibiotic resistance identifies populations that have similar distributions of antibiotic resistance and therefore may have similar approaches to antibiotic prescribing. Due to the challenge of interpreting

multidimensional dissimilarity matrices, ordination approaches visually present this information to efficiently recognize such populations.

3.5 Figures and Tables

Table 3.1: Example *Salmonella* antibiotic resistance data showing the categorization of each isolate and the calculation of p_i under the line-listed and the aggregated approach (see Section 3.2.1). In the antibiotics columns 1 indicates an isolate that is resistant and 0 indicates an isolate that is susceptible to the corresponding antibiotic. The antibiotic abbreviations are Amoxicillin-clavulanic acid (Am), Cephalothin (Ce), Ciprofloxacin (Cp), Ceftriaxone (Cr), Kanamycin (Ka), and Streptomycin (St) respectively. Resistance Type is the complete combination of antibiotics that an isolate is resistant to, for example, the abbreviation Am indicates resistance to Ampicillin only, while AmKaSt indicates resistance to Ampicillin, Kanamycin and Streptomycin and these are examples of two different resistance types.

Isolate No.	**Am**	**Ce**	**Cp**	**Cr**	**Ka**	**St**	**Resistance Type**	**Line-Listed p_i**
1	1	0	0	0	0	0	Am	1/6
2	0	0	0	0	0	1	St	1/6
3	1	0	0	0	1	1	AmKaSt	1/6
4	1	1	1	1	1	1	AmCeCpCrKaSt	1/6
5	0	0	0	0	0	1	St	1/6
6	0	0	0	0	1	1	KaSt	1/6
Aggregated p_i	3/14	1/14	1/14	1/14	3/14	5/14	-	-

Table 3.2: Percentage Distribution of the number of clinical resistant *S. enterica* isolates, collected by the Canadian Integrated Program for Antimicrobial Resistance Surveillance (CIPARS), into, Single and Multiple Resistance for different provinces in Canada (A) and for various serotypes (B) with species richness for line-listed approach (Resistance types). Number of 30 isolates groups (N) represents the sample size for each Province and each Serotype following the 30 isolate sampling strategy for measuring SW index for different serotypes in each province (refer to Table A3.1 in Appendix for raw distribution of clinical resistant isolates for serotypes in each province). Resistant isolates are those that are resistant to at least one antibiotic.

A) Provinces	**No. of Clinical Resistant *Salmonella* Isolates**	**No. of 30 Isolates groups (N)**	**Single Resistance (%)**	**Multiple Resistance (%)**	**Species Richness (Resistance Types)**
British Columbia	747	22	33.33	66.67	124
Alberta	867	26	32.64	67.36	133
Saskatchewan	283	7	44.88	55.12	57
Manitoba	442	13	36.88	63.12	80
Ontario	2382	77	32.75	67.25	193
Quebec	1351	43	31.09	68.91	153
New Brunswick	306	8	29.08	70.92	70
Prince Edward Island	36	0	47.22	52.78	17
Nova Scotia	224	6	53.12	46.88	49
Newfoundland & Labrador	79	1	43.04	56.96	25
B) Serotypes					
Enteritidis	1620	50	23.52	76.48	114
Heidelberg	2127	66	29.24	70.76	143
I 4,[5],12:i:-	550	14	28.73	71.27	78
Newport	195	4	20	80	61
Typhimurium	2245	69	9.93	90.07	210

Table 3.3: Model selection results for the Shannon-Weiner diversity index. Model Description lists the model variables, where SW stands for Shannon-Weiner index as the response variable, and P stands for Provinces, S stands for Serotypes, and D stands for Days as the predictor variables, and c is a constant term; LL is the log-likelihood of the models; K stands for the number of parameters in the models; delta AICc is the difference in the AICc of the model with the lowest AICc; wAICc is the model weights; Multiple R^2 shows the amount of variance explained by the predictor variables.

Model No.	Model Description	LL	K	AICc	delta AICc	wAICc	Multiple R^2
1	SW ~ P + S + D + S x D + c	-73.03	17	185.40	0	0.99	0.55
2	SW ~ P+S+D+P x D + S x D + c	-70.35	24	197.49	12.09	0	0.54
3	SW ~ P + S + P x S + c	-66.79	33	197.49	12.09	0	0.54
4	SW ~ P + S + D + c	-84.22	13	198.38	12.99	0	0.53
5	SW ~ P + S + c	-86.40	12	200.45	15.05	0	0.52
6	SW ~ S + D + S x D + c	-89.77	10	202.70	17.31	0	0.49
7	SW ~ P + S + D + P x D + c	-81.52	20	209.67	24.28	0	0.50
8	SW ~ S + D + c	-99.76	6	213.95	28.56	0	0.46
9	SW ~ S + c	-101.35	5	215.01	29.61	0	0.45
10	SW ~ P + S + D + P x S x D + c	-38.26	60	305.63	120.24	0	0.58
11	SW ~ P + S + D + P x S + c	-64.40	34	305.63	120.24	0	0.58
12	SW ~ P + D + c	-150.75	9	322.45	137.05	0	0.1
13	SW ~ D + c	-158.26	2	322.60	137.19	0	0.03
14	SW ~ c	-161.52	1	327.06	141.66	0	0
15	SW ~ P + c	-154.47	8	327.70	142.30	0	0.07
16	SW ~ P + D + P x D + c	-148.90	16	334.74	149.35	0	0.05

Table 3.4: Parameter estimates for the best model (i.e. Model 1 in Table 3.3) with 95% Confidence Interval (C.I.). (Model: SW ~ Province + Serotype + Days + Serotype x Days)

Parameter	Parameter Estimate	95% C.I.	
		2.5%	97.5%
Intercept	1.385	1.105	1.665
BC	-0.002	-0.211	0.207
MB	-0.066	-0.311	0.179
NB	0.008	-0.285	0.301
NS	-0.228	-0.555	0.100
ON	-0.359	-0.523	-0.196
QC	-0.210	-0.389	-0.031
SK	-0.133	-0.439	0.174
Heidelberg	0.701	0.401	1.002
I 4,[5],12:i:-	0.911	-0.035	1.858
Typhimurium	0.753	0.458	1.048
Newport	1.762	0.760	2.765
Days	-1.078×10^{-5}	-1.114×10^{-4}	8.983×10^{-5}
Heidelberg : Days	-1.610×10^{-4}	-2.865×10^{-4}	-3.525×10^{-5}
I 4,[5],12:i:- : Days	-1.255×10^{-4}	-4.148×10^{-4}	1.908×10^{-4}
Typhimurium : Days	9.165×10^{-5}	-3.735×10^{-5}	2.206×10^{-4}
Newport : Days	-1.560×10^{-4}	-4.901×10^{-4}	1.780×10^{-4}

Table 3.5: Comparison of the Line-Listed and the Aggregated Approach for measuring Alpha diversity. Scenario A – Multidrug Resistance shows that all the isolates (A1, A2, and A3) are resistant to multiple antibiotics. Scenario B – Single Resistance shows that all isolates (B1, B2, and B3) are resistant to only one antibiotic. SW diversity is the same for both scenarios under the aggregated approach, while for the line-listed approach the SW diversity is higher for Scenario B. For the aggregated approach, 'p_i' is the proportion of isolates resistant to a particular antibiotic *i* relative to the total frequency of resistance summed across all antibiotics; and for the line-listed approach, 'p_i' is the proportion of isolates belonging to a resistance type *i*.

Scenario A – Multidrug resistance

Line-Listed Approach

Isolates	AmCpKa
A1	1
A2	1
A3	1
Total	3
$\mathbf{p_i}$	3/3
SW = 0	

Aggregated Approach

Isolates	Am	Cp	Ka
A1	1	1	1
A2	1	1	1
A3	1	1	1
Total	3	3	3
$\mathbf{p_i}$	3/9	3/9	3/9
SW = 1.10			

Scenario B – Single resistance

Line-Listed Approach

Isolates	Am	Cp	Ka
B1	1	0	0
B2	0	1	0
B3	0	0	1
Total	1	1	1
$\mathbf{p_i}$	1/3	1/3	1/3
SW = 1.10			

Aggregated Approach

Isolates	Am	Cp	Ka
B1	1	0	0
B2	0	1	0
B3	0	0	1
Total	1	1	1
$\mathbf{p_i}$	1/3	1/3	1/3
SW = 1.10			

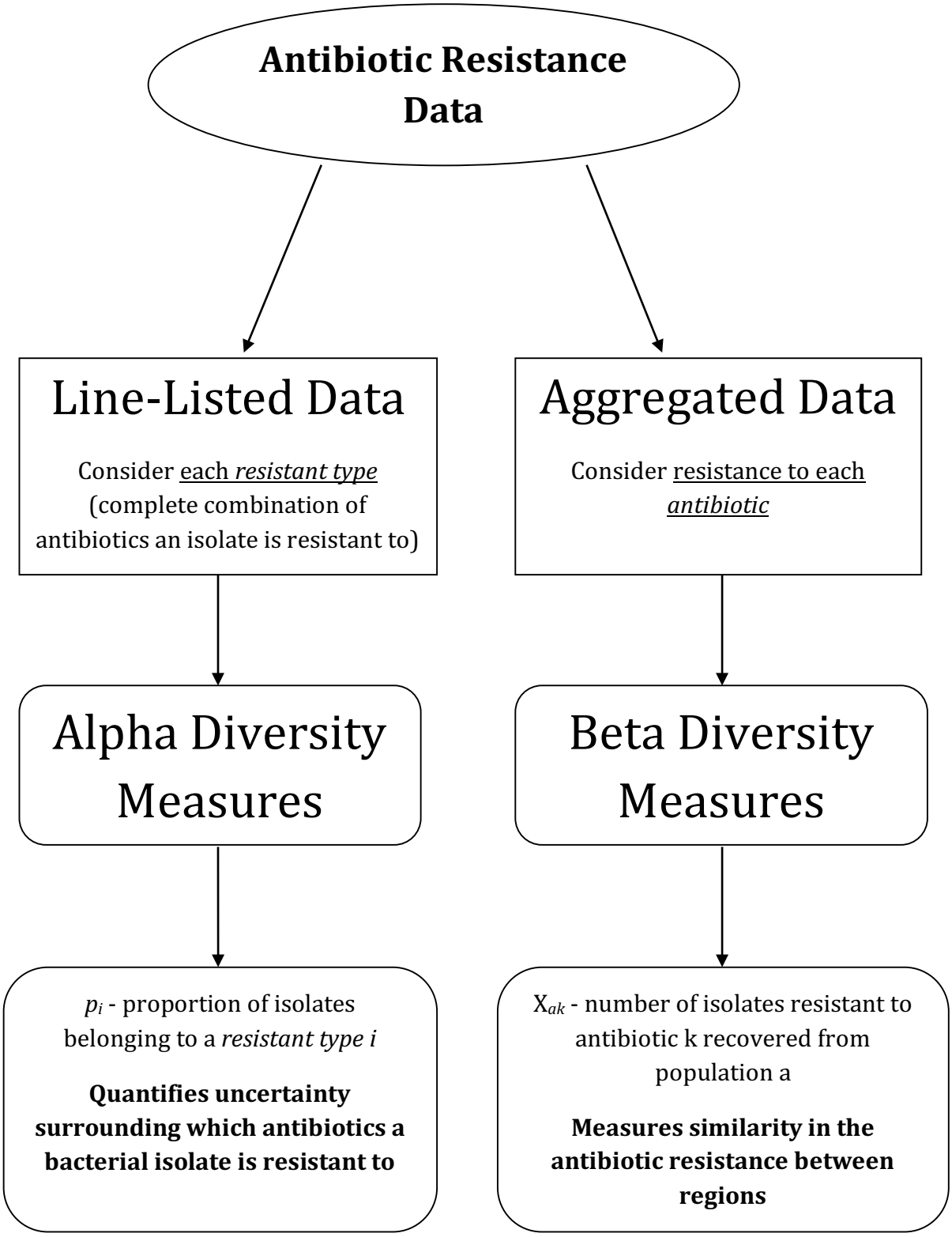


Figure 3.1: A flow chart for measuring alpha and beta diversity for antibiotic resistance data. p_i appears in equation 1 and 2 in Chapter 2; X_{ak} appears in equation 4 in Chapter 2**.**

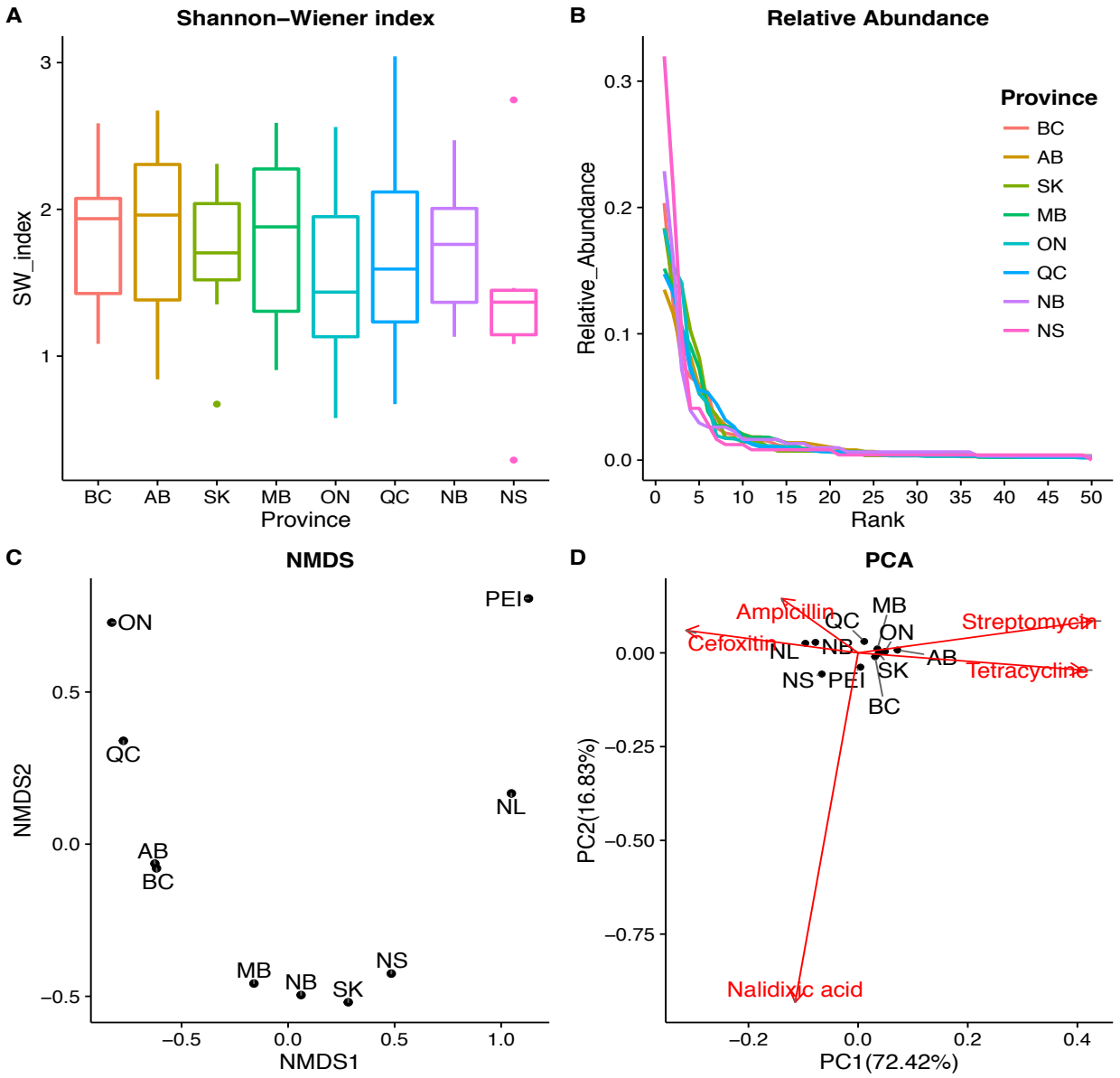


Figure 3.2: Alpha and Beta Diversity for Provinces. (A) Boxplot showing the SW index for provinces in Canada from west (left) to east (right). The solid lines denote the mean, the boxes represent the interquartile range, the whiskers are 1.5 times the interquartile range and the circles are outliers. The sample size (N; number of 30 isolate groups) for each province can be found in Table 3.2. (B) Relative abundance of the ranked resistance types in various provinces. (C) NMDS (Stress = 0.0008) using Bray-Curtis distance on raw antibiotic resistance data. Stress represents the mismatch between the rank order of the pairwise dissimilarities, between provinces, due to Bray-Curtis distance and in the NMDS ordination. (D) PCA for provinces performed on the variance-covariance matrix of the antibiotic resistance data transformed into proportions. PC1 explains 72.42% and PC2 explains 16.83% of the variance in the composition of antibiotic resistance with respect to the provinces. Vectors point in the direction of higher than average relative abundance of particular antibiotics.

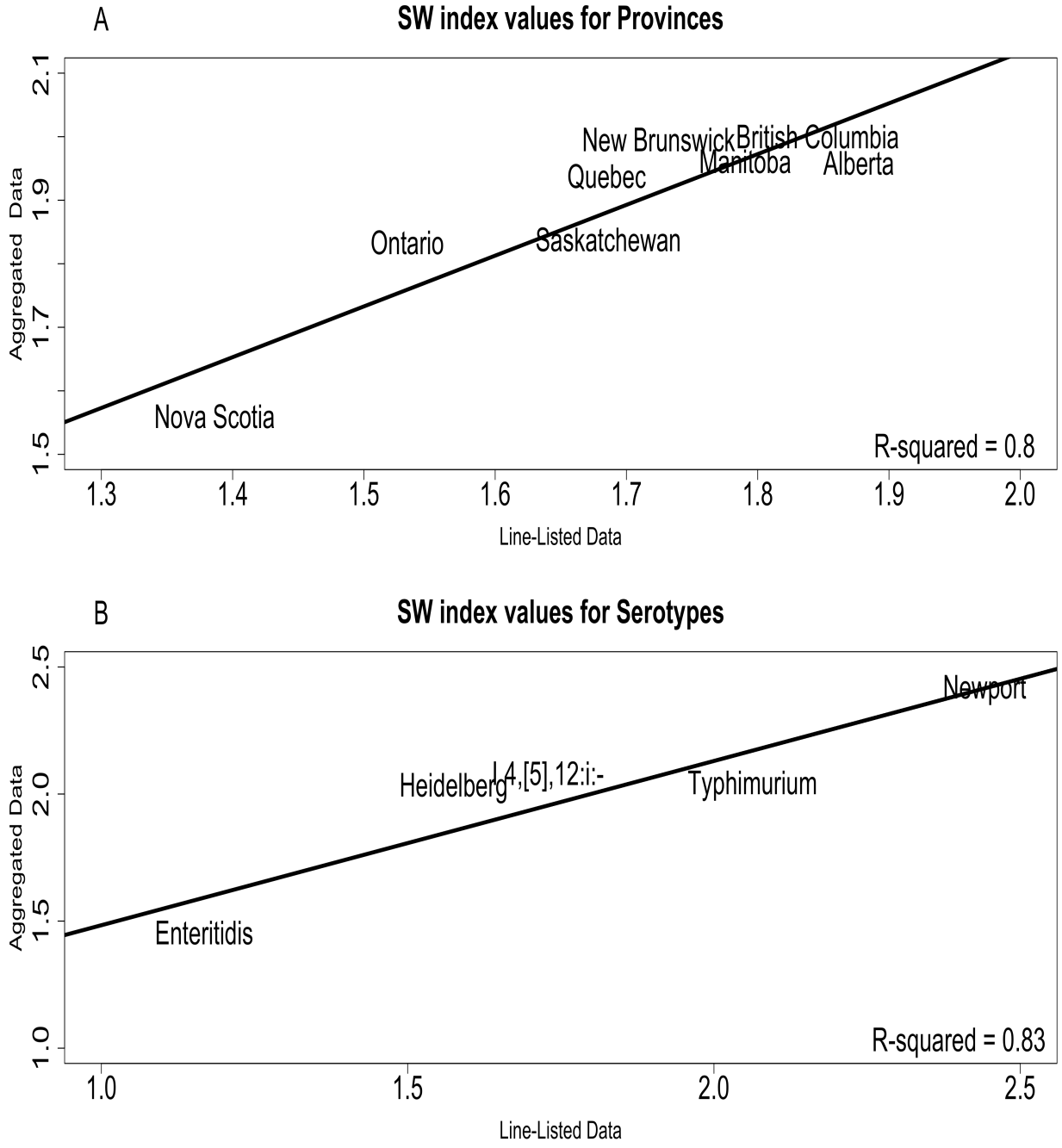


Figure 3.3: Correlation between Shannon-Weiner (SW) index values measured using line-listed and aggregated approaches. The line in each graph corresponds to the fitted regression line for SW index values between line-listed and aggregated approaches. The upper panel (A) shows the high correlation (r^2=0.8) for provinces in Canada. The lower panel (B) shows the high correlation (r^2=0.83) for *S. enterica* serotypes**.**

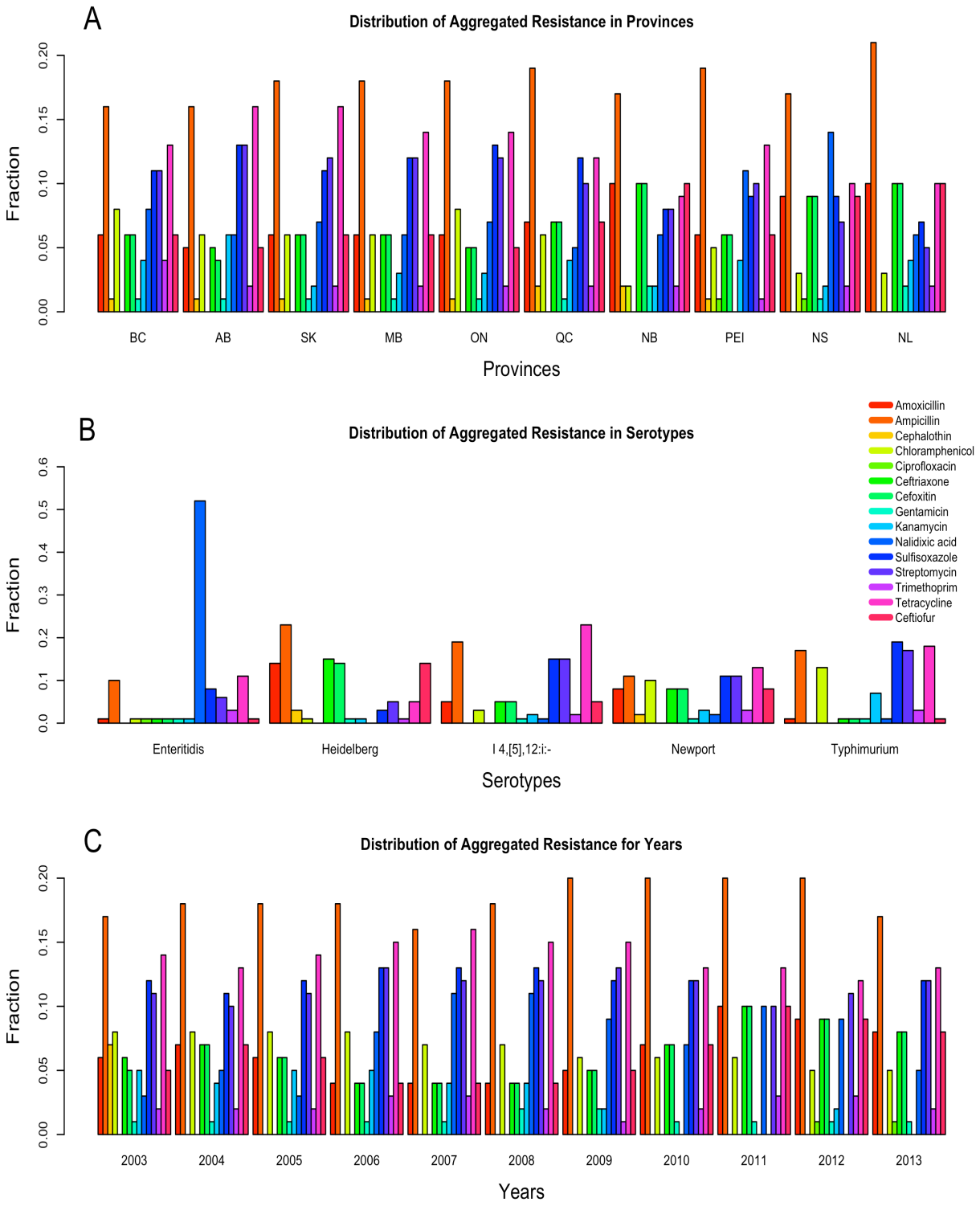


Figure 3.4: The distribution of aggregated resistance for provinces (province names are abbreviated; A), serotypes (B), and years (C). Note that in all the panels (A), (B), and (C) fraction of antibiotic resistance sum to greater than 1 for provinces, serotypes, and years because some isolates are resistant to multiple antibiotics.

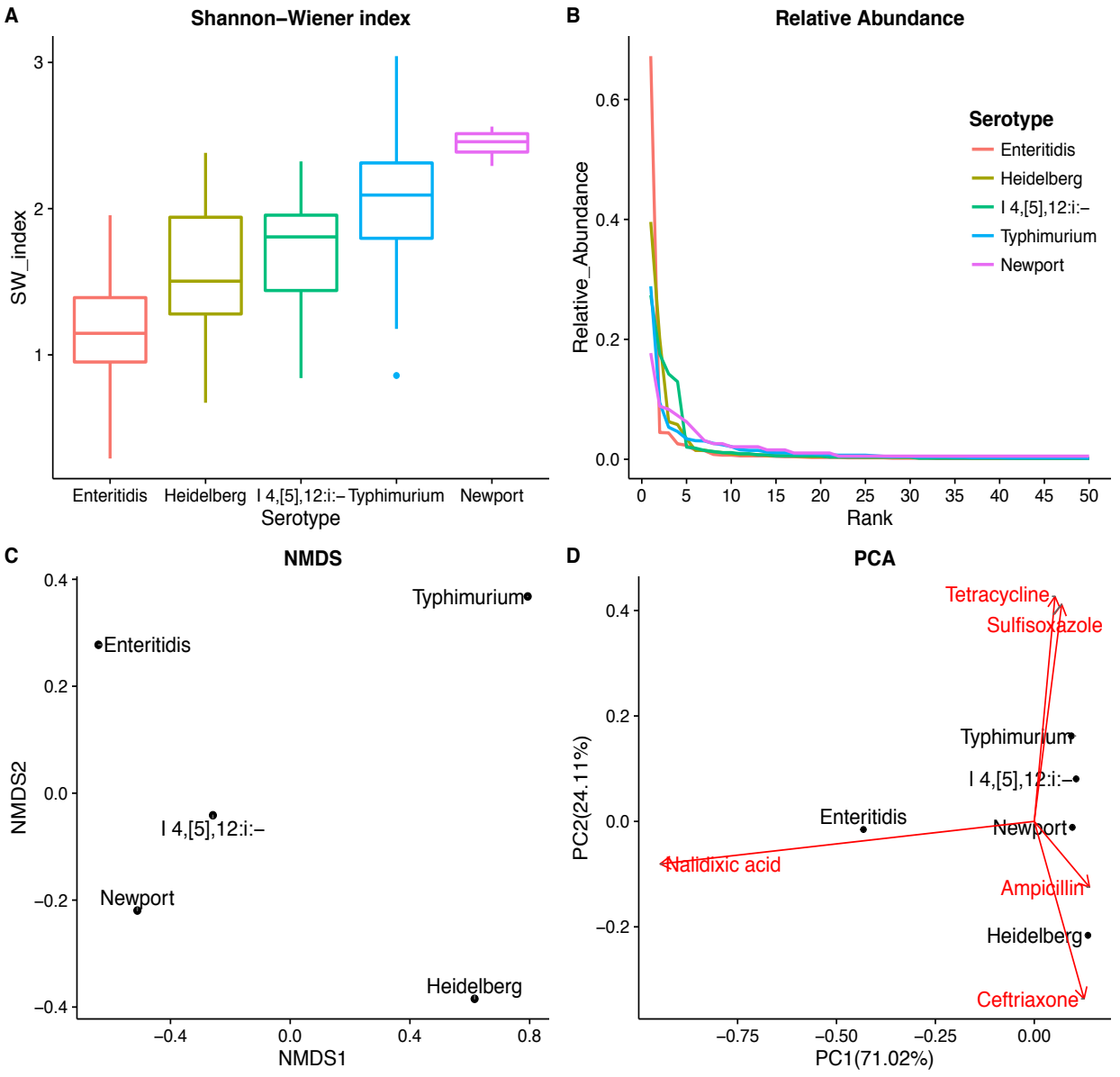


Figure 3.5: Alpha and Beta Diversity for Serotypes. (A) Boxplot showing the SW index for different *S. enterica* serotypes. The solid lines denote the mean, the boxes represent the interquartile range, the whiskers are 1.5 times the interquartile range and the circles are outliers. The sample size (N; number of 30 isolate groups) for each serotype can be found in Table 3.2. (B) Relative abundance of the ranked resistance types for serotypes. (C) NMDS (Stress = 0) using Bray-Curtis distance on raw antibiotic resistance data. Stress represents the mismatch between the rank order of the pairwise dissimilarities, between serotypes, due to Bray-Curtis distance and in the NMDS ordination. (D) PCA performed on the variance-covariance matrix of the antibiotic resistance data transformed into proportions. PC1 explains 71.02% and PC2 explains 24.11% of the variance in the composition of antibiotic resistance with respect to the serotypes. Vectors point in the direction of higher than average relative abundance of particular antibiotics.

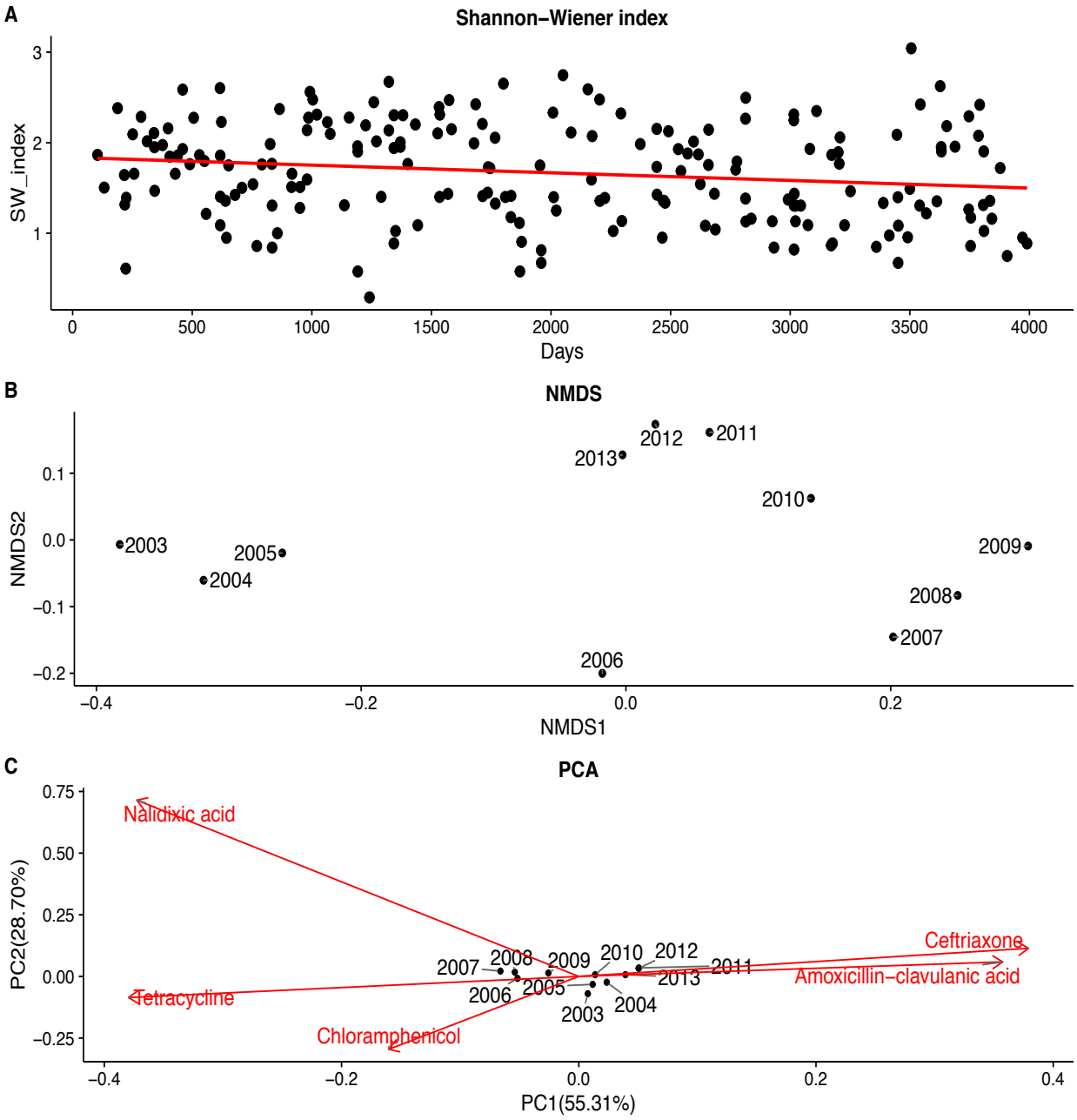


Figure 3.6: Alpha and Beta Diversity for time periods. (A) SW index change over the entire surveillance period (4016 days). The points denote SW diversity values at different time periods and the red line represents the fitted regression line. (B) NMDS for Years (2003-2013) using Bray-Curtis distance on raw antibiotic resistance data (stress = 0.01). Stress represents the mismatch between the rank order of the pairwise dissimilarities, between years, due to Bray-Curtis distance and in the NMDS ordination. (C) PCA for Years (2003-2013) performed on the variance-covariance matrix of antibiotic resistance data transformed into proportions. PC1 explains 55.31% and PC2 explains 28.07% of the variance in the composition of antibiotic resistance with respect to the years. Vectors point in the direction of higher than average relative abundance of particular antibiotics.

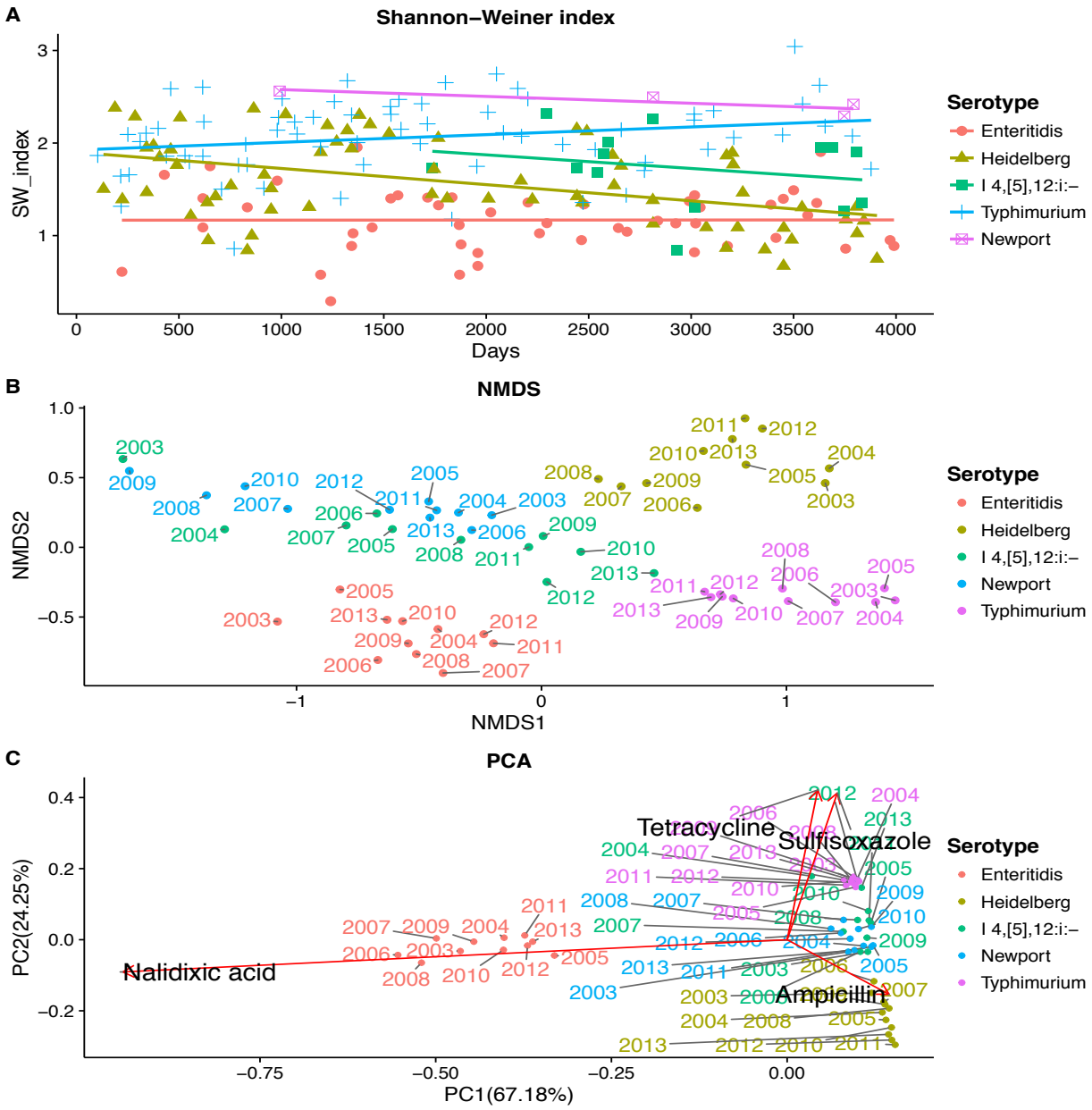


Figure 3.7: Alpha and Beta Diversity for interaction between Serotype and Time Periods. (A) SW index changes for each serotype over the entire surveillance period (4016 days). The points represent SW index values for different time periods grouped by serotype. The lines represent the fitted regressions for each serotype. (B) NMDS using Bray-Curtis distance on raw antibiotic resistance data for Serotype-Year combinations (stress = 0.06). Stress represents the mismatch between the rank order of the pairwise dissimilarities due to Bray-Curtis distance and in the NMDS ordination. (C) PCA performed on the variance-covariance matrix of antibiotic resistance data for Serotype-Year combination transformed into proportions. PC1 explains 67.18% and PC2 explains 24.25% of the variance in the composition of antibiotic resistance with respect to the serotype-year combinations. Vectors point in the direction of higher than average relative abundance of particular antibiotics.

3.6 References

Abay, S., Kayman, T., Otlu, B., Hizlisoy, H., Aydin, F. & Ertas, N. (2014). Genetic diversity and antibiotic resistance profiles of Campylobacter jejuni isolates from poultry and humans in Turkey. International Journal of Food Microbiology, 178(16), 29-38

Anderson, D.R. (2008). *Model based inference in the life sciences: A primer on evidence*. Springer, New York, New York.

Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L., et al. (2011). Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist: Roadmap for beta diversity. Ecology Letters, 14(1), 19–28. http://doi.org/10.1111/j.1461-0248.2010.01552.x

Barwell, L., Isaac, N., & Kunin, W. (2015). Measuring β - diversity with species abundance data. *Journal of Animal Ecology, 84*(4), 1112-1122.

Beceiro, A., Tomás, M., & Bou, G. (2013). Antimicrobial Resistance and Virulence: a Successful or Deleterious Association in the Bacterial World? *Clinical Microbiology Reviews*, *26*(2), 185–230. http://doi.org/10.1128/CMR.00059-12

Beck, J., Holloway, J.D. & Schwanghart, W. (2013) Undersampling and the measurement of beta diversity (ed D Orme). *Methods in Ecology and Evolution*, 4, 370–382.

Bennett, J. R., and Gilbert, B. (2016). Contrasting beta diversity among regions: how do classical and multivariate approaches compare?. *Global Ecology and Biogeography*.

Blaak, H., Van Hoek, A., Hamidjaja, R., Van Der Plaats, R., Kerkhof-de Heer, L., De Roda Husman, A., & Schets, F. (2015). Distribution, Numbers, and Diversity of ESBL-Producing E. coli in the Poultry Farm Environment. *PloS One, 10*(8), E0135402.

Bolker, B. (2008). Ecological models and data in R. New Jersey, USA: Princeton University Press.

Bondo, K. J., Pearl, D. L., Janecko, N., Boerlin, P., Reid-Smith, R. J., Parmley, J. and Jardine, C. M. (2015), Epidemiology of *Salmonella* on the Paws and in the Faeces of Free-Ranging Raccoons (*Procyon Lotor*) in Southern Ontario, Canada. Zoonoses and Public Health. doi: 10.1111/zph.12232

Caucci, S., Karkman, A., Cacace, D., Rybicki, M., Timpel, P., Voolaid, V., et al. (2016). Seasonality of antibiotic prescriptions for outpatients and resistance genes in sewers and wastewater treatment plant outflow. *FEMS Microbiology Ecology, 92*(5). doi:10.1093/femsec/fiw060

Chao, A., Chazdon, R. L., Colwell, R. K., and Shen, T.-J. (2005). A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecology Letters 8, 148–159.

Chao, A., Gotelli, N., Hsieh, T., Sander, E., Ma, K., Colwell, R., and Ellison, A. (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs, 84*(1), 45-67.

Chen, H., Wang, Y., Su, L., & Chiu, C. (2013). Nontyphoid salmonella infection: Microbiology, clinical features, and antimicrobial therapy. *Pediatrics and Neonatology, 54*(3), 147-52.

Chiarucci, A., Bacaro, G., & Scheiner, S. (2011). Old and new challenges in using species diversity for assessing biodiversity. *Philosophical Transactions of the Royal Society B, 366*(1576), 2426-2437.

Ciccolini, M., Spoorenberg, V., Geerlings, S., Prins, J., & Grundmann, H. (2015). Using an index-based approach to assess the population-level appropriateness of empirical antibiotic therapy. *Journal of Antimicrobial Chemotherapy, 70*(1), 286-293.

De la Pedrosa, E., Baquero, F., Loza, E., Nadal-Serrano, J., Fenoll, A., Del Campo, R., & Cantón, R. (2009). High clonal diversity in erythromycin-resistant Streptococcus pneumoniae invasive isolates in Madrid, Spain (2000–07). *Journal of Antimicrobial Chemotherapy, 64*(6), 1165-1169.

Di Giulio, M., Di Campli, E., Di Bartolomeo, S., Cataldi, V., Marzio, L., Grossi, L., et al. (2016). In vitro antimicrobial susceptibility of Helicobacter pylori to nine antibiotics currently used in Central Italy. *Scandinavian Journal of Gastroenterology, 51*(3), 263-9.

Dyar, OJ, Hoa, NQ, Trung, NV, Phuc, HD, Larsson, M, Chuc, NTK, & Lundborg, CS. (2012). High prevalence of antibiotic resistance in commensal Escherichia coli among children in rural Vietnam. *Bmc Infectious Diseases, 12*, 92.

Eguale, T., Engidawork, E., Gebreyes, W., Asrat, D., Alemayehu, H., Medhin, G., et al. (2016). Fecal prevalence, serotype distribution and antimicrobial resistance of Salmonellae in dairy cattle in central Ethiopia. *BMC Microbiology, 16*, 20.

Glynn, M., Bopp, C., Dewitt, W., Dabney, P., Mokhtar, M., & Angulo, F. (1998). Emergence of Multidrug-Resistant Salmonella enterica SerotypeTyphimurium DT104 Infections in the United States. *The New England Journal of Medicine, 338*(19), 1333-1339.

Government of Canada. Canadian Integrated Program for Antimicrobial Resistance Surveillance (CIPARS). (2013). Annual Report – Chapter 1,2. Antimicrobial Resistance. Public Health Agency of Canada, Guelph, Ontario, 2014.

Helms M, Simonsen J, Molbak K. Quinolone resistance is associated with increased risk of invasive illness or death during infection with Salmonella serotype Typhimurium. J Infect Dis. 2004;190(9):1652–4.

Hill, T., Walsh, K., Harris, J., and Moffett, B. (2003). Using ecological diversity measures with bacterial communities. *FEMS Microbiology Ecology, 43*(1), 1-11.

Holm, A. M., Cordoba, G. R., Siersma, V., Bjerrum, L., Sørensen, T., & Jessen, L. (2015). Point of care susceptibility testing in primary care - Does it lead to a more appropriate prescription of antibiotics in patients with uncomplicated urinary tract infections? Protocol for a randomized controlled trial. *BMC Family Practice, 16*(1), BMC Family Practice, 21 August 2015, Vol.16(1).

Hong, P., Lee, B., Aw, M., Shek, L., Yap, G., Chua, K., et al. (2010). Comparative Analysis of Fecal Microbiota in Infants with and without Eczema (Fecal Microbiota of Infants). *PLoS ONE, 5*(4), E9964.

Johnson, J., Thomas, J., Graham, T., Townshend, I., Byrne, J., Selinger, L., & Gannon, V. (2003). Prevalence of Escherichia coli O157:H7 and Salmonella spp. in surface waters of southern Alberta and its relation to manure sources. *Canadian Journal of Microbiology,49*(5), 326-335.

Jokinen, C. C., Koot, J., Cole, L., Desruisseau, A., Edge, T. A., Khan, I. U. H., et al. (2015). The distribution of Salmonella enterica serovars and subtypes in surface water from five agricultural regions across Canada. *Water Research*, *76*, 120–131. http://doi.org/10.1016/j.watres.2015.02.038

Kim, W., Karabasil, N., Bulajic, S., Dunkley, K., Callaway, T., Poole, T., et al. (2005). Comparison of Spontaneous Antibiotic Resistance Frequency of Salmonella Typhimurium Growth in Glucose Amended Continuous Culture at Slow and Fast Dilution Rates. *Journal of Environmental Science and Health, Part B, 40*(3), 475-484.

Krebs, C. J. (1999). Ecological methodology (2nd ed). Menlo Park, Calif:Benjamin/Cummings.

Laupland, K., Schonheyder, H., Kennedy, K., Lyytikainen, O., Valiquette, L., Galbraith, J., & Collignon, P. (2010). Salmonella enterica bacteraemia: A multi-national population-based cohort study. *Bmc Infectious Diseases, 10*, Bmc Infectious Diseases, 2010 Apr 14, Vol.10.

MacLean, R. C., Hall, A. R., Perron, G. G., & Buckling, A. (2010). The population

genetics of antibiotic resistance: integrating molecular mechanisms and treatment contexts. *Nature Reviews Genetics*, *11*(6), 405–414. doi:10.1038/nrg2778

Magiorakos, A. P., Srinivasan, A., Carey, R.B., Carmeli, Y., Falagas, M.E., Giske, C.G., et al. (2012). Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: An international expert proposal for interim standard definitions for acquired resistance. *Clinical Microbiology and Infection, 18*(3), 268-281.

Magurran, A.E. (2004). Measuring Biological Diversity. Blackwell, Oxford.

Martin LJ, Fyfe M, Doré K, Buxton JA, Pollari F, Henry B, Middleton D, Ahmed R, Jamieson F, Ciebin B, et al. Increased burden of illness associated with antimicrobial-resistant Salmonella enterica serotype typhimurium infections. J Infect Dis. 2004;189(3):377–84.

Mazerolle, M.J. (2016) AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). R package version 2.0-4. http://CRAN.R-project.org/ package=AICcmodavg.

Mead, P., Slutsker, L., Dietz, V., McCaig, L., Bresee, J., Shapiro, C., et al. (1999). Food-related illness and death in the United States. *Emerging Infectious Diseases, 5*(5), 607-625.

Michel, P., Martin, L., Tinga, C., & Doré, K. (2006). Regional, Seasonal, and Antimicrobial Resistance Distributions of Salmonella Typhimurium in Canada: A Multi-Provincial Study.*Canadian Journal of Public Health / Revue Canadienne De Sante'e Publique, 97*(6), 470-474.

Morris, E., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T., et al. (2014). Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. *Ecology and Evolution, 4*(18), 3514-3524.

Mutlu, Ece A., Keshavarzian, Ali, Losurdo, John, Swanson, Garth, Siewe, Basile, Forsyth, Christopher, et al. (2014). A compositional look at the human gastrointestinal microbiome and immune activation parameters in HIV infected subjects. *PLoS Pathogens, 10*(2), PLoS Pathogens, Feb, 2014, Vol.10(2).

Nesbitt, A., Ravel, A., Murray, R., McCORMICK, R., Savelli, C., Finley, R., et al. (2012). Integrated surveillance and potential sources of Salmonella Enteritidis in human cases in Canada from 2003 to 2009. *Epidemiology and Infection*, *140*(10), 1757–1772. http://doi.org/10.1017/S0950268811002548

Nys, Okeke, Kariuki, Dinant, Driessen, & Stobberingh. (2004). Antibiotic resistance of faecal Escherichia coli from healthy volunteers from eight developing countries. *Journal of Antimicrobial Chemotherapy,54*(5), 952-955.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara , R. B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H. , and Wagner, H. (2016). vegan: Community Ecology Package. R package version 2.3-5. http://CRAN.R-project.org/package=veganGGplot

Pakyz, A., Powell, J., Harpe, S., Johnson, C., Edmond, M., & Polk, R. (2008). Diversity of antimicrobial use and resistance in 42 hospitals in the United States. *Pharmacotherapy, 28*(7), 906-12.

Pielou, E. C. (1975). Ecological diversity. New York : Wiley InterScience

Price, L., Liu, C., Melendez, J., Frankel, Y., Engelthaler, D., Aziz, M., et al. (2009). Community Analysis of Chronic Wound Bacteria Using 16S rRNA Gene-Based Pyrosequencing: Impact of Diabetes and Antibiotics on Chronic Wound Microbiota (Chronic Wound Microbiota). *PLoS ONE, 4*(7), E6462.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Shankar, V., Agans, R., Holmes, B., Raymer, M., & Paliy, O. (2013). Do gut microbial communities differ in pediatric IBS and health? *Gut Microbes, 4*(4), 347-52.

Sigala, J., & Unc, A. (2013). Pyrosequencing estimates of the diversity of antibiotic resistant bacteria in a wastewater system. *Water Science and Technology: A Journal of the International Association on Water Pollution Research, 67*(7), 1534-43.

Slowikowski, K. (2016). ggrepel: Repulsive Text and Label Geoms for 'ggplot2'. R package version 0.5. http://CRAN.R-project.org/package=ggrepel

Soetaert, K., & Heip, C. (1990). Sample-size dependence of diversity indices and the determination of sufficient sample size in a high-diversity deep-sea environment. *Marine Ecology Progress Series, 59*, 305-307.

Soetaert, K. (2016). plot3D: Plotting Multi-Dimensional Data. R package version 1.1. http://CRAN.R-project.org/package=plot3D

Su, L., Chiu, C., Chu, C., & Ou, J. (2004). Antimicrobial Resistance in Nontyphoid Serotypes: A Global Challenge. *Clinical Infectious Diseases, 39*(4), 546-551.

Talebi Bezmin Abadi, A., Mobarez, A., Taghvaei, T., & Wolfram, L. (2010). Antibiotic Resistance of Helicobacter pylori in Mazandaran, North of Iran.*Helicobacter, 15*(6), 505-509.

Tanwar, J., Das, S., Fatima, Z., & Hameed, S. (2014). Multidrug resistance: An emerging

crisis. *Interdisciplinary Perspectives on Infectious Diseases, 2014*, 541340.

Thomas, J. L., Slawson, R. M., & Taylor, W. D. (2013). *Salmonella* serotype diversity and seasonality in urban and rural streams. *Journal of Applied Microbiology*, *114*(3), 907–922. http://doi.org/10.1111/jam.12079

Laxminarayan, R., & Klugman, K. (2011). Communicating trends in resistance using a drug resistance index. *BMJ Open, 1*(2), BMJ Open 2011, Vol.1(2).

Van Den Bruel, A., Aertgeerts, B., Bruyninckx, R., Aerts, M., & Buntinx, F. (2007). Signs and symptoms for diagnosis of serious infections in children: A prospective study in primary care. *The British Journal of General Practice : The Journal of the Royal College of General Practitioners, 57*(540), 538-46.

Varma JK, Molbak K, Barrett TJ, Beebe JL, Jones TF, Rabatsky-Ehr T, Smith KE, Vugia DJ, Chang HG, Angulo FJ. Antimicrobial-resistant nontyphoidal Salmonella is associated with excess bloodstream infections and hospitalizations. J Infect Dis. 2005;191(4):554–61.

Warnes, G.R., Bolker, B., Lumley, T., Johnson, R.C. Contributions from Randall C. Johnson are Copyright SAIC-Frederick, Inc. Funded by the Intramural Research Program, of the NIH, National Cancer Institute and Center for Cancer Research under NCI Contract NO1-CO-12400. (2015). gmodels: Various R Programming Tools for Model Fitting. R package version 2.16.2. http://CRAN.R-project.org/package=gmodels

Wilke, C.O. (2016). cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 0.6.1.9999. https://github.com/wilkelab/cowplot

Wilkes, G., Edge, T. A., Gannon, V. P. J., Jokinen, C., Lyautey, E., Neumann, N. F., et al. (2011). Associations among pathogenic bacteria, parasites, and environmental and land use factors in multiple mixed-use watersheds. *Water Research*, *45*(18), 5807–5825. http://doi.org/10.1016/j.watres.2011.06.021

Wright JG, Tengelsen LA, Smith KE, Bender JB, Frank RK, Grendon JH, Rice DH, Thiessen AM, Gilbertson CJ, Sivapalasingam S, et al. Multidrug-resistant Salmonella Typhimurium in four animal facilities. Emerg Infect Dis. 2005;11(8):1235–41.

Zhang, Bo, Zhang, Zi-wei, Meng, Fan-Yu, Wu, Qiong, Xu, Shi-Wen, & Wang, Xiao-Long. (2014). Analysis of the structure of bacteria communities and detection of resistance genes of quinolones from pharmaceutical wastewater. *Annals of Microbiology, 64*(1), 23-29.

Chapter 4

Thesis Summary

Contributions to theory:

i) We propose a new method of applying biodiversity theory to summarize the antibiotic resistance data. In particular, we provide direction for calculating p_i (as it appears in equation 1 and 2 in Chapter 2) for measuring alpha diversity indices when applied to antibiotic resistance data. We recommend using a line-listed approach for measuring alpha diversity and defining p_i as the proportion of isolates belonging to a resistance type *i*. For measuring beta diversity, we recommend using aggregated approach and defining X_{ak} or X_{bk} (as it appears in equation 4 in Chapter 2) as the number of isolates that are resistant to the antibiotic k recovered from population a or population b, respectively.

ii) Our work is a novel contribution to public health in understanding and summarizing the antibiotic resistance across broad range of antibiotics.

Contributions to practice:

i) Calculating alpha diversity using the line-listed approach will provide information on the diversity of resistance types. For highly diverse populations, in the absence of any microbiological testing, the uncertainty surrounding the combination of antibiotics that an isolate is resistant to is highest. This implies that the treatment for such populations is more likely to fail, in the absence of any other information, relative to less diverse populations.

ii) Beta Diversity using the aggregated approach will provide information on the antibiotics that are likely to succeed or fail in different regions, when infections are treated without susceptibility tests.

Our method summarizes overall antibiotic resistance for a population, which provides direction for prescribing antibiotics when the antibiotic resistance of the infecting pathogen is unknown. This will also likely improve the treatment at the population level and help reduce the failure rate of treatments for the infections that are treated without prior susceptibility tests.

To apply our method you need:

i) For Alpha Diversity:

   a. A large line-listed antibiotic resistance dataset for pathogens that are prescribed antibiotic treatment.

   b. Alpha diversity metrics.

ii) For Beta Diversity:

a. Aggregated antibiotic resistance dataset.

b. Beta diversity metrics and software for ordinations (such as the vegan package in R, basic R package, etc.)

Limitations and Future Directions:

There are many limitations of our analysis in Chapter 3, however, diversity measures that reduce high dimensional information into a few dimension will always need to rely on strong assumptions. When applied to ecology, every diversity index makes an assumption that all species are equal despite some species being more similar to each other. There are some approaches in ecology that attempt to overcome this limitation, for example taxonomic diversity indices (Desrochers and Anand, 2004) and the approach by Leinster et al. (2012). The same limitation applies to antimicrobial resistance as some resistance types or antibiotics are more similar than others. In this sense, the limitations of applying biodiversity measures to antimicrobial resistance are similar to the limitations associated with its use in ecology. One of the limitations of our analysis is due to the limited size of our data set, which might have caused underestimation of the total number of resistance types due to inadequate sampling.

S. enterica infections in humans are typically not prescribed antibiotic treatment. We anticipate that when applied to pathogens that are treated with antibiotics, our method will provide valuable information on whether the antibiotic therapy will likely succeed or fail based on the choice of antibiotic and the type of resistance expressed by the infecting isolate. Future research needs to investigate the feasibility of using the aggregated

approach to measure and interpret the results of alpha diversity in different scenarios. A future direction would be trying to create models to recover the line-listed data from the aggregated data. Such models would help in measuring alpha diversity when only aggregated data is available. Future studies should also test the application of other biodiversity measures on antibiotic resistance study and their meaningful interpretations.

The need for indices summarizing antibiotic resistance across antibiotics, pathogen species, and disease syndromes has been supported by some recent studies (Laxminarayan and Klugman, 2011, Ciccolini et al. 2015). Our proposed methods for measuring alpha and beta diversity indices (Figure 3.1) are useful tools for antimicrobial resistance studies as these methods can be used to measure and understand antibiotic resistance across a broad range of antibiotics.

References

Ciccolini, M., Spoorenberg, V., Geerlings, S., Prins, J., & Grundmann, H. (2015). Using an index-based approach to assess the population-level appropriateness of empirical antibiotic therapy. *Journal of Antimicrobial Chemotherapy, 70*(1), 286-293.

Di Giulio, M., Di Campli, E., Di Bartolomeo, S., Cataldi, V., Marzio, L., Grossi, L., et al. (2016). In vitro antimicrobial susceptibility of Helicobacter pylori to nine antibiotics currently used in Central Italy. *Scandinavian Journal of Gastroenterology, 51*(3), 263-9.

Dyar, OJ, Hoa, NQ, Trung, NV, Phuc, HD, Larsson, M, Chuc, NTK, & Lundborg, CS. (2012). High prevalence of antibiotic resistance in commensal Escherichia coli among children in rural Vietnam. *Bmc Infectious Diseases, 12*, 92.

Laxminarayan, R., & Klugman, K. (2011). Communicating trends in resistance using a drug resistance index. *BMJ Open, 1*(2), BMJ Open 2011, Vol.1(2).

Magurran, A. E. (1988). Ecological diversity and its measurement. Princeton, N.J: Princeton University Press.

Nys, Okeke, Kariuki, Dinant, Driessen, & Stobberingh. (2004). Antibiotic resistance of faecal Escherichia coli from healthy volunteers from eight developing countries. *Journal of Antimicrobial Chemotherapy,54*(5), 952-955.

Talebi Bezmin Abadi, A., Mobarez, A., Taghvaei, T., & Wolfram, L. (2010). Antibiotic Resistance of Helicobacter pylori in Mazandaran, North of Iran.*Helicobacter, 15*(6), 505-509.

Western Pacific Region (WPR) - World Health Organization (WHO). 2016. WHO supports nationwide campaign to combat antimicrobial resistance. Press Release. Web. 2016. Retrieved from http://www.wpro.who.int/vietnam/mediacentre/releases/2015/antibiotics_resistance_week2015/en/

Whittaker, R. H. (1972). Evolution and Measurement of Species Diversity. *Taxon*, *21*(2/3), 213. http://doi.org/10.2307/1218190

Appendix

Table A3.1: Distribution of clinical antibiotic resistant *S. enterica* isolates (before the 30 isolate sampling for different serotypes in each province) between provinces and serotypes. The rows represent provinces (abbreviations) and the column represent different *S. enterica* serotypes. As seen in the table, PEI has less than 30 isolates for each serotype and NL has only one serotype that has more than 30 isolates. The total number of resistant isolate is 6737. Resistant isolates are those that are resistant to at least one antibiotic.

	Enteritidis	Heidelberg	I 4, [5], 12:i:-	Newport	Typhimurium	**Row Total**
BC	234	179	65	32	237	747
AB	190	221	108	29	319	867
SK	66	77	53	6	81	283
MB	90	154	51	9	138	442
ON	583	668	136	81	914	2382
QC	244	530	109	33	435	1351
NB	77	159	11	2	57	306
PEI	15	10	4	2	5	36
NS	106	86	8	0	44	224
NL	15	43	5	1	15	79
Column Total	1620	2127	550	195	2245	**6737**

Table A3.2: Testing association between the predictor variables province and serotype using the Contingency Table. The table shows sample sizes, i.e. the number of SW index values (response variable), for different serotypes in each province following the 30 isolates sampling. The rows represent provinces (abbreviations) and the columns represent serotypes. Fisher's exact test (with simulated p-values based on 2000 replication) was performed because expected values for different serotypes in each province was <5 in most cases. The test shows that province and serotypes do not have any association at the significance level of $\alpha=0.05$ (p-value =0.972; two-sided test). Hence, the categorical predictor variables province and serotype are independent.

	Enteritidis	Heidelberg	I 4,[5],12:i:-	Typhimurium	Newport	**Row Total**
AB	6	7	3	10	0	26
BC	7	5	2	7	1	22
MB	3	5	1	4	0	13
NB	2	5	0	1	0	8
NS	3	2	0	1	0	6
ON	19	22	4	30	2	77
QC	8	17	3	14	1	43
SK	2	2	1	2	0	7
Column Total	50	65	14	69	4	**202**

Fisher's Exact Test for Count Data:

p-value = 0.971

Table A3.3: Mean SW diversity index for Provinces in Canada (with sample size N) with 95% Confidence Interval (C.I.) around the mean.

Provinces	Sample Size (N)	Mean SW index	95% C.I.	
			2.5%	97.5%
British Columbia	22	1.84	1.65	2.03
Alberta	26	1.88	1.66	2.09
Saskatchewan	7	1.68	1.16	2.21
Manitoba	13	1.79	1.43	2.15
Ontario	77	1.53	1.42	1.65
Quebec	43	1.69	1.52	1.86
New Brunswick	8	1.71	1.29	2.12
Nova Scotia	6	1.38	0.55	2.22

Table A3.4: Mean SW diversity index for *S. enterica* Serotypes (with sample size N) with 95% Confidence Interval (C.I.) around the mean.

Serotypes	Sample Size (N)	Mean SW index	95% C.I.	
			2.5%	97.5%
Enteritidis	50	1.17	1.07	1.26
Heidelberg	66	1.57	1.46	1.69
I 4,[5],12:i:-	14	1.73	1.49	1.96
Typhimurium	4	2.06	1.97	2.16
Newport	69	2.44	2.26	2.63

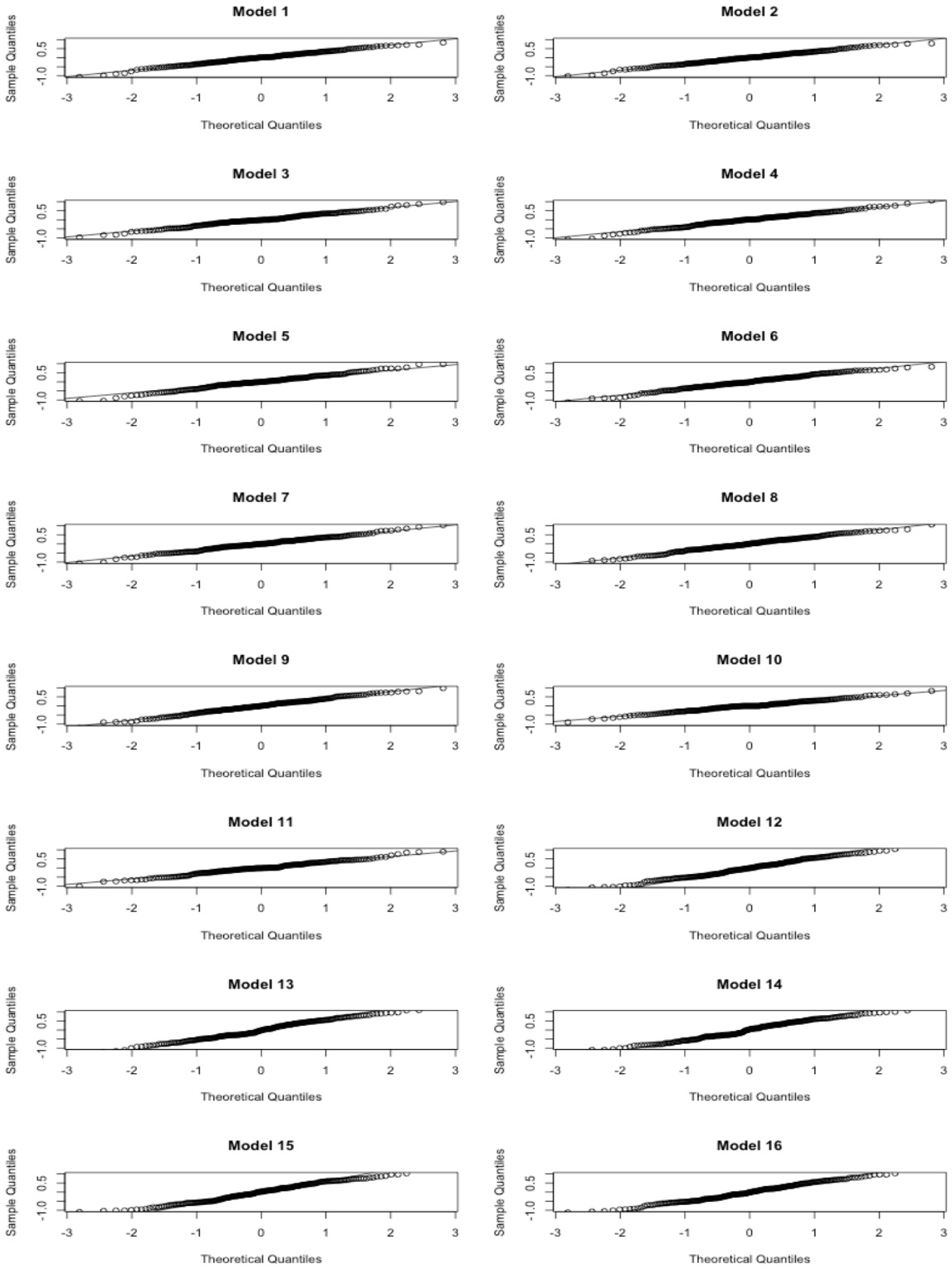


Figure A3.1: Normality Plots (QQ-Plots) for the Shannon-Weiner index. The theoretical quantiles are given by a normal distribution. Model descriptions are provided in Table 3.3. This figure shows that the normality assumption of each model is satisfied for the data.

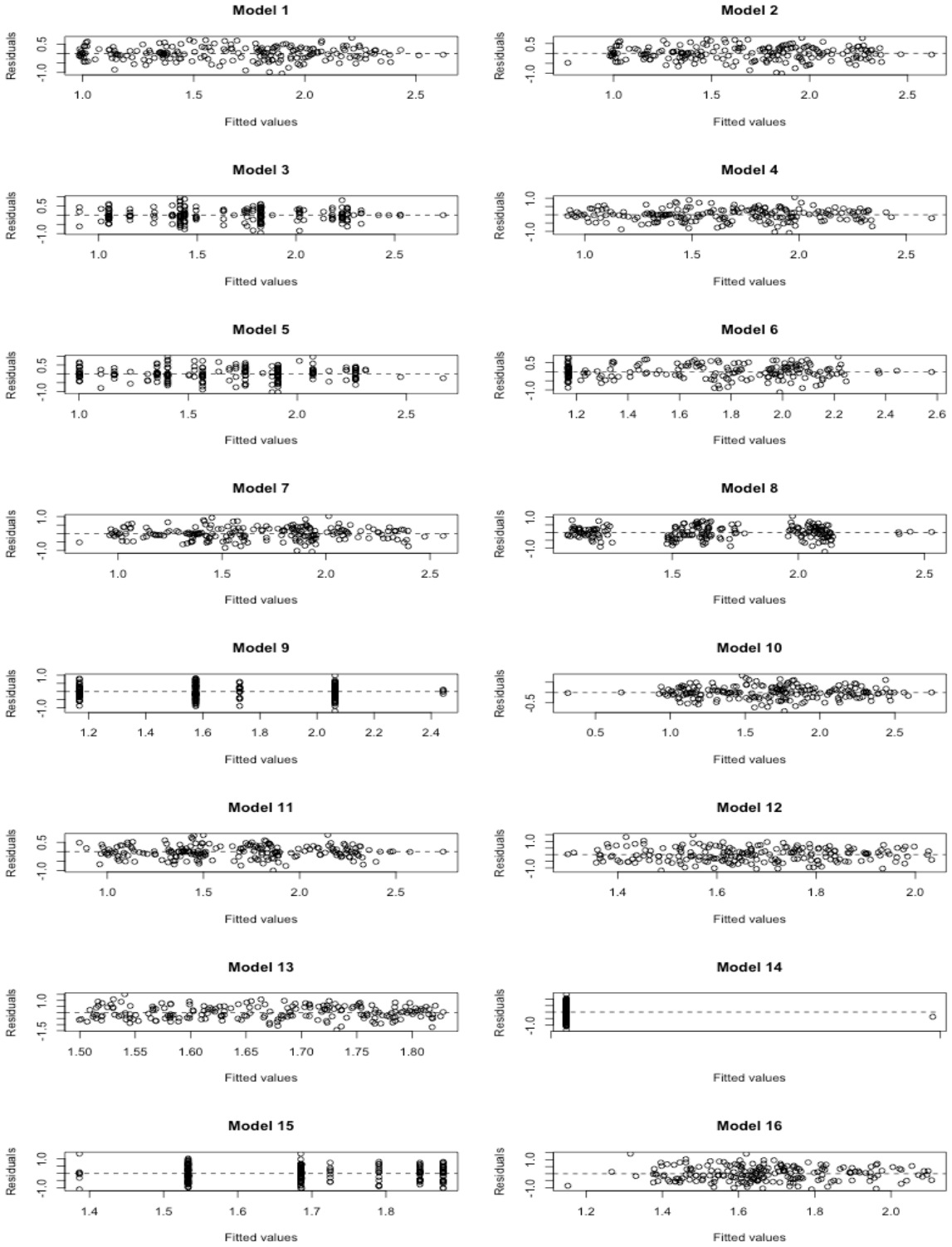


Figure A3.2: Residuals versus fitted coefficient plots. For model descriptions refer to Table 3.3. This figure shows that the linearity assumption for the predictor variables in each model is reasonable, the variance of the residuals are equal (homogeneity of variance), and there are few outliers in each model.

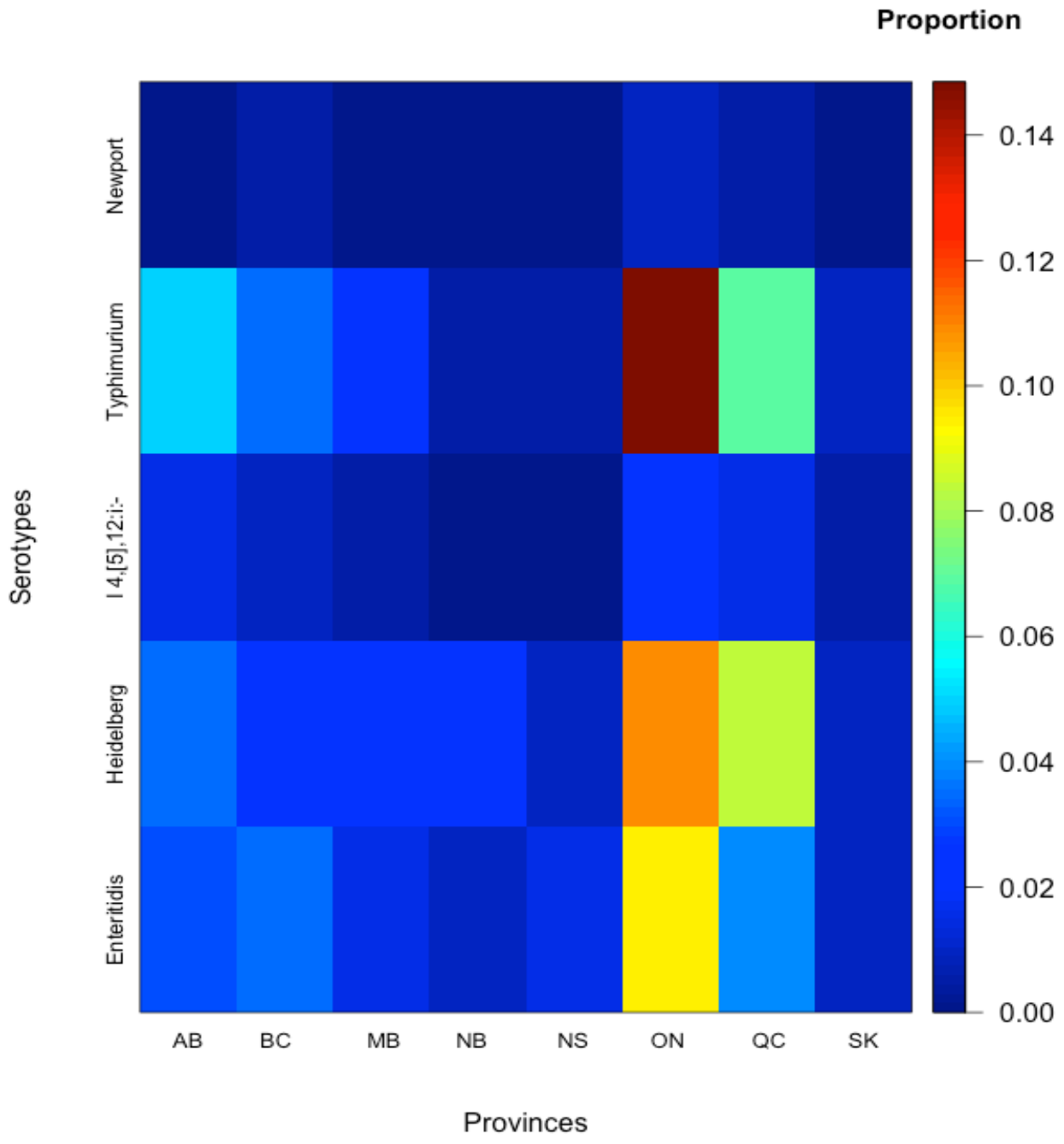


Figure A3.3: Heat Map showing the distribution of *S. enterica* serotypes in each of the provinces. Different colors indicate different proportions as explained by the legend on the right side of the map. This map shows no obvious relationship between provinces and serotypes and hence, the use of province and serotype as independent predictor variables in our models is warranted.

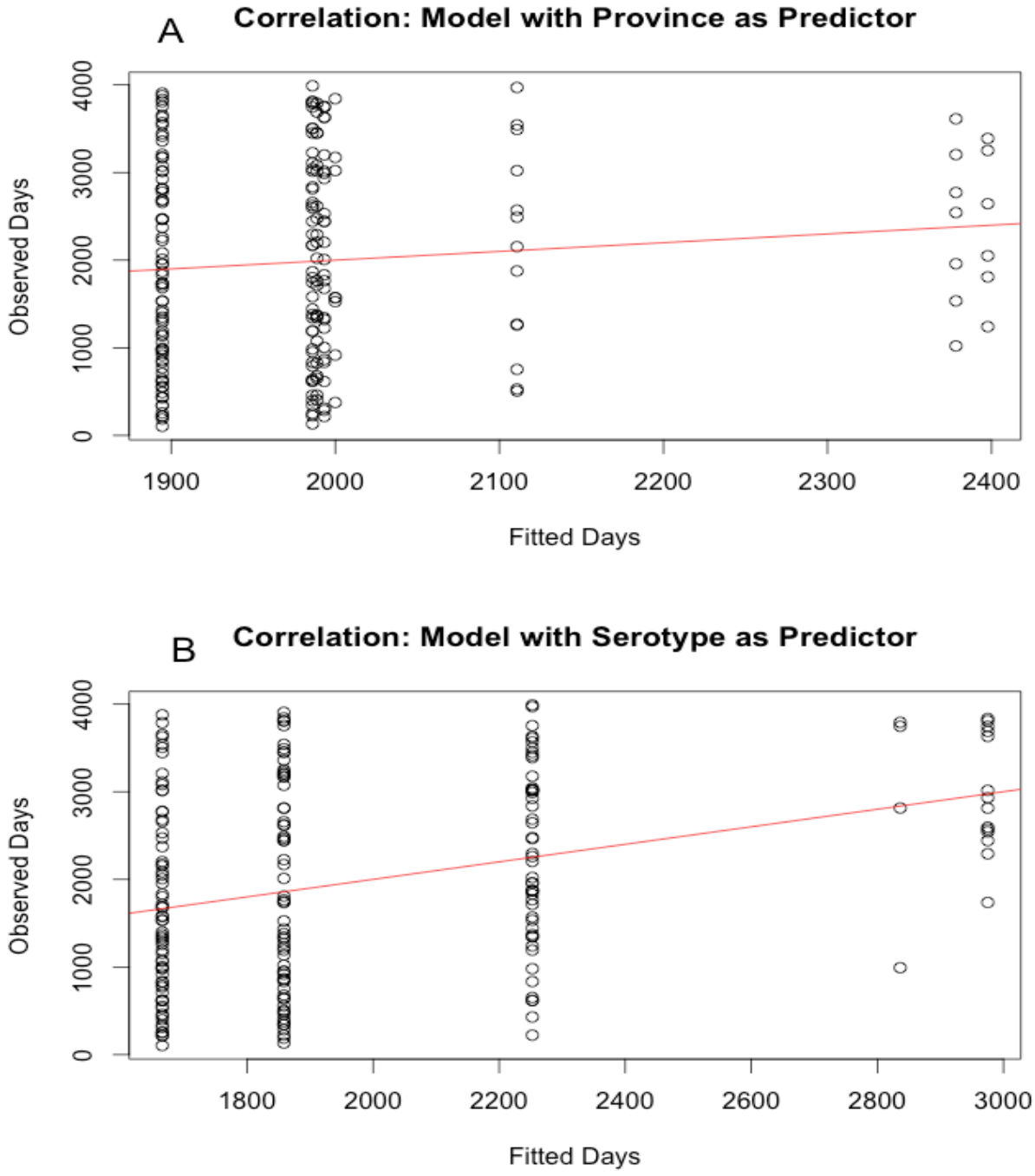


Figure A3.4: Correlation between Days and Provinces or Serotypes. The upper panel (A) shows a weak correlation between observed Days and model fitted Days with Province as a predictor variable (Pearson correlation coefficient, r^2=0.11). The lower panel (B) shows weak correlation between observed Days and model fitted Days with Serotype as a predictor variable (Pearson correlation coefficient, r^2=0.33). The red line in each graph represents the goodness of model fit.

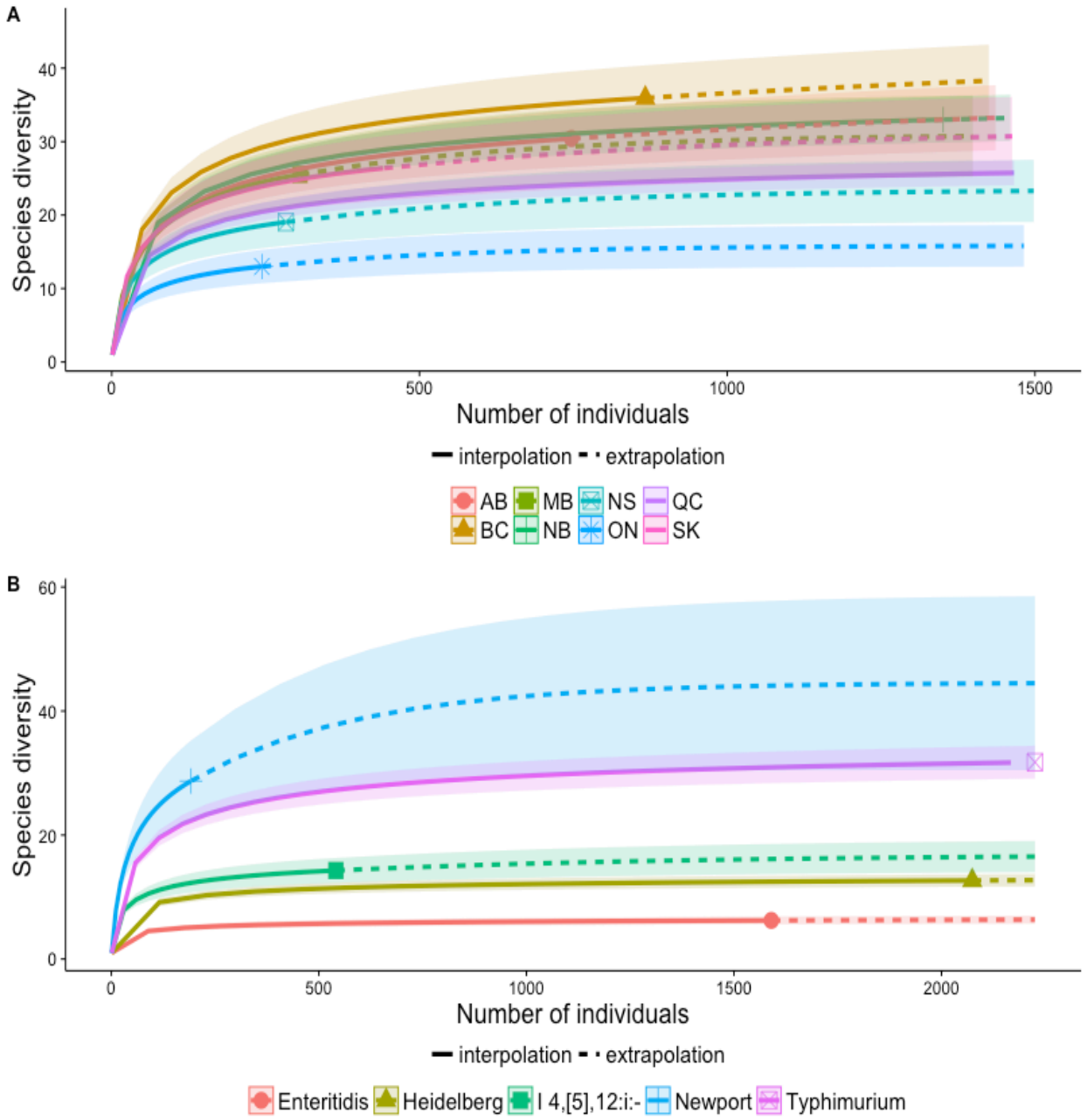


Figure A3.5: Rarefaction and Extrapolation Curve for comparison of species diversity between provinces and between serotypes at different sample sizes. The upper panel (A) shows the rarefaction and extrapolation curve for provinces. The solid lines represent rarefaction and the dotted lines represent extrapolation to the largest sample size for province. Different colors represent different provinces as shown in legend under the graph. The lower panel (B) shows the rarefaction and extrapolation curve for serotypes. The solid lines represent rarefaction and the dotted lines represent extrapolation to the largest sample size for serotypes. Different colors represent different serotypes as shown in legend under the graph.

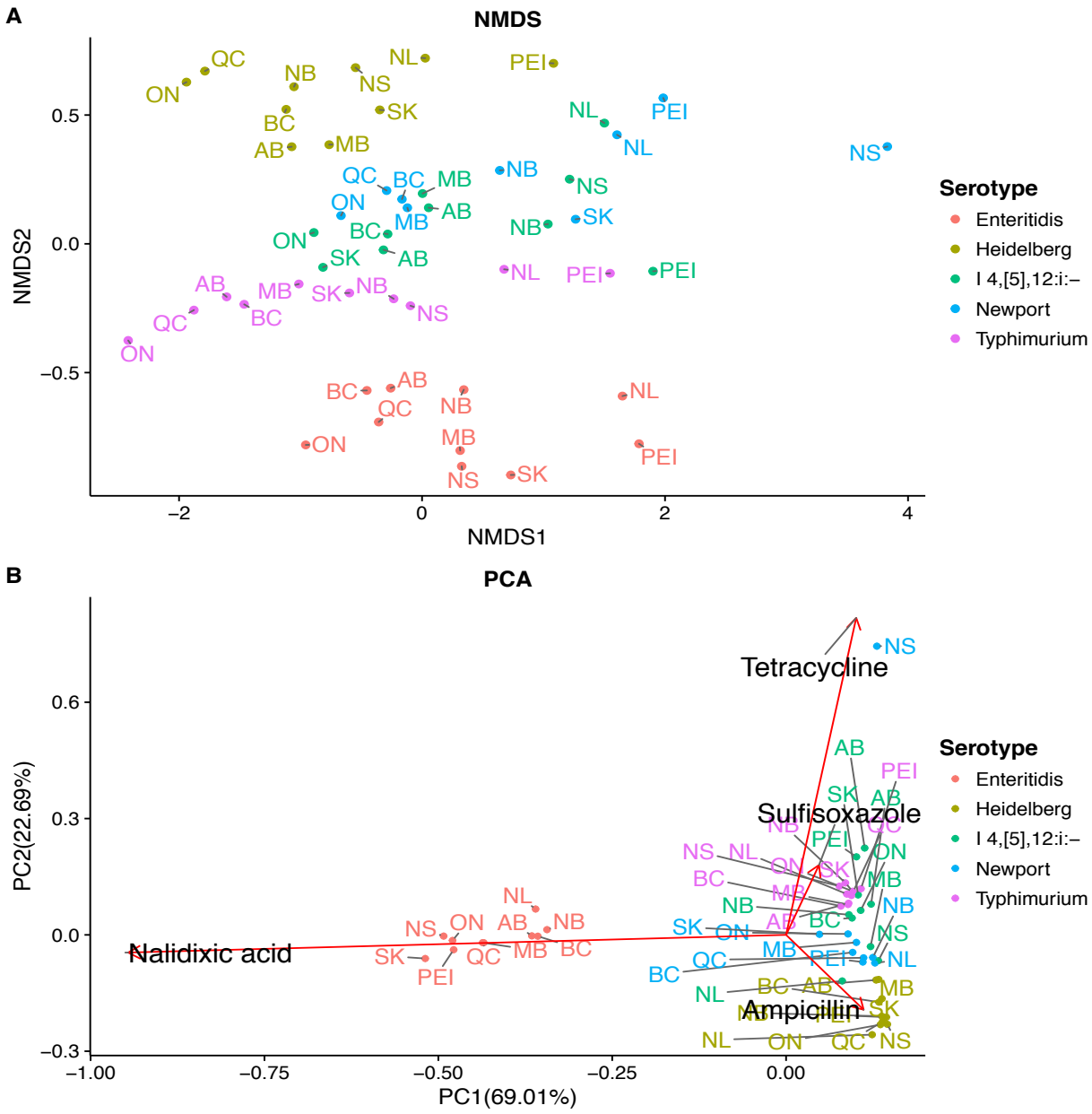


Figure A3.6: Beta diversity for *Salmonella* serotypes within Canadian provinces. Points are labeled with province abbreviations and grouped with respect to serotypes (colors, see the legend to the right). Panel (A) shows the NMDS for province-serotype combination (Stress = 0.06) on raw antibiotic resistance data. Stress represents the mismatch between the rank order of the pairwise dissimilarities due to Bray-Curtis distance and in the NMDS ordination. Panel (B) shows the PCA for province-serotype combinations performed on the variance-covariance matrix of antibiotic resistance data transformed into proportions. PC1 explains 69.01% and PC2 explains 22.69% of the variance in the distribution of antibiotic resistance. Vectors point in the direction of higher than average relative abundance of particular antibiotics.

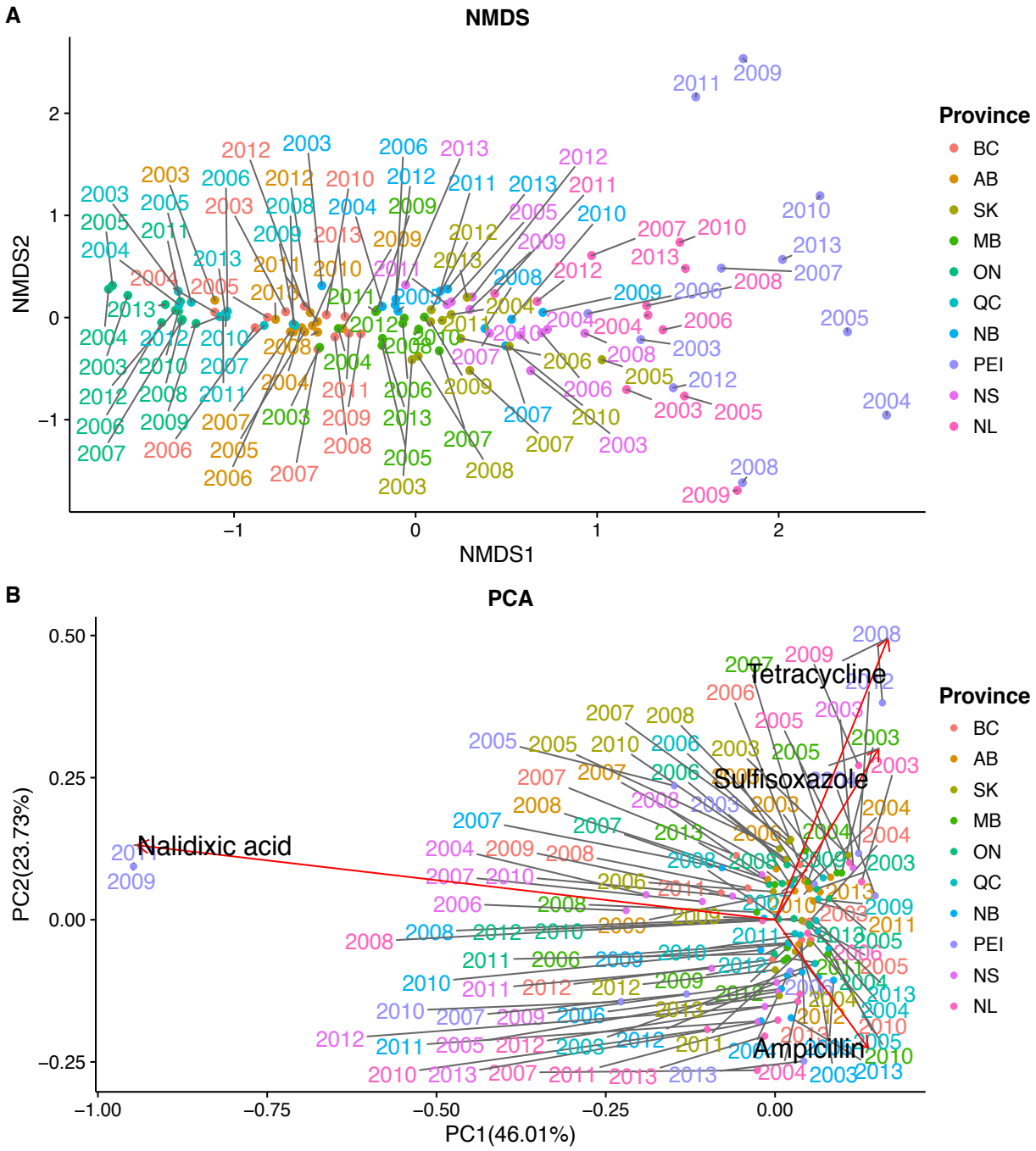


Figure A3.7: Beta diversity for Canadian Provinces in different years. Points are labeled as years and grouped with respect to provinces (colors, see the legend to the right). Panel (A) shows the NMDS for province-year combination (Stress = 0.07) on raw antibiotic resistance data. Stress represents the mismatch between the rank order of the pairwise dissimilarities due to Bray-Curtis distance and in the NMDS ordination. Panel (B) shows the PCA for province-year combinations performed on the variance-covariance matrix of antibiotic resistance data transformed into proportions. PC1 explains 46.01% and PC2 explains 23.73% of the variance in the distribution of antibiotic resistance. Vectors point in the direction of higher than average relative abundance of particular antibiotics.

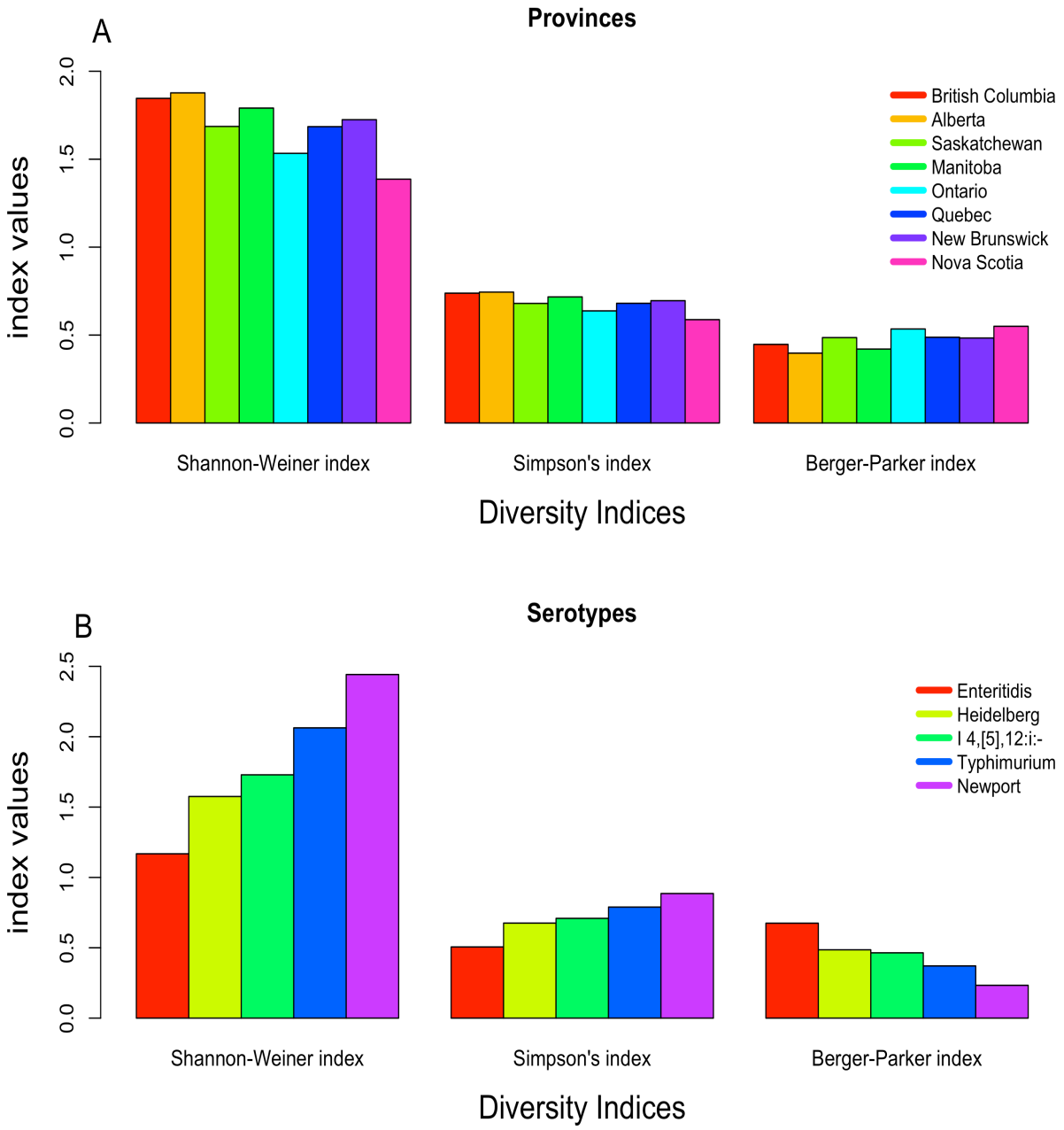


Figure A3.8: Bar plots for comparing three different alpha diversity indices for Provinces (A) and Serotypes (B) in the *S. enterica* dataset following the 30 isolate sampling strategy. Different color in (A) represents provinces and in (B) represents serotypes, as shown in the legend. This figure shows that different results are obtained using different alpha diversity indices as they have different properties (such as emphasis on rare versus common resistance types; refer to Table 2.2 for diversity indices and their properties).

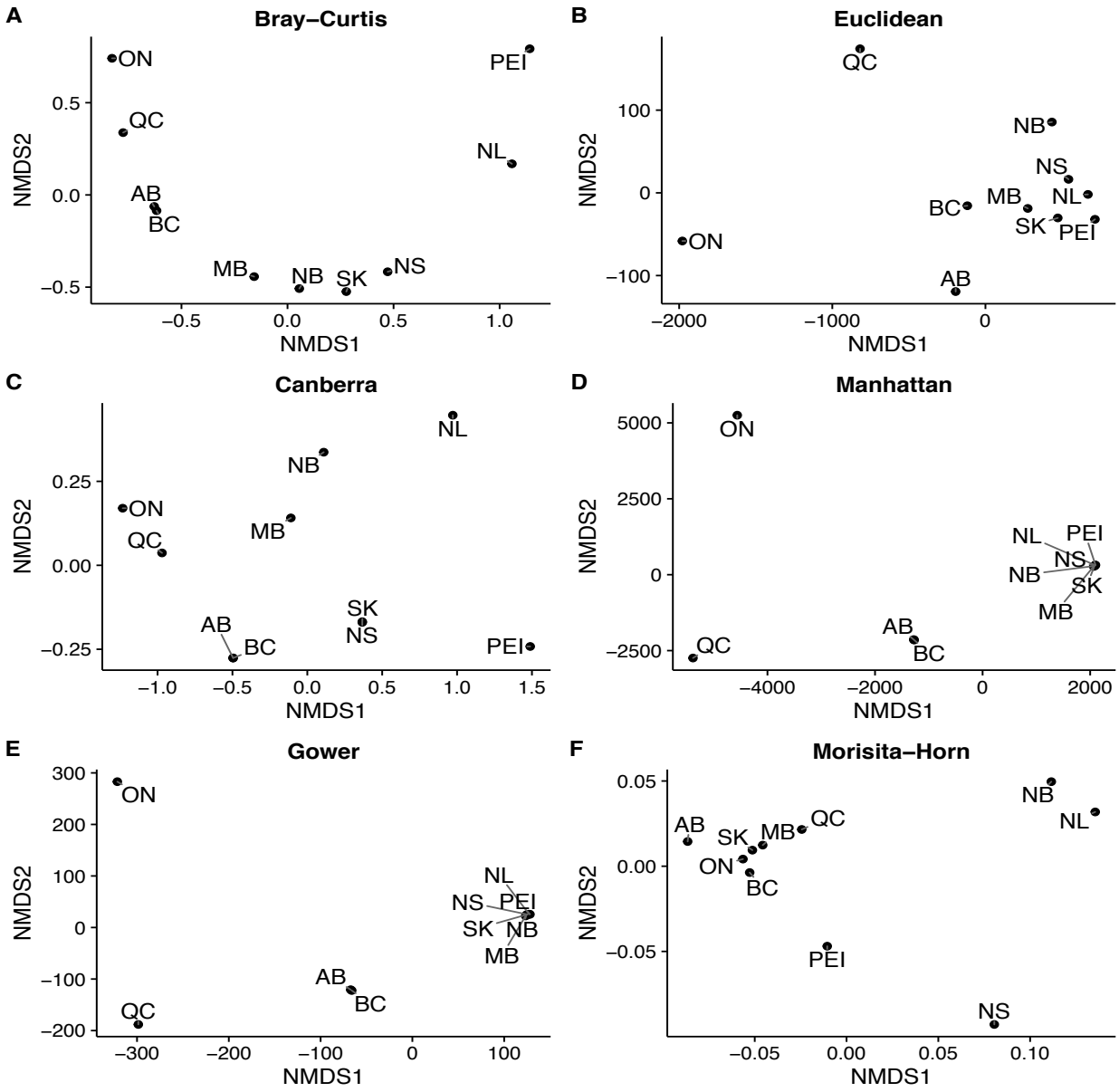


Figure A3.9: NMDS comparing different beta diversity measures for Provinces in the *Salmonella* dataset. The beta diversity measure with which NMDS was performed is mentioned at the top of each graph. Each point in each NMDS represents a province. Stresses for each NMDS are: (A) 2 x 10^{-3}, (B) 2.43 x 10^{-5}, (C) 4 x 10^{-4}, (D) 6.83 x 10^{-5}, (E) 6.20 x 10^{-5}, and (F) 0.08. Stress for each NMDS represents the mismatch between the rank order of the pairwise dissimilarities, between provinces, due to beta diversity measure and in the NMDS ordination. This figure shows that similar or different results may be obtained using different beta diversity measures as various measures have similar or different properties (exclusion/inclusion of joint absence, emphasis on rare/common species; refer to Table 2.3 for indices and their properties).

DATA Request Form to CIPARS:

Centre for Foodborne, Environmental and Zoonotic Infectious Diseases (CFEZID) and Canadian Integrated Program for Antimicrobial Resistance Surveillance (CIPARS)

Data Request Form

In order to keep track of requests for data held by CIPARS and to ensure that you receive the information that you are looking for, please check off the appropriate boxes on the following pages and provide any other additional details when indicated. Please also fill in your contact information below.

Contact Information:

Name: **Amy Hurford**

Organization/Committee: **Memorial University of Newfoundland**

Phone number: **(709) 726-1561 (before Dec 22, 2014) or (709) 864-8301 (after Jan 1, 2015)**

Email address: **ahurford@mun.ca**

CFEZID/CIPARS contacts:

Rita Finley
PH : 519-826-2213
E-mail : Rita.Finley@phac-aspc.gc.ca

1. **Purpose of Request**
   - a. Outbreak Investigation ☐
   - b. Supplemental Data ☐
   - c. Research Project[1] ✓
   - d. Publication ✓
   - e. Other ✓
     - i. Please specify: **MSc thesis**

2. **AMR Data request details (please complete as much detail as possible)**
   - a. Salmonella serotype(s): **All available**
   - b. Phage type(s): **All available**
   - c. Region(s) of concern
     - i. Canada ✓

[1] *Please attach a copy of the project proposal.*

ii. Province/region ☐
    1. Please specify: _____

d. Time frame
    i. Year(s): **All available**
    ii. Month(s): **All available**

3. Data variables requested:

a. Laboratory data
    i. Bacterial genus ✓
    ii. Serotype ✓
    iii. Phage type ✓
    iv. PFGE patterns ☐
    v. Antimicrobial susceptibility patterns ✓
    vi. Other ☐
        1. Please specify: _____

b. Epidemiological data
    i. Project/surveillance component ✓
    ii. Year ✓
    iii. Month ✓
    iv. Province/region ✓
    v. Other ☐
        1. Please specify: _____

4. Preferred format of the data:

a. Standard CIPARS summary tables ✓
b. Standard CIPARS summary figures ☐
c. "One-off" summary tables (see next page) ☐

Please provide specific details of other formats or in terms of how the r x c tables etc. should be categorized and stratified here:

None

5. AMU Data request details (please complete as much detail as possible)

a. Human AMU Data
    i. *Pharmacy* ☐
    ii. *Hospital* ☐
    iii. Physician ☐

b. Antimicrobials of interest
   i. *All* ☐
   ii. *Cat. I* ☐
   iii. *Cat. II* ☐
   iv. *Cat. III* ☐
   v. *Cat. IV* ☐

c. Others (Specify): _____
   _____

d. Region(s) of concern
   i. Canada ☐
   ii. Province/region ☐
      1. Please specify: _____

e. Time frame
   i. Year(s): **All available**
   ii. Month(s): **All available**

6. Data variables requested:
   i. Total prescriptions ☐
   ii. Total cost ☐
   iii. Total kilograms ☐
   iv. Total number of units ☐
   v. Defined daily doses ☐
   vi. Province ☐
   vii. Total diagnoses ☐
   viii. Total antimicrobial recommendations ☐
   ix. Age groups ☐
   x. Disease classification/code ☐
   xi. Other:
      1. Please specify: _____

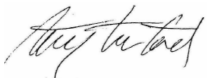
7. Preferred format of the data: Microsoft Excel or CSV

PLEASE NOTE:
You are receiving confidential information, as such, it should be kept in strict confidence. This confidential information can be used only for the purpose outlined above. The data should not in any way be manipulated to allow the re-identification of any patient/case.

This data should be protected to prevent any unauthorized use, dissemination or publication outside of its intended use. There will be no additional copies, extracts or reproductions made of the confidential information with the exception for sharing on a need-to-know basis for the uses outlined above.

By signing below, you are agreeing to keep this information confidential and not using it for any purposes outside of those outlined above. The data will be deleted from all systems immediately after the completion of the project.

Signed by:



Amy Hurford Date: **December 10, 2014**