INVESTIGATING THE k-NEAREST NEIGHBORS RESOLUTION ALGORITHMS FOR PYROPRINTS AND CLUSTERING FOR BACTERIAL STRAINS

A Thesis

presented to

the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Computer Science

by

Jeffrey D. McGovern

March 2016

COMMITTEE MEMBERSHIP

TITLE:	Investigating The *k*-Nearest Neighbors Resolution Algorithms for Pyroprints and Clustering for Bacterial Strains
AUTHOR:	Jeffrey D. McGovern
DATE SUBMITTED:	March 2016
COMMITTEE CHAIR:	Alexander Dekhtyar, Ph.D. Professor of Computer Science
COMMITTEE MEMBER:	Chris Kitts, Ph.D. Professor of Biological Sciences
COMMITTEE MEMBER:	Foaad Khosmood, Ph.D. Professor of Computer Science

ABSTRACT

Investigating The *k*-Nearest Neighbors Resolution Algorithms for Pyroprints and Clustering for Bacterial Strains

Jeffrey D. McGovern

Fecal contamination in bodies of water is an issue that frequently plagues public and environmental water supplies. Finding the source of the fecal matter can help prevent further contamination and more quickly curb future contamination. Microbial Source Tracking (MST) aims to determine the source host species of strains of microbiological lifeforms and library-dependent MST is one method that can assist in this fecal matter sourcing. Recently, the Biology Department and the Computer Science Department at California Polytechnic State University San Luis Obispo (Cal Poly) teamed up to build a database called The Cal Poly Library of Pyroprints (CPLOP). Students collect fecal samples, culture *E. coli* isolates from the samples to pyrosequence the two internal transcribed spacer (*ITS*) DNA regions of *E. coli*, and insert this data, called pyroprints, into CPLOP. This work investigates two MST methodologies that use CPLOP: a bacterial strain-based approach and an isolate-based approach. By using DBSCAN to build bacterial strains, we found that between 41% and 51% of the clustered data fell into pure strains, while another 34% to 43% fell into a cluster where its host species was the most dominant, validating the effectiveness of using *E. coli* in CPLOP. We also verified the expected existence of transient strains and found them to be few in number. Unfortunately, between 27% and 53% of the data remained unclustered. As a fallback, we turn to the *k*-Nearest Neighbors Resolution Algorithms for Pyroprints (*k*-RAP), which consists of four strategies to resolve the multiple *k*-nearest neighbors lists that result from querying the two *ITS* regions of an *E. coli* isolate. It provides us a variety of resolution strategies that garner between 65% and 85% overall accuracy and over 75% accuracy for well-represented host species.

ACKNOWLEDGMENTS

Thanks to:

- My parents, Bill and Katrina McGovern, and my sister Ashleigh for their love, patience, and support throughout my life.
- Chris and Kimberly Paterson Hunt, who have been my friends since the beginning of this academic journey.
- Cameron and Aline Christensen, for their friendship, beer knowledge, and memes.
- George and Carole Leone, for their unending support and guidance throughout my academic career.
- Andrew Guenther, for uploading this template.
- Roommates Kimberly Paterson Hunt, Eriq Augustine, Chris Hunt, Maria Pikusova, Nicole Martin, and Cody Hunt for the strength of our roommunity.
- Alex Dekhtyar for his insight and advisement in the fields of data science, Eurasian indie bands, and novel science fiction and fantasy used book stores.
- Chris Kitts, Michael Black, and Jennifer VanderKelen for editing my papers and teaching me everything I needed to know about poop to succeed.
- Ilene, Tom, and Brian Jones; Andrew Wang and Daniel Gerrity; and Andrew, Katia, and Finn Guenther for housing me while I presented parts of this thesis at bioinformatics conferences around the country.
- The Outrun, Retrowave, and Synthwave genres of music and artists Lazerhawk, FM-84, and Carpenter Brut, whose driving synths and 80's aesthetic impelled me to finish this thesis.
- And Paul Biggins, for his love and unwavering dedication.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

LIST OF ALGORITHMS

Chapter 1

INTRODUCTION

Fecal contamination in public water sources is an issue that health officials and city and county governments must frequently combat. Pathogens present in fecal matter pose severe health risks to humans and pets and the decomposition of fecal coliform bacteria can upset the balance of aquatic ecosystems by depleting dissolved oxygen to low enough levels that it may kill other species in the water [28]. Such severe threats to the health of humans, pets, and the local ecosystem motivates public health officials to take action in order to mitigate its consequences. Often times, no one observes the cause of the fecal contamination, but rising levels of fecal coliform bacteria indicate that fecal contamination has occurred. In these situations, usually the only course of action that natural resource managers have is to simply restrict public access until contamination levels reach an acceptable level, which may not prevent further contamination [59]. Identifying the source of fecal contamination in water supplies is an important initial step to prevent further contamination [66].

Microbial Source Tracking (MST) is the field of research that aims to discover the host species that microbial lifeforms originate from and aids the process of sourcing fecal contamination. Microbes thrive inside the gut of animals, as well as in masses of plant matter, and routinely make their way into the environment via fecal matter deposition. Biologists conjecture that strains of microbes or bacteria present in fecal matter, called fecal indicator bacteria (FIB), remain relatively unique to the species of the host they originated from. A strain of a species of microbe is a subtype of that species where the microbes in that strain are closely related in some meaningful way. How researchers specifically define strains often differs, since each definition of a strain depends on the characterization of the microbe in question and the methods used to derive such characterizations [66]. Typically, the objective is to discover which microbes of a bacterial isolate came from the same parent microbe [36]. A strain, then, can be thought of as consisting of individuals descending from the same individual to generate a “group” or “family.” Researchers put significant effort into choosing the relevant microbes and appropriately characterizing them in order to discover which strains tend to belong to which species.

A common method of MST known as library-based MST involves collecting fecal matter from a known host species, culturing isolates of the relevant microbes in the fecal matter, and building a digital representation of the collected isolates for storage into a database and analysis. Storing an appropriate digital representation allows researchers to perform rigorous analysis and comparison between FIB isolates collected from different host animals and host species, as well as isolates collected from the same host animal, but at different times. The data inserted into such a database may range from collection metadata about the microbiome, to a specific microbe characterization, or to any other useful set of metrics that can appropriately profile an entry [58].

In this way, researchers build a “library” of known-host species isolates. Using this library, researchers can take an environmental sample with FIB from an unknown source, process the microbial isolates using the same procedure and the known-host species isolates in the library, and compare the strain representation of the environmental sample to those in the library to find any close matches. Since the researchers know the host species of the isolates in the library, they can make a reasonable determination of the source of the isolates in the environmental sample. The methods used to compare isolates and make assertions depends entirely upon the FIB, their method of collection, and their digital representation.

Library-based MST is usually only effective within the region in which the known-host species isolates came from, making it difficult to build a “one size fits all” library. Companies exist that can, for fees in the ballpark of $100, attempt to determine the host species of a provided sample and while these companies exist nationwide, they are few in number and usually cannot build a representative sampling of every region for accurate host species determination. Additionally, when investigating an incident of fecal contamination, investigators want to send out multiple samples to build reliable evidence for a determination of the source. As a result, the cost of outsourcing becomes too prohibitive and determinations too inaccurate for it to be an option. Thus, there exists a need for a cost-effective and accurate method of MST in order to properly tackle the problem of preventing fecal contamination in water supplies.

In 2009, the California Polytechnic State University San Luis Obispo (Cal Poly) Biology and Computer Science Departments built The Cal Poly Library of Pyroprints (CPLOP) [67], a database of *Escherichia coli* (*E. coli*) isolate fingerprints, called pyroprints. Students collect fecal samples from a variety of host species from the San Luis Obispo area and build the pyroprints using a low cost DNA sequencing method called pyrosequencing on two intergenic regions of an *E. coli* isolate. Building pyroprints ends up costing roughly two orders of magnitude less than outsourcing samples, cutting the cost of building an effective MST library by as much as 60% [9]. It is through CPLOP that Cal Poly researchers hope to better understand bacterial strains, effectively differentiate between them, and provide a cost-effective MST methodology.

In order to be an effective MST fingerprinting method, pyroprinting must contain information that allows for the accurate discrimination between closely related strains of *E. coli* bacteria. Internal Transcribed Spacers (*ITS*) in bacteria are regions of DNA that do not contain instruction for building proteins and thus have high variability, since variability across generations of bacteria does not affect the survivability of the microbe. Because of this high variability, researchers can use these regions to differentiate between strains of the same species of microbe. *E. coli* isolate pyroprints stored in CPLOP represent the polymerase chain reaction (PCR)-amplified regions of DNA between the *16S* and *23S* genes and *23S* and *5S* genes, referred to as *ITS-1* and *ITS-2* respectively. *ITS-1* and *ITS-2*, along with the entire *E. coli* genome, repeat seven times, giving us seven highly variable regions for each *ITS*. Any offspring inherit mostly accurate copies[1] of the *ITS* regions of the parent microbe, encoding the notion of a "group" or "family" and allowing researchers to use them to differentiate between strains [68]. By building pyroprints out of these regions, CPLOP researchers hope to form a reliable notion of an *E. coli* strain that they can use for MST.

A pyroprint is a vector comprised of the peak heights of pyrosequences of multiple copies of a repeated region of DNA. By dispensing a series of nucleotides at specific times and observing the resulting light emitted, CPLOP researchers can build a fingerprint of that DNA sequence. In traditional pyrosequencing, the DNA sequenced is an amplified version of a single sequence of DNA, allowing researchers to reconstruct the exact sequence of nucleotides that make up the DNA. Since C-PLOP researchers pyroprint segments of DNA that repeat but are highly variable, researchers cannot reconstruct the exact sequences of the *ITS* sequences. Instead, CPLOP contains pyroprints that represent the random variability in the genome of *E. coli* isolates. Previous work in [65] optimized the pyroprinting process, including the dispensation sequence and peak height determination, for each *ITS* to best delineate between different strains of *E. coli* using the Pearson correlation coefficient to compare pyroprints.

The Pearson correlation coefficient ρ normalizes the covariance of two vectors by the standard deviation of each, providing a notion of relative co-variability between

[1] Some variation may occur, but researchers assume it is small for immediately related microbes and large for distantly related microbes.

the vectors that remains invariant of noise and scaling — a core reason why CPLOP researchers use it to compare pyroprints. In order to compare two *E. coli* isolates in CPLOP, researchers must separately compare the *ITS-1* pyroprints to each other and the *ITS-2* pyroprints to each other using Pearson correlation coefficient. It is meaningless to compare different *ITS* to each other since they represent entirely different sections of DNA obtained through a different sequence of dispensations. This effectively gives us two comparison metrics between isolates: the Pearson correlation coefficient between two *ITS-1* and the Pearson correlation coefficient pyroprints between two *ITS-2* pyroprints — $\rho_{ITS\text{-}1}$ and $\rho_{ITS\text{-}2}$. Using these values, CPLOP researchers rigorously define the notion of a strain as isolates with ρ higher than 0.990 in both *ITS* regions.

CPLOP supports numerous research projects, ranging from longitudinal studies of a host animal to large studies of one or more host species, in order to understand the evolution and transmission of *E. coli* strains and verify that pyroprinting provides an accurate representation of *E. coli* strains. Previous work on CPLOP include formation and validation of the pyroprinting process and exploration of the evolution and transference of *E. coli* strains within and between host animal and host species, and new algorithms designed specifically for CPLOP to better understand its data [41, 31].

Much of the work done so far using CPLOP has been exploring the composition, evolution, and transference of strains among host animals and host species. While part of this is to validate the MST methodology that leverages CPLOP data, researchers gain a large amount of insight into how *E. coli* strains get into and evolve in fecal matter by using pyroprints to rigorously study changes. Clustering methods become very useful in this case, owing their effectiveness to the notion of a strain being similar to a "group" or "family" of a closely related subtype of a species of microbe.

Two pieces of previous work, [41, 42] and [31], worked toward building clustering algorithms that can provide meaningful insight into the *E. coli* isolates in CPLOP. The former, *OhClust!*, is an agglomerative clustering algorithm where a biologist-provided metadata-ontology guides the agglomeration. The latter, by Eric Johnson, is a density-based clustering algorithm — DBSCAN — optimized by a fast range query for nearby isolates.

While *OhClust!* takes advantage of all of the information available in CPLOP, DBSCAN encodes our notion of a strain the closest and allows for strain discovery without needing to guess what ontology provides the best insight. Moreover, the range query optimizations made in [31] allow for efficient, low-memory querying of isolates while still encoding the notion of Pearson correlation coefficient between isolates, an improvement over *OhClust!*'s need to precompute and store distances in order to mitigate the consequences of agglomerative clustering's need for a high number of distance computations. In a nutshell, DBSCAN uses a distance metric, a minimum neighbors value `MinPts`, and an ε range to categorize data points as one of three types — core point, border point, or noise — building clusters from every core point within ε and their associated border points.

While the original purpose of CPLOP was to support MST and some manual MST studies have been conducted, little research has been done on building an automated MST method. Most studies performed with CPLOP focused on validating and exploring the various biological features captured by the pyroprinting process and the comparison metric used to compare pyroprints, the Pearson correlation coefficient. Building objective, repeatable classification metrics that use the data in CPLOP to assist MST can help biologists inform investigators of a possible source that caused, or is causing, fecal contamination.

Since strains are core to CPLOP, it is appropriate for us to incorporate them into our MST methodology and prudent to have a fallback if that methodology fails to make a determination. The first MST method classifies an unknown-source isolate thus: incorporate the isolate into CPLOP, construct strains, and classify the host species of the unknown isolate as the dominant host species of the strain it ends up in. The second, fallback technique simply compares the unknown isolate to every isolate in CPLOP and classifies as the dominant host species of the most similar isolates. For the former, we investigate DBSCAN from [31] and used for MST in [38]. For the latter, fallback method, we investigate The k-Nearest Neighbors Resolution Algorithms for Pyroprints (k-RAP) from [37].

In [38], we constructed the notion of a bacterial strain purely from the clusters produced by DBSCAN — i.e. we defined bacterial strains to be the clusters produced by DBSCAN. We studied the cluster purity — the proportion of isolates in a cluster that are of the same species — of the entire clustering at different `MinPts` values. For `MinPts` between 1 and 5, respectively, we are able to cluster between 72.9% and 52.1% of the isolates in CPLOP, with between 51.0% and 41.2% falling into pure clusters and another 34.4% to 43.8% falling into clusters where their host species is the most dominant. In this investigation, we also confirmed the presence of so-called transient *E. coli* strains — strains of *E. coli* that show up in many different host species — that tend to confound MST. More importantly, it showed that CPLOP has relatively few of these transient strains and a large number of pure strains. Its limitation comes in the coverage of CPLOP — how much of our dataset we actually cluster.

As a first step towards building an effective classification technique, we chose to use the k-Nearest Neighbors (k-NN) classification algorithm on CPLOP to measure how accurately we can classify samples that we know the host species of. k-NN classifies an unknown-class datum by querying a library of known data — each datum has a class, or classification — for “nearby” data; sorts the list by nearness, limiting it to k many “neighbor” data points; and classifies from this “k-nearest neighbors” list by picking the most plural classification present among the neighbors, breaking ties by average position.

A somewhat unique obstacle arises with the *E. coli* isolates in CPLOP: in order to compare isolates, we must use two different comparison metrics — $\rho_{\mathit{ITS\text{-}1}}$ and $\rho_{\mathit{ITS\text{-}2}}$. For k-NN on CPLOP *E. coli* isolates, this means that we produce two k-nearest neighbors lists that we must classify from. Resolving multiple k-NN lists can be useful for any data that has multiple meaningful-yet-exclusive ways to compare one datum to another. Biologists using k-NN will likely want to restrict the list further than k, since their definition of a strain relies heavily on bounding the Pearson correlation coefficient between two isolates for both *ITS* — so we also add an α threshold to further limit the involved k-NN lists.

The four resolution algorithms, called the k-NN Resolution Algorithms for Pyroprints (k-RAP) and previously published in [37], are termed: Meanwise Resolution, Resolution by Winner, Resolution by Union, and Resolution by Intersection. Meanwise Resolution takes the average of the comparison value to form a single k-NN list. Resolution by Winner finds the most plural classification in each k-NN list and picks the classification with the most instances of that class in its list. Resolution by Union combines each k-NN list into a single set — performing effectively a union on all of the k-NN lists — and finds the most plural classification in the resulting set, breaking ties by average original position. Resolution by Intersection forms a new set that is exactly the isolates that appear in every k-NN list — effectively performing an intersection at the isolate level — expanding both lists and adding to the set until the set itself is of size k and choosing the most plural class of the new set.

Investigating k-RAP in [37] showed us that classification accuracy for the entire database stayed 65% and 85% for all of the algorithms. Precision and recall for well-represented host species also stayed safely above 0.30, which is far better than random and notably better than our outsourced baseline. Underrepresented host species predictably performed poorly in classification, with k-RAP accuracy dropping below 0.30 for precision and recall. Furthermore, α thresholding noticeably improved performance on some resolution algorithms, namely Resolution by Intersection, which saw an improvement from 74.7% for no filtering to 85.9% for the strictest filtering at $\alpha = 0.99$. A close second in accuracy was Resolution by Union at 76%, but saw no noticeable improvement when adding α filtering.

This thesis consists of the work done in [37] and [38], with deeper explanations and investigation into the results. Namely, the contributions of this work are the following:

- k-NN Resolution Algorithms for Pyroprints: Modifications to the k-NN classification algorithm that can resolve multiple comparison metrics
- A modification to k-NN that adds α thresholding to further restrict the individual k-NN lists
- An empirical study measuring the accuracy of identifying the host species for the *E. coli* isolates stored in CPLOP, investigating how values of k and α affect the accuracy with each resolution metric
- An investigation of the efficient density-based clustering algorithm in [31] that is scalable and meaningfully encodes the comparison-metric used in CPLOP
- A set of validation measures for clustering bacterial isolates into strains
- An evaluation of our strain discovery procedure based on the defined set of measures

The rest of this thesis is organized as follows: Chapter 3 provides an overview of relevant work in the field of MST and an introduction to work done using CPLOP;

Chapter 2 details CPLOP and the background necessary to understand the algorithms presented; Chapter 4 abstractly describes the strain-based and isolate-based approaches to MST that we investigate; Chapter 5 describes the DBSCAN clustering method, lists our evaluation criteria, and discusses the results; Chapter 6 describes the k-Nearest Neighbors Resolution Algorithms for Pyroprints, lists our evaluation criteria, and discusses the results; and Chapter 8 concludes, offering suggestions for future work.

Chapter 2

BACKGROUND

2.1 Biological

This chapter provides an overview of the problem that clustering for bacterial strain and k-Nearest Neighbors Resolution Algorithms for Pyroprints attempts to solve. Fecal contamination is a serious health concern that public officials work hard to mitigate the effects of. Microbial Source Tracking is a solution that these officials typically employ in order to discover the source of fecal contamination and avert more contamination. Bacterial strain typing is just one of these techniques that The Cal Poly Library of Pyroprints employs in a cost-effective manner.

2.1.1 Fecal Contamination

Fecal contamination is dangerous for humans and animals alike. Pathogenic bacteria reside in fecal matter and contamination of publicly accessible and environmental water sources expose humans and animals to their dangerous effects. Often, it is up to natural resource managers and public health officials to both eliminate and prevent fecal contamination from occurring. Preventing contact with such bacteria is key to maintaining good public and environmental health.

Fecal matter can contain many different types of microbes, some of which are pathogenic. A *pathogen* is a microbe that causes disease in an animal. Pathogenic microbes that make their way into fecal matter include *Escherichia coli* (*E. coli*), *Salmonella typhii*, and *Campylobacter*. Diseases they might cause can include nose and ear infections, dysentary, and typhoid fever. Zoonotic pathogens — those that wild animals, pets, or livestock can transmit to humans — are of particular concern, since it can be challenging to determine the source of the contamination. The use of antibiotics in livestock increases the resistance of some pathogens to traditional therapeutic techniques to treat the disease in humans [59]. It behooves public health officials and natural resource managers alike to mitigate the cause of fecal contamination for the safey and health of the public.

2.1.2 Microbial Source Tracking with Fecal Indicator Bacteria

Microbial Source Tracking (MST) aims to discover the host species of microbial lifeforms, often employing Fecal Indicator Bacteria (FIB) to discover the source of fecal contamination. Many techniques exist that leverage unique characteristics of the FIB present in the fecal matter. Ultimately, choosing the right FIB requires that researchers and investigators understand their resource and budget constraints, and tailor their MST process accordingly.

MST techniques using FIB generally fall into two broad categories: library-dependent and library-independent. Certain FIB and bacteriophages are known to be from specific host species, making host species assessment of fecal matter possible without the need to build a library. However, most library-independent MST cannot establish a host species source "beyond humans and a few domestic species" [59].

Library-dependent techniques build a library of known host species samples of FIB and use this data to determine host species for an unknown sample. The effort needed to build such a library can be considerable, often requiring the genetic sequencing of a considerate number of samples to be effective [59]. Commonly, these libraries are only useful for the environment in which the researchers sampled from. What they gain by building such a library is the flexibility to determine the source of fecal matter from multiple host species [59]. These library-dependent techniques rely on the concept of bacterial strain typing, discussed in the following section.

2.1.3 Delineating Bacterial Strains

When using FIB for MST, distinguishing between bacterial strains (BS) is a dynamic technique that can allow researchers to determine the source of fecal matter from a variety of host species. Other methods that rely on MST are usually limited in the host species they are effective at sourcing. By building a library of FIB samples, investigators can leverage bacterial strain typing to create a flexible and widely applicable method of sourcing fecal matter. Nevertheless, it has some drawbacks that, with some effort, MST investigators and researchers can overcome.

Bacterial strain typing for MST relies on the notion that strains of FIB remain mostly unique to the host species from which they came. A *strain* of a species of microbe is a subtype of that species where the microbes in that strain are closely related to each other in some meaningful way. Generally, how one defines a strain differs between research groups and application problem domains, as each definition of a strain depends on the characterization of the microbe in question and the methods used to derive such characterizations. A strain, then, can be thought of as consisting of individuals descending from the same individual to generate a "group" or "family."

Escherichia coli (*E. coli*) inhabits the gut of many animals, but only certain strains are pathogenic. The guts of animals and humans are a great environment for flora like *E. coli* to thrive and multiply. *E. coli* can get into the gut in many ways based on the environment the host animal lives in, the food they eat, and the other animals they interact with. Since *E. coli* resides in so many different species, researchers typically use it as the FIB of choice for library-based MST.

As [59] points out, the strains in *E. coli* can change considerably within the same host species. Geography, time, rainfall, and habitat all can have an effect on the strains present in a particular host species and even host animal. The transient nature of some *E. coli* strains forces researchers to build a library that contains a considerable number of samples to encapsulate the full spread of strains in the relevant host species, adding to the cost and efforts needed to perform MST and usually limiting the environmental scope of the library built to the environment in which it was sampled from. The Cal Poly Library of Pyroprints attempts to remedy this by building a cheap yet reliable method of distinguishing between strains using pyroprints on the two *ITS* regions of *E. coli*.

2.2 The Cal Poly Library of Pyroprints

This section details the aspects of The Cal Poly Library of Pyroprints (CPLOP) relevant to microbial source tracking (MST). It explains the nature of pyroprints and the pyroprinting process, including what segments of *E. coli* DNA CPLOP researchers use and how they collect the isolates used in the process. Finally, it describes the steps necessary to properly compare two *E. coli* isolates for strain identification and MST and provides an overview of how CPLOP stores the pyroprint data to facilitate it.

2.2.1 Pyroprints

Pyroprints are the core data structure in CPLOP used to represent *E. coli* isolates. Using an inexpensive DNA sequencing technique called pyrosequencing, we can build a fingerprint[1] that allows us to effectively differentiate between *E. coli* strains. Building pyroprints requires careful use of the pyrosequencing process.

Pyrosequencing is a DNA sequencing technique appropriate for sequencing short DNA fragments (up to around 150-200 base pairs) [60]. A machine dispenses a predefined series of nucleotides and carefully measures the light output of the reaction. The amount of light emitted is directly proportional to the amount of the corresponding

[1] hence, the portmanteau "pyroprint"

nucleotide present in the DNA.

The output of the machine is a time series graph depicting the measured light before and after each dispensation, called a pyrogram. Previous work in [41, 65] helped us determine the optimal dispensation order and length and exactly which portions of this graph to use. Determining the optimal dispensation order and its length are crucial to effective sequencing: too few dispensations may not encode enough information, while too many may degrade the quality of the data.

A *pyroprint* is a vector representing the peak light values of the pyrosequencing of one of the *ITS* regions in the seven loci of the *E. coli* genome. As explained in Section 2.2.2, *ITS-1* and *ITS-2* offer keen insight into *E. coli* strains, since random variation can occur in them without affecting the survivability of the *E. coli* microbes. We pyroprint each *ITS* separately, building at least two pyroprints for each isolate: one for *ITS-1* and another for *ITS-2*.

2.2.2 Internal Transcribed Spacers

Choosing the proper region of DNA to fingerprint is crucial to effective strain delineation. Generally, when fingerprinting FIB, researchers avoid using regions that code for functional products and focus instead on non-coding regions of DNA, since variations within do no affect the survivability of the microbe. When differentiating between *E. coli* strains, researchers use the two *ITS* regions between three genes: *16S*, *23S*, *5S*.

The *16S*, *23S*, and *5S* ribosomal RNA operons (rDNA) are genes that help code for proteins in *E. coli* bacteria. Any changes to these regions may affect the rate or nature of protein synthesis and thus affect the survivability of the bacteria. As a result, we consider these regions to remain *conserved* across *E. coli* strains. Using these segments directly to differentiate between strains would be fruitless, since even wildly different *E. coli* strains will still have nearly identical copies of these three regions.

Between these three genes are two non-coding, and thus *unconserved*, regions, each called an internal transcribed spacer (*ITS*). Since *ITS* do not code for functional products, random variations occur in *ITS* regions that do not affect the survivability or reproducibility of the bacteria. Researchers frequently use *ITS* for strain delineation, due to their unconserved nature. Importantly, any offspring of a microbe inherit the *ITS-1* and *ITS-2* regions, allowing biologists to use them to differentiate strains [68]. The two *ITS* regions that bridge *16S-23S* and *23S-5S* we respectively refer to as *ITS-1* and *ITS-2*. Amplifying the *ITS* regions of DNA becomes a straightforward and inexpensive process, due to the highly conserved regions flanking each *ITS*. Primers can reliably attach to the rDNA immediately next to each *ITS*, because of their conserved nature, allowing for polymerase chain reaction (PCR) amplification of the *ITS-1* and *ITS-2* regions.

Applying PCR to an *E. coli* isolate requires awareness of the following observation, crucially affecting how we can interpret a fingerprint:

> The *ITS* regions of *E. coli* and the rDNA — referred to collectively as *loci* — repeat around the *E. coli* genome seven times.

Figure 2.1 depicts these seven loci and the relative position of the *ITS* regions between the rDNA. The primers used to attach to the rDNA attach to each of the seven instances, resulting in PCR amplification of all seven copies of *ITS-1* and *ITS-2*. What results from PCR is an amplified mixture of these seven unconserved *ITS* regions that we can use as a fingerprint for the isolate.

The repetition of the *ITS* regions makes pyroprints different from traditional pyrosequencing. Traditional pyrosequencing allows researchers to figure out the se-

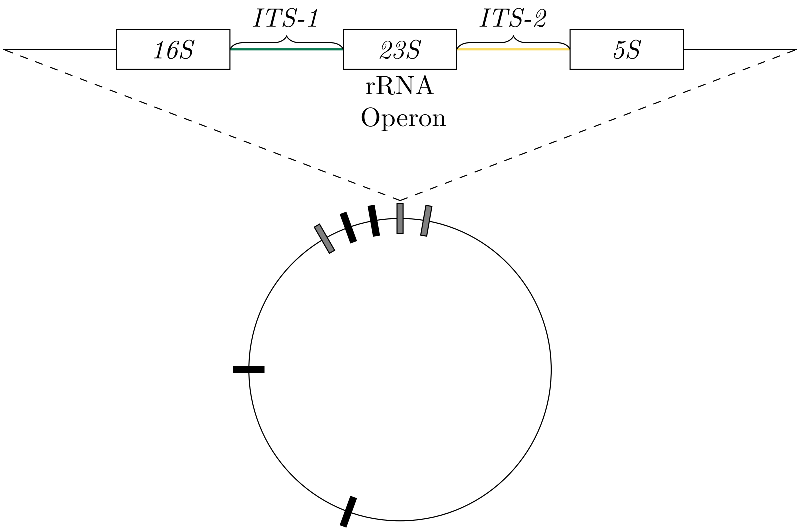


Figure 2.1: A diagram of a simplified segment of *E. coli* DNA, outlining the *ITS-1* and *ITS-2 ITS* regions, which repeat 7 times around the *E. coli* genome.

quence of nucleotides that make up the segment of DNA, because they only pyrosequence the PCR amplification of a single segment — or multiple conserved segments — at a time. Pyroprinting considers the PCR amplification of more than one unconserved segment of DNA at a time, encoding more information with a single pyroprint. As a result, CPLOP researchers cannot use a pyroprint to figure out the nucleotide sequence of the pyroprinted *ITS* region.

2.2.3 Obtaining & Comparing *Escherichia coli* Isolates

Cal Poly students collect *Escherichia coli* (*E. coli*) isolates from the fecal samples of a variety of different sources and compare them for a multitude of different studies. Culturing and *E. coli* extraction occurs in an introductory cell and molecular biology class. Comparing two isolates requires separate consideration of each *ITS* region.

The data stored in CPLOP are obtained as follows. Biologists collect fecal samples from a host-subject of a known species. They extract *E. coli* and culture individual bacterial cells from the bacterial material contained in each sample. A bacterial

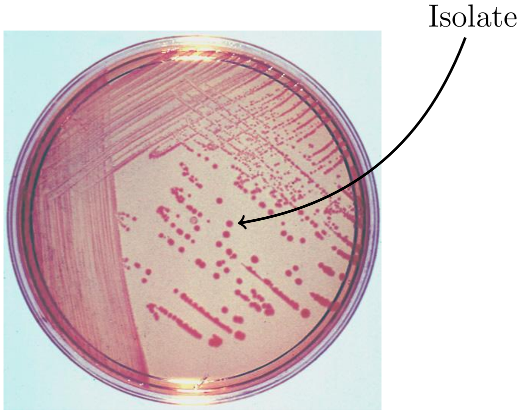


Figure 2.2: In order to obtain an isolate, researchers streak fecal matter onto this dish and make dots from the streaks that they then culture.

isolate is an individual culture grown from a fecal sample, as shown in Figure 2.2. Each isolate undergoes PCR to amplify the DNA in the two *ITS* regions of DNA, after which the pyrosequencing of each region produces pyroprints that are stored in the CPLOP database.

Sources of isolates in CPLOP are often animal species, but include many isolates cultured from environmental sources, like creeks and the ocean. A large portion of CPLOP consists of isolates derived from Cow and Human sources. The disproportionate number of Cows in CPLOP is due to a study investigating the strain demographics and transmission in cattle [17, 18]. Every year, Cal Poly houses cattle from around the state starting in May for testing and vaccination before they leave for auction in September. Cal Poly researchers obtained fecal sample from every cow as they arrived and when they departed, comparing the isolates for similarity before and after cohabitation. Isolates derived from Humans make up a large proportion of CPLOP because Cal Poly students investigated *E. coli* strain characteristics in a variety of studies [44, 46, 47]. Such disproportionate representation of host species in library-based MST is a common problem and Section 6.3 discusses the effect with respect to CPLOP.

The nature of separately pyroprinting the *ITS-1* and *ITS-2* regions of an isolate requires adherence to the following principle:

Principle 2.2.1 (Comparing Isolates)**.** The only valid comparison between isolates using the a comparison metric ρ is a separate comparison of the *ITS-1* pyroprints using ρ and *ITS-2* pyroprints using ρ: Given two isolates Isolate A and Isolate B, where Isolate A has pyroprints $\mathit{ITS\text{-}1}_A$ and $\mathit{ITS\text{-}2}_A$ and Isolate B has pyroprints $\mathit{ITS\text{-}1}_B$ and $\mathit{ITS\text{-}2}_B$, $\rho(\mathit{ITS\text{-}1}_A, \mathit{ITS\text{-}1}_B)$ and $\rho(\mathit{ITS\text{-}2}_A, \mathit{ITS\text{-}2}_B)$ are the only valid comparisons between the two isolates.

In other words, in order to compare two isolates, one must consider the pyroprints of each *ITS* region separately, because these pyroprints represent different regions of DNA. Figure 2.3 depicts the comparison process. Comparing an *ITS-1* pyroprint to an *ITS-2* pyroprint with ρ is completely meaningless, since the two pyroprints represent different segments of DNA in the *E. coli*.

2.2.4 Pearson Correlation Coefficient

Strain delineation underlies the goals of CPLOP — the ability to encode the concept of strains of the FIB *E. coli* and distinguish between different ones is fundamental to library-based MST. Towards that end, CPLOP researchers need some way to compare isolates using the pyroprints of their *ITS* regions that effectively distinguishes between different strains. Picking the right comparison function to compare the pyroprints of two isolates is crucial and we found that the Pearson Correlation Coefficient is an effective metric.

Given two D-dimensional vectors, $\vec{x} = (x_1, \ldots, x_D)$ and $\vec{y} = (y_1, \ldots, y_D)$, the

Pearson correlation coefficient ρ is:

$$\rho(\vec{x},\vec{y}) = \frac{1}{D}\sum_{i=1}^{D}\frac{(x_i-\mu_{\vec{x}})(y_i-\mu_{\vec{y}})}{\sigma_{\vec{x}}\cdot\sigma_{\vec{y}}} = \frac{cov(\vec{x},\vec{y})}{\sigma_{\vec{x}}\cdot\sigma_{\vec{y}}} \tag{2.1}$$

where $\mu_{\vec{x}}$, $\mu_{\vec{y}}$ are the means of x_i's and y_i's respectively, $\sigma_{\vec{x}}$, $\sigma_{\vec{y}}$ are their standard deviations, and $cov(\vec{x},\vec{y})$ is the covariance between the two vectors. The Pearson correlation coefficient encodes a notion of similarity: as the final portion of Equation 2.1 shows, ρ calculates the covariance of the two vectors, normalizing the value by the standard deviation of both.

The *covariance* of two vectors measures the joint variability of the vectors, i.e. how much one vector varies with respect to another. Given two D-dimensional vectors, $\vec{x} = (x_1,\dots,x_D)$ and $\vec{y} = (y_1,\dots,y_D)$:

$$cov(\vec{x},\vec{y}) = \frac{1}{D}\sum_{i=1}^{D}(x_i-\mu_{\vec{x}})(y_i-\mu_{\vec{y}}) \tag{2.2}$$

Positive covariance between two vectors means the two behave similarly, negative means they behave in the opposite manner, and zero means they behave independently of one another.

The *standard deviation* measures the amount of deviation a set of values has from its average. Given a D-dimensional vector $\vec{x} = (x_1,\dots,x_D)$, its standard deviation is:

$$\sigma_{\vec{x}} = \sqrt{\frac{1}{D}\sum_{i=1}^{D}(x_i-\mu_{\vec{x}})^2} \tag{2.3}$$

One can also define the standard deviation as the square root of the covariance:

$$\sigma_{\vec{x}} = \sqrt{cov(\vec{x},\vec{x})} \tag{2.4}$$

Using the Pearson correlation coefficient as a measure of similarity is straightforward, due to it being a combination of the covariance and the standard deviation. Since one can define the covariance of a vector with itself in terms of the standard deviation:

$$cov(\vec{x}, \vec{x}) = \sigma_{\vec{x}}^2 \tag{2.5}$$

it is clear that given two vectors, $\vec{x}$ and $\vec{x}\,'$, where $\vec{x} = \vec{x}\,'$:

$$\rho(\vec{x}, \vec{x}\,') = \frac{cov(\vec{x}, \vec{x}\,')}{\sigma_{\vec{x}}\sigma_{\vec{x}\,'}} = \frac{cov(\vec{x}, \vec{x})}{\sigma_{\vec{x}}\sigma_{\vec{x}}} = \frac{\sigma_{\vec{x}}^2}{\sigma_{\vec{x}}^2} = 1 \tag{2.6}$$

Due to the normalizing effect of the σ's, it is impossible for $\rho > 1$. Conversely, two very dissimilar vectors will have a $\rho < 1$. Specifically, Pearson correlation coefficient measures a linear correlation between vectors, so two vectors with a $\rho = 0$ means that the two vectors have no linear relationship[2]. Work done in [65] determined that multiple pyroprints of the same isolate obtain a $\rho > 0.995$ and CPLOP researchers use this value for quality control [9, 33].

It is from this measure of similarity ρ that we define a comparison metric between pyroprints. The nature of this function works perfectly for what pyroprints encode. The values in a pyroprint represent peak light intensity values from chemical reactions, the intensity of which is proportional to the nucleotide content of the DNA pyroprinted. Peak intensity differences and noise from the machine are accounted for by Pearson correlation coefficient, since it measures how the pyroprint vector values change with respect to each other and machine variations will be similar between pyroprintings.

Defining strains using the Pearson correlation coefficient requires a similarity threshold, above which we may consider two isolates to be part of the same strain. Two isolates are considered to be of the same strain if the pyroprints of both regions

[2] For completeness, if $\rho \approx -1$, then there is an *inverse correlation* between the two vectors. It is easy to see if given $\vec{x} = (x_1, \ldots, x_D)$ and $\vec{x}\,' = (x'_1, \ldots, x'_D) = (-x_1, \ldots, -x_D)$, then $cov(\vec{x}, \vec{x}\,') = -cov(\vec{x}, \vec{x}) \Rightarrow \rho(\vec{x}, \vec{x}\,') = -\rho(\vec{x}, \vec{x}) = -1$. By similar reasoning, $-1 \leq \rho \leq 1$

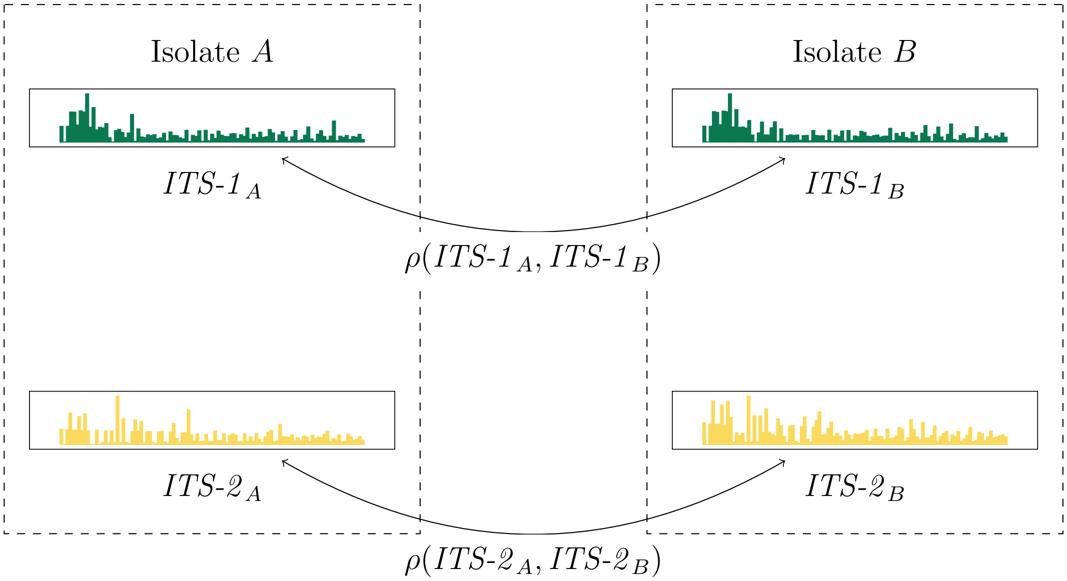
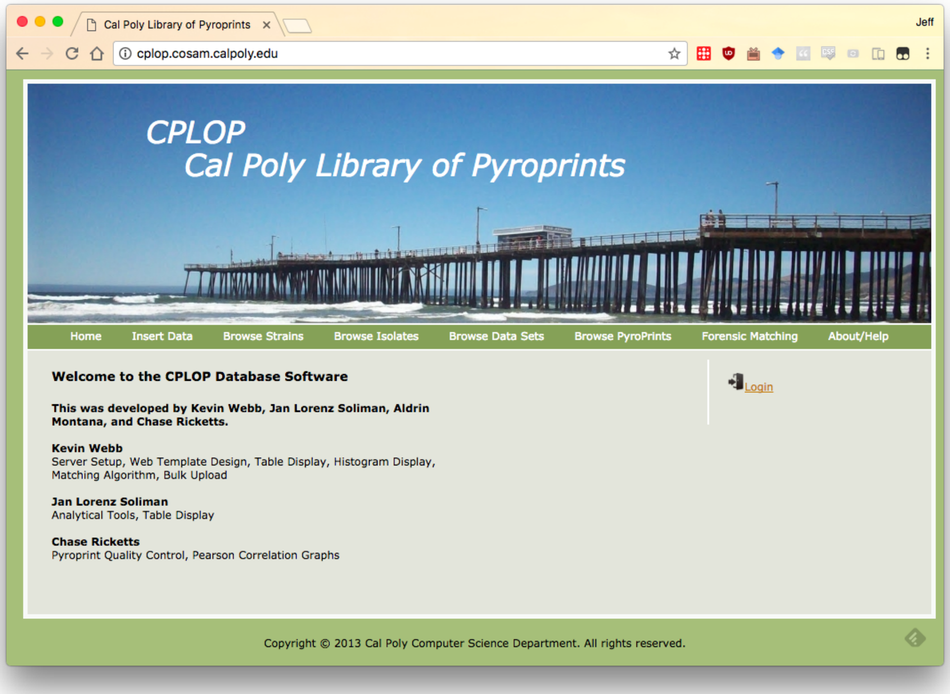


Figure 2.3: Comparing isolates involves comparing the pyroprints of each isolate using ρ, the Pearson correlation coefficient, with the stipulation that one can only compare pyroprints from the same *ITS* in their respective isolate. The green bar plots represent a pyroprint of *ITS-1*, while gold bar plots represent a pyroprint of *ITS-2*.

have a $> \alpha$, where $\alpha = 0.990$ [65, 9]. Work done in [65] determined this α threshold by simulating the pyroprinting process with tools from [41] on known *E. coli* strains from the National Center for Biotechnology Information database. Simulations performed in [10] further confirmed the usefulness of this α value.

2.2.5 Database

There are three components that comprise CPLOP: the physical cold storage of fecal samples and isolates, the back end data store, and the front end web interface. Cold storage allows CPLOP researchers to perform The back end data store holds the pyroprints and metadata about their host species and collection. The web interface, shown in Figure 2.4, allows researchers to perform queries and test whether isolates match. Cold storage holds the collected fecal samples and the isolates cultured from them. It allows CPLOP researchers to culture additional isolates and re-pyroprint



Welcome to the CPLOP Database Software

This was developed by Kevin Webb, Jan Lorenz Soliman, Aldrin Montana, and Chase Ricketts.

Kevin Webb
Server Setup, Web Template Design, Table Display, Histogram Display, Matching Algorithm, Bulk Upload

Jan Lorenz Soliman
Analytical Tools, Table Display

Chase Ricketts
Pyroprint Quality Control, Pearson Correlation Graphs

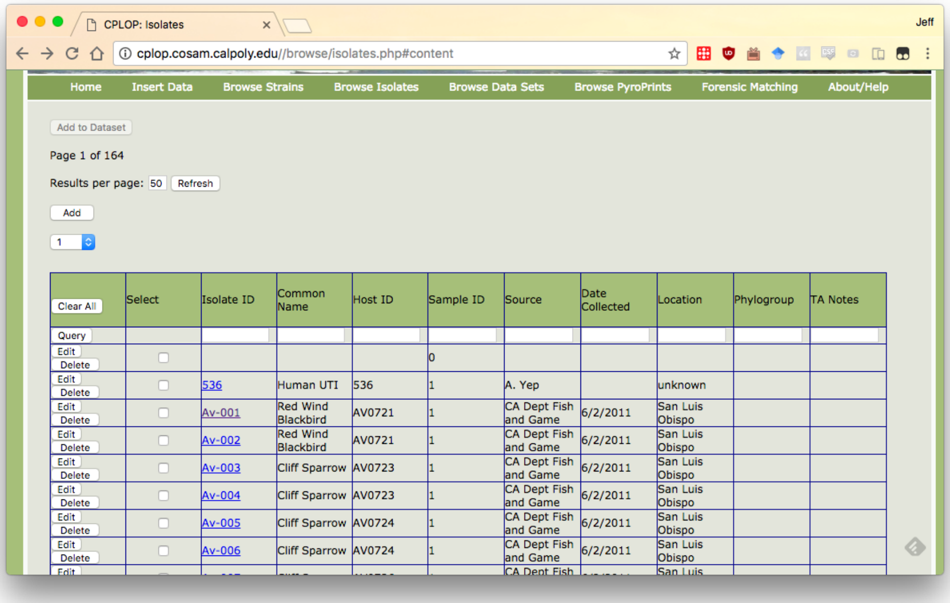
Copyright © 2013 Cal Poly Computer Science Department. All rights reserved.

Figure 2.4: Researchers use CPLOP through a web-accessible frontend.

existing isolates. Often, researchers refer to the cold storage as the "library" in library-based MST.

The data store for CPLOP pyroprints is a MySQL database. It stores metadata for each collected sample, including who collected it and where and when they collected it. Importantly, the name of the host species and a unique designation for the host animal marks the sample that the pyroprints of an isolate came from. For computationally intense procedures, researchers export the data and run it one a different machine.

The web front end, written in PHP, allows CPLOP to access the information in the database from the Internet to: perform queries for isolates and pyroprints, browse isolate and pyroprint datasets, and perform forensic matching. The isolates in CPLOP are visible from this interface (see Figure 2.5) as are the pyroprints. Isolates may have multiple pyroprints, which the CPLOP front end allows researchers to browse, which Figure 2.6 shows. Certain isolates come from particular collection runs, be



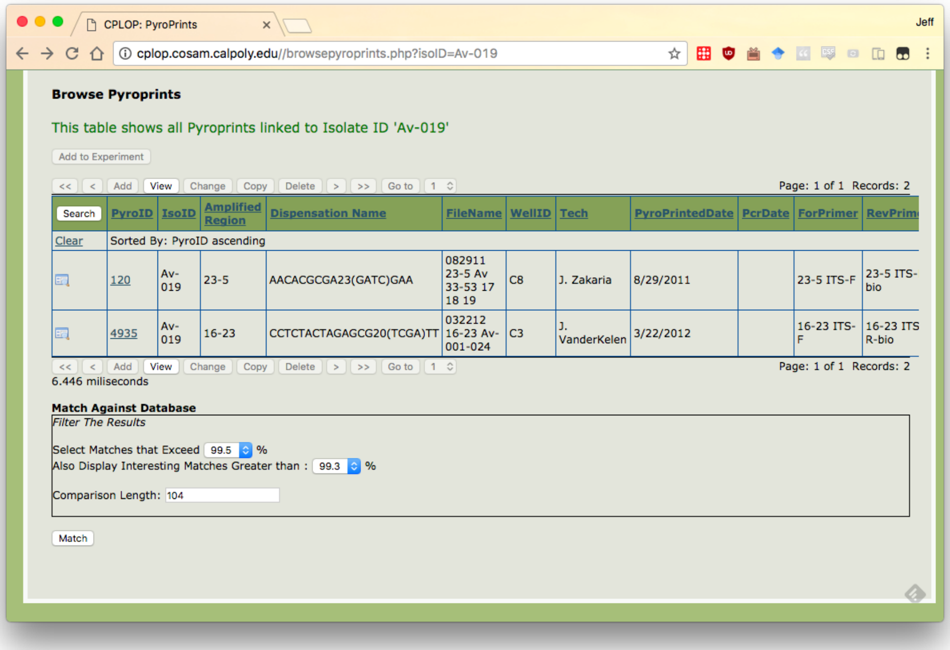
Page 1 of 164

Results per page: 50 Refresh

1

	Select	Isolate ID	Common Name	Host ID	Sample ID	Source	Date Collected	Location	Phylogroup	TA Notes
Query										
Edit Delete	□				0					
Edit Delete	□	536	Human UTI	536	1	A. Yep		unknown		
Edit Delete	□	Av-001	Red Wind Blackbird	AV0721	1	CA Dept Fish and Game	6/2/2011	San Luis Obispo		
Edit Delete	□	Av-002	Red Wind Blackbird	AV0721	1	CA Dept Fish and Game	6/2/2011	San Luis Obispo		
Edit Delete	□	Av-003	Cliff Sparrow	AV0723	1	CA Dept Fish and Game	6/2/2011	San Luis Obispo		
Edit Delete	□	Av-004	Cliff Sparrow	AV0723	1	CA Dept Fish and Game	6/2/2011	San Luis Obispo		
Edit Delete	□	Av-005	Cliff Sparrow	AV0724	1	CA Dept Fish and Game	6/2/2011	San Luis Obispo		
Edit Delete	□	Av-006	Cliff Sparrow	AV0724	1	CA Dept Fish and Game	6/2/2011	San Luis Obispo		
Edit						CA Dept Fish		San Luis		

Figure 2.5: CPLOP allows researchers to explore its isolates



Browse Pyroprints

This table shows all Pyroprints linked to Isolate ID 'Av-019'

Page: 1 of 1 Records: 2

Search	PyroID	IsoID	Amplified Region	Dispensation Name	FileName	WellID	Tech	PyroPrintedDate	PcrDate	ForPrimer	RevPrim
Clear	Sorted By: PyroID ascending										
	120	Av-019	23-5	AACACGCGA23(GATC)GAA	082911 23-5 Av 33-53 17 18 19	C8	J. Zakaria	8/29/2011		23-5 ITS-F	23-5 ITS-bio
	4935	Av-019	16-23	CCTCTACTAGAGCG20(TCGA)TT	032212 16-23 Av-001-024	C3	J. VanderKelen	3/22/2012		16-23 ITS-F	16-23 ITS R-bio

Page: 1 of 1 Records: 2

6.446 miliseconds

Match Against Database

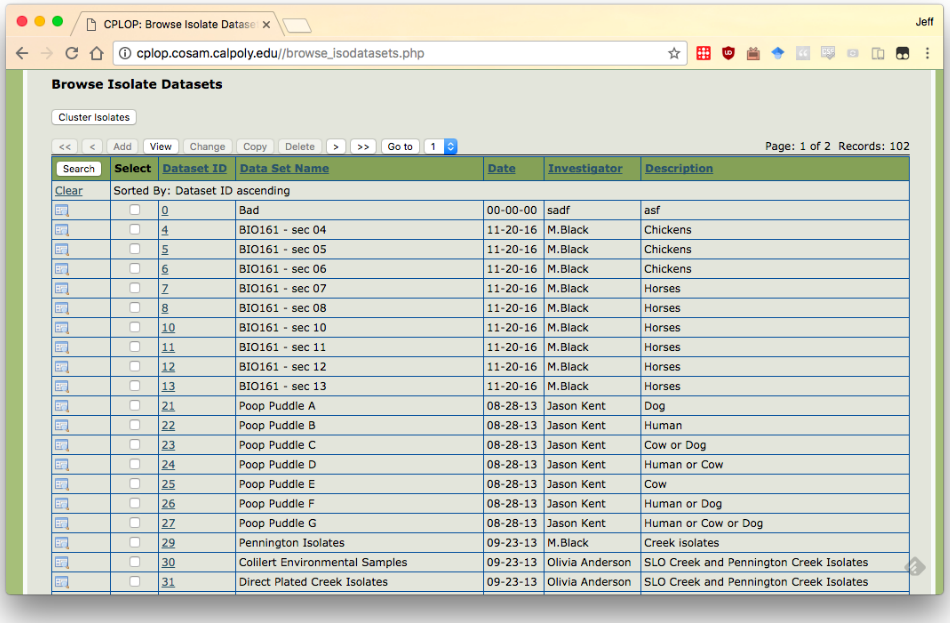
Filter The Results

Select Matches that Exceed 99.5 %

Also Display Interesting Matches Greater than : 99.3 %

Comparison Length: 104

Figure 2.6: The CPLOP frontend allows researchers to browse the pyroprints of an isolate.



CPLOP: Browse Isolate Datase
Jeff
cplop.cosam.calpoly.edu//browse_isodatasets.php

Browse Isolate Datasets

Cluster Isolates
<< | < | Add | View | Change | Copy | Delete | > | >> | Go to | 1

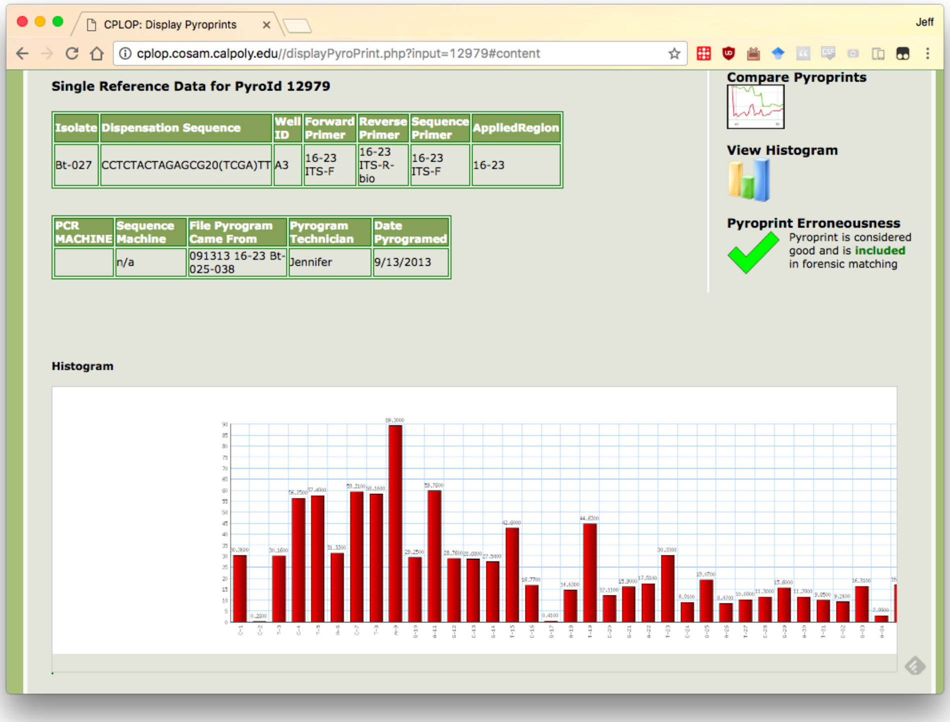
Page: 1 of 2 Records: 102

Search	Select	Dataset ID	Data Set Name	Date	Investigator	Description
Clear	Sorted By: Dataset ID ascending					
	□	0	Bad	00-00-00	sadf	asf
	□	4	BIO161 - sec 04	11-20-16	M.Black	Chickens
	□	5	BIO161 - sec 05	11-20-16	M.Black	Chickens
	□	6	BIO161 - sec 06	11-20-16	M.Black	Chickens
	□	7	BIO161 - sec 07	11-20-16	M.Black	Horses
	□	8	BIO161 - sec 08	11-20-16	M.Black	Horses
	□	10	BIO161 - sec 10	11-20-16	M.Black	Horses
	□	11	BIO161 - sec 11	11-20-16	M.Black	Horses
	□	12	BIO161 - sec 12	11-20-16	M.Black	Horses
	□	13	BIO161 - sec 13	11-20-16	M.Black	Horses
	□	21	Poop Puddle A	08-28-13	Jason Kent	Dog
	□	22	Poop Puddle B	08-28-13	Jason Kent	Human
	□	23	Poop Puddle C	08-28-13	Jason Kent	Cow or Dog
	□	24	Poop Puddle D	08-28-13	Jason Kent	Human or Cow
	□	25	Poop Puddle E	08-28-13	Jason Kent	Cow
	□	26	Poop Puddle F	08-28-13	Jason Kent	Human or Dog
	□	27	Poop Puddle G	08-28-13	Jason Kent	Human or Cow or Dog
	□	29	Pennington Isolates	09-23-13	M.Black	Creek isolates
	□	30	Colilert Environmental Samples	09-23-13	Olivia Anderson	SLO Creek and Pennington Creek Isolates
	□	31	Direct Plated Creek Isolates	09-23-13	Olivia Anderson	SLO Creek and Pennington Creek Isolates

Figure 2.7: Certain isolates may be part of a collection of isolates, which CPLOP has the ability to sort by.

they certain studies or classroom examples, and appear collectively as datasets on the website. CPLOP also provides the ability to view the pyroprint histogram, as Figure 2.8 shows. Forensic matching (Figure 2.9) is a crucial feature of CPLOP, allowing researchers to query a dataset against the CPLOP database to find matching isolates.

Cal Poly servers host the CPLOP website at `http://cplop.cosam.calpoly.edu/`. The servers are limited in computational ability, only containing 4GB of RAM. Such limitations make it difficult to implement algorithms like *OhClust!* [41, 42, 44], which require more than 4GB of RAM for efficient computation, or [2], which requires access to a cluster of computers for *MapReduce* ability. Future work will assess the feasibility of moving over to more dynamic systems, like Amazon Web Services.


CPLOP: Display Pyroprints
Jeff
cplop.cosam.calpoly.edu//displayPyroPrint.php?input=12979#content

Single Reference Data for PyroId 12979

Isolate	Dispensation Sequence	Well ID	Forward Primer	Reverse Primer	Sequence Primer	AppliedRegion
Bt-027	CCTCTACTAGAGCG20(TCGA)TT	A3	16-23 ITS-F	16-23 ITS-R-bio	16-23 ITS-F	16-23

PCR MACHINE	Sequence Machine	File Pyrogram Came From	Pyrogram Technician	Date Pyrogramed
	n/a	091313 16-23 Bt-025-038	Jennifer	9/13/2013

Compare Pyroprints

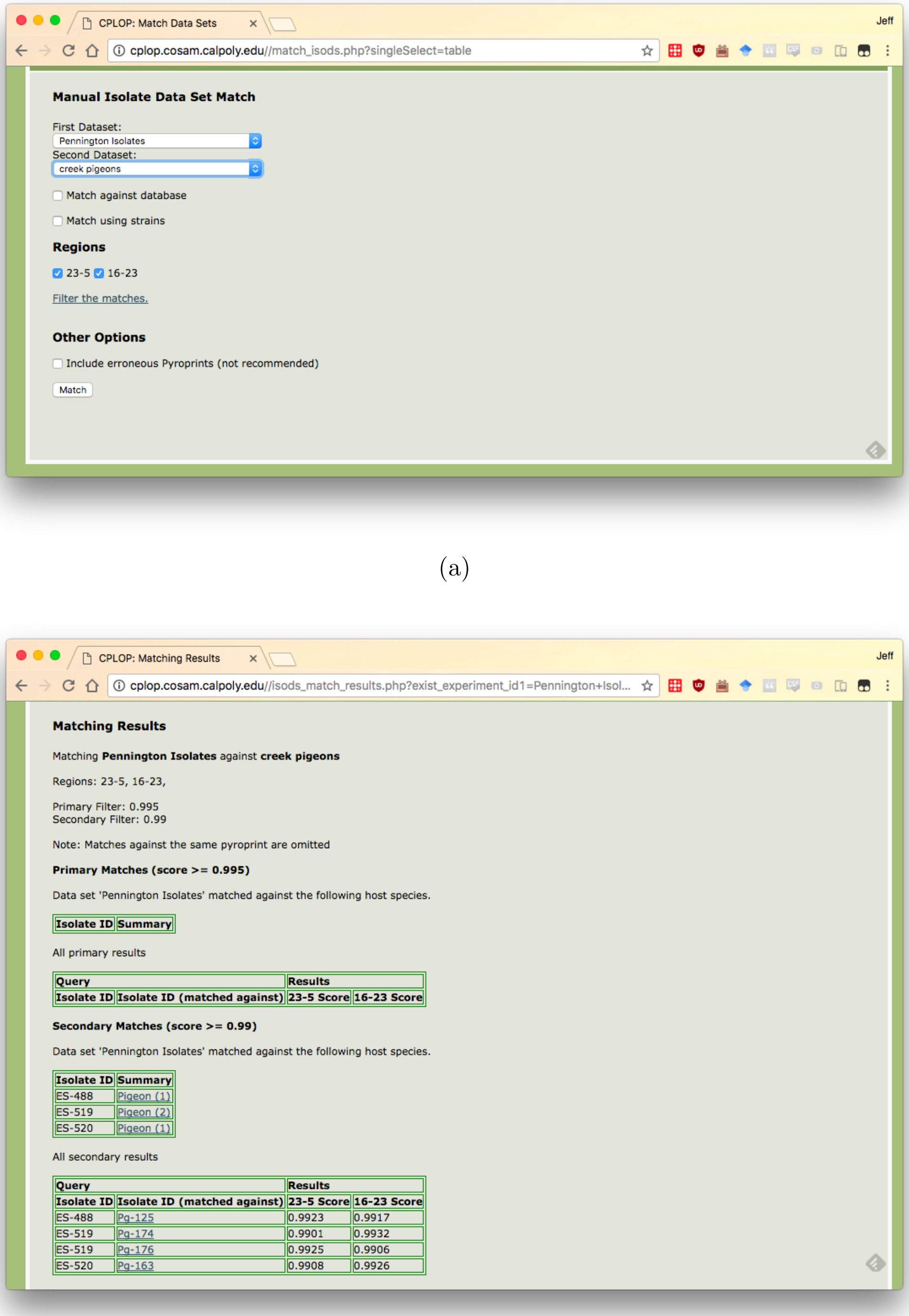
View Histogram

Pyroprint Erroneousness

✓ Pyroprint is considered good and is **included** in forensic matching

Histogram

Figure 2.8: Researchers can view the histogram of an individual pyroprint using CPLOP.



CPLOP: Match Data Sets
Jeff
cplop.cosam.calpoly.edu//match_isods.php?singleSelect=table

Manual Isolate Data Set Match

First Dataset: Pennington Isolates

Second Dataset: creek pigeons

□ Match against database

□ Match using strains

Regions

☑ 23-5 ☑ 16-23

Filter the matches.

Other Options

□ Include erroneous Pyroprints (not recommended)

(a)

CPLOP: Matching Results
Jeff
cplop.cosam.calpoly.edu//isods_match_results.php?exist_experiment_id1=Pennington+Isol...

Matching Results

Matching **Pennington Isolates** against **creek pigeons**

Regions: 23-5, 16-23,

Primary Filter: 0.995
Secondary Filter: 0.99

Note: Matches against the same pyroprint are omitted

Primary Matches (score >= 0.995)

Data set 'Pennington Isolates' matched against the following host species.

Isolate ID	Summary

All primary results

Query		Results	
Isolate ID	Isolate ID (matched against)	23-5 Score	16-23 Score

Secondary Matches (score >= 0.99)

Data set 'Pennington Isolates' matched against the following host species.

Isolate ID	Summary
ES-488	Pigeon (1)
ES-519	Pigeon (2)
ES-520	Pigeon (1)

All secondary results

Query		Results	
Isolate ID	Isolate ID (matched against)	23-5 Score	16-23 Score
ES-488	Pg-125	0.9923	0.9917
ES-519	Pg-174	0.9901	0.9932
ES-519	Pg-176	0.9925	0.9906
ES-520	Pg-163	0.9908	0.9926

(b)

Figure 2.9: Forensic matching is a key feature of CPLOP, allowing researchers to choose subsets of CPLOP data (a) to find strain-level matches (b).

2.2.6 CPLOP Makeup

Figure 2.10 shows the distribution of CPLOP isolates considered in this study among its 53 different host species.

Table 2.1: The number of isolates per host species.

Cow	1749
Human	1643
Ground Squirrel	196
Pigeon	196
Dog	179
Sheep	94
Wild Turkey	72
Pig	66
Horse	52
Cat	46
Chicken	44
Bat	37
Mountain Lion	32
Cliff Sparrow	28
Deer	20
White Crowned Sparrow	15
Opossum	12
Seagull	11
Sea Otter	10
Pelican	8
Bear	6

Host species	**Number of Isolates**
Owl	6
California Sea Lion	6
Red Tailed Hawk	5
Grey Fox	4
Coyote	4
Red-shoulder Hawk	4
Rabbit	4
Elephant Seal	4
Bobcat	4
Common Loon	4
Great Horned Owl	4
Racoon	4
Golden Eagle	3
American Kestrel	3
Mallard Duck	3
Deer Mouse	2
Tree Swallow	2
Red Wind Blackbird	2
Eared Grebe	2
Crow	2
Clark Grebe	2
Red Shoulder Hawk	2
Surf Scoter	2
Red Throated Loon	2
Western Kingbird	2

Host species	**Number of Isolates**
Turkey Vulture	2
Red-Winged Blackbird	2
Sea Lion	2
Guinni	2
Common Murre	2
Cougar	1
Orangutan	1

There are a total of 4,610 isolates in our dataset[3]. As seen from Figure 2.10, the organic growth of CPLOP yielded disproportionately many *E. coli* isolates originating from humans and cows (however, as shall be seen below, these isolates belong to a large number of strains). Each isolate is represented in CPLOP with two pyroprints — one each for *ITS-1* and *ITS-2* region. We ignore species with fewer than 4 instances, since their representation in CPLOP means they are reasonably unlikely to ever dominate a k-nearest neighbors list or cluster.

2.3 Computer Algorithms

Two computer science concepts are core to the clustering and classification techniques investigated in this thesis. One is a clustering algorithm that builds clusters by categorizing the datapoints into three different types, clustering some and denoting the unclustered as noise. Another is a classification technique that searches for nearby datapoints in order to classify an unknown datapoint.

[3] A simplified version of CPLOP containing isolate IDs, host species, and z-score normalizations can be found at `https://github.com/jmcgover/cplop-acm-bcb-2016`.

2.3.1 Density-Based Clustering of Isolates

Density-based clustering algorithms build clusters based on two parameters: the minimum number of neighbors `MinPts`, a point must have to be a core point of a cluster, and ε, the radius that those neighbors must be within. These algorithms define clusters with respect to core points and border points — points within ε of a core point — labeling everything else — the singletons — as noise. For this work, we chose to use DBSCAN as the clustering technique for grouping isolates because dense groupings of similar isolates fits our intuition of bacterial isolate strains. Closely related "families" of isolates will appear in the same cluster and we want these clusters to have sufficient purity to aid us in MST.

DBSCAN[21] provides the framework for our clustering algorithm. It uses a distance metric, a minimum neighbors value `MinPts`, and an ε range to categorize data points as one of three types: a core point, a border point, or noise. A *core point* is a point that has at least `MinPts` data points within ε of it. A *border point* is a point that is within ε of a core point, but that does not have `MinPts` points within ε of it. Every other point is *noise.* A *cluster* is a group of neighboring core points with their associated border points. According to this definition of a cluster, all clusters must have at least `MinPts` points in them. Figure 2.11 depicts this process.

Density-based clustering techniques require a distance metric, often times the Euclidean distance, between data points in order to cluster. Performing fast range queries greatly improves the speed of clustering. If the range query can finish in $O(\log n)$ time, then DBSCAN can run in $O(n \log n)$ time. Organizing the data into a spatial index can optimize these spatial queries.

Spatial indexes structure the data into a search tree, similar to a binary search tree, organizing the points by distance. When querying for nearby points, the algorithm can traverse this search tree, ignoring certain points along the way. While this can speed up the range query to a $O(\log n)$, many spatial indexes degenerate into a $O(n)$ operation. In the former case, this makes DBSCAN run in $O(n \log n)$ time.

In DBSCAN, the `RangeQuery` function handles range queries by taking as parameters the data point and a distance and returning all data points within range of the query point. Mathematically, a *range query* is a function $RangeQuery : D \times \mathbb{R} \rightarrow \{D\}$ that takes a query point $q \in D$ and a real-valued $\varepsilon \in \mathbb{R}$ and returns $\{d \in D | Dist(q, d)\} < \varepsilon\}$ — the set of all other points within ε of the query point — where $Dist$ is some distance metric $Dist : D \times D \rightarrow \mathbb{R}$. If the distance metric forms a proper metric space, then data structures like quad trees and octrees can speedup ε range queries for a point. One can imagine the range query as a hypersphere centered at the query point with a radius of the query range. In order to make `RangeQuery` fast, we had to make some optimizations.

2.3.2 k-Nearest Neighbors

The k-Nearest Neighbors classification algorithm (k-NN) is a straightforward algorithm to classify an unclassified object using a library. Using a comparison function, it compares the unclassified object to "nearby" classified objects. It uses the concept of a comparison function to formulate an idea of "closeness," asserting that the similarity an unknown object has to a class of objects relates to the class of the unknown objects itself. Figure 2.13

To outline the process: Given an unclassified object u, a library of classified objects $\mathbb{L}$, and a comparison function, $\mathbb{C}$:

1. Compare u to each object in $\mathbb{L}$ using $\mathbb{C}$
2. Add the classified object and the result to a list of neighbors, N
3. Sort N by most similar

4. Consider only the top k entries in N, called the k-nearest neighbors

5. Classify u as the *most plural* classification in the k-nearest neighbors list

Algorithm 1 describes this process in pseudocode.

Algorithm 1 k-Nearest Neighbors

Input:

- $u \in \mathbb{U}$: an unknown isolate
- $\mathbb{L} \subseteq \mathcal{I}$: a library of known isolates

Output:

- $\mathcal{S}$ value, classifying the unknown isolate u

Requires:

- k, α, and the comparison function C are predetermined

```
procedure ClassifyKNN($u, \mathbb{L}$)
    $N \leftarrow \emptyset$                  /* Make nearest neighbors list.      */
    for $p \in \mathbb{L}$ do                 /* For each library element:         */
        $sim \leftarrow C_i(u, p)$                  /* Compare.                */
        add $(sim, p)$ to $N$                       /* Add to neighbors.       */
    end for
    sort $N$ by $sim$                 /* Sort by most similar.             */
    $s \leftarrow$ FindMostPluralSpecies($\{n_1, \ldots, n_k \mid n_i \in N\}$)
    return $s$
end procedure
```

The motivation is that the unclassified object must be "close" to some of the classified objects in our database, using an appropriate measure of closeness — the *comparison function* — for the data. By choosing the *most plural* or *dominant* classification — the classification that shows up the highest number of times — in the k-nearest neighbors we can, with some accuracy, classify our unknown object.

Figure 2.13 depicts an example graph of datapoints in a coordinate space. All but one of the has a class associated with it, any of A, B, or C. The class of one point, denoted by ?, requires determination. Comparison functions like Euclidean distance or Manhattan distance may be the most appropriate way to compare these datapoints.

Figure 2.13 shows how using k-NN with Euclidean distance and various k can classify this point. Figures 2.13a, 2.13b, and 2.13c show the k-nearest neighbors lists as k changes from 4, to 6, to 9. At k=4, we see that k-NN classifies the unknown as A because there are 2 As, but only one each of B and C. For k=6, B is the resulting classification, since B shows up 3 times — more than the 2 A and 1 C. Finally, as we extend k all the way out to 9, k-NN classifies it as C, since such an extension exposes the k-nearest neighbors list to all 4 C's, more than any other classification.

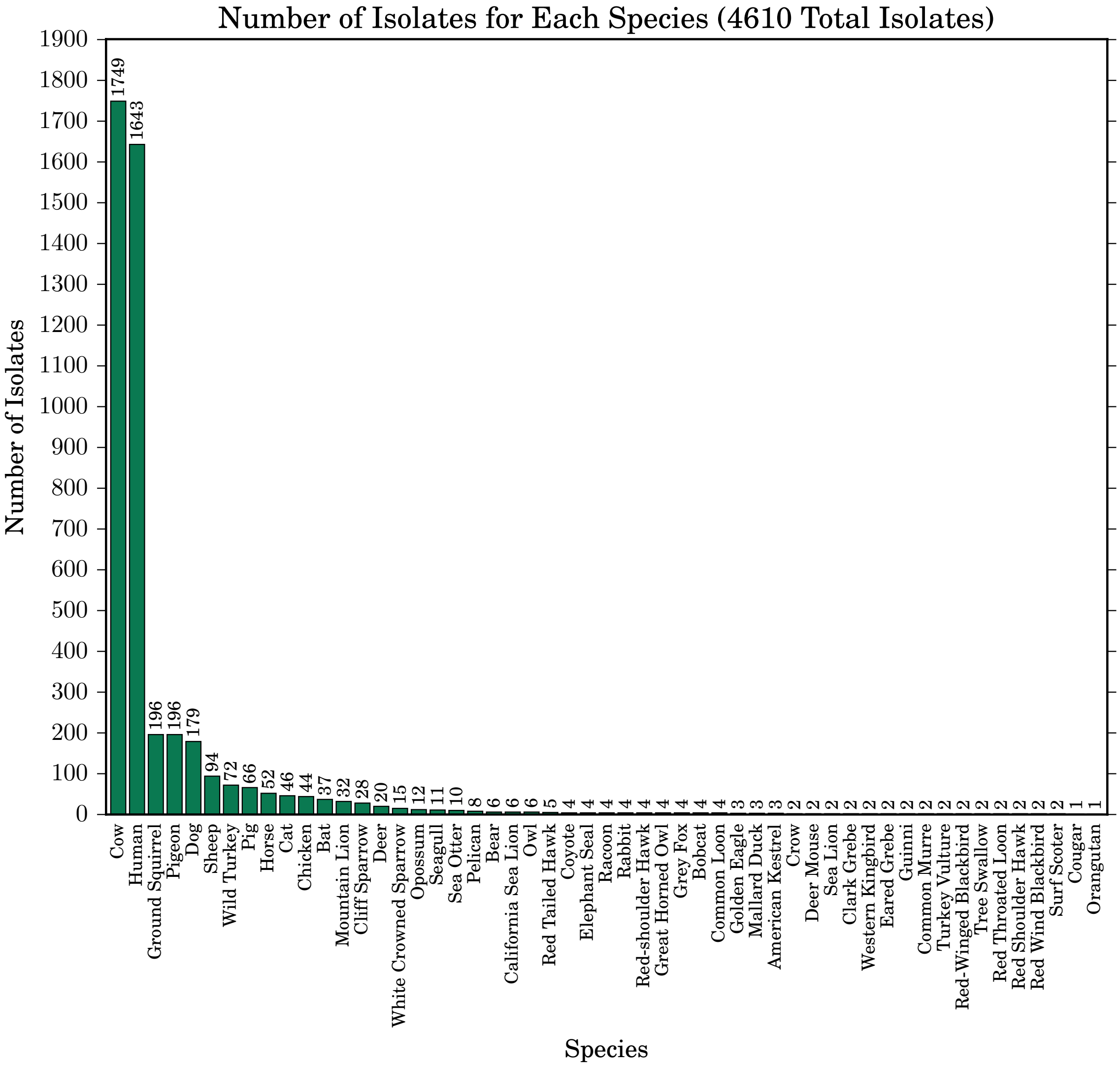


Figure 2.10: A histogram of the number of isolates of each species in our study, taken from CPLOP. There are 4,610 total isolates from 53 different host species.

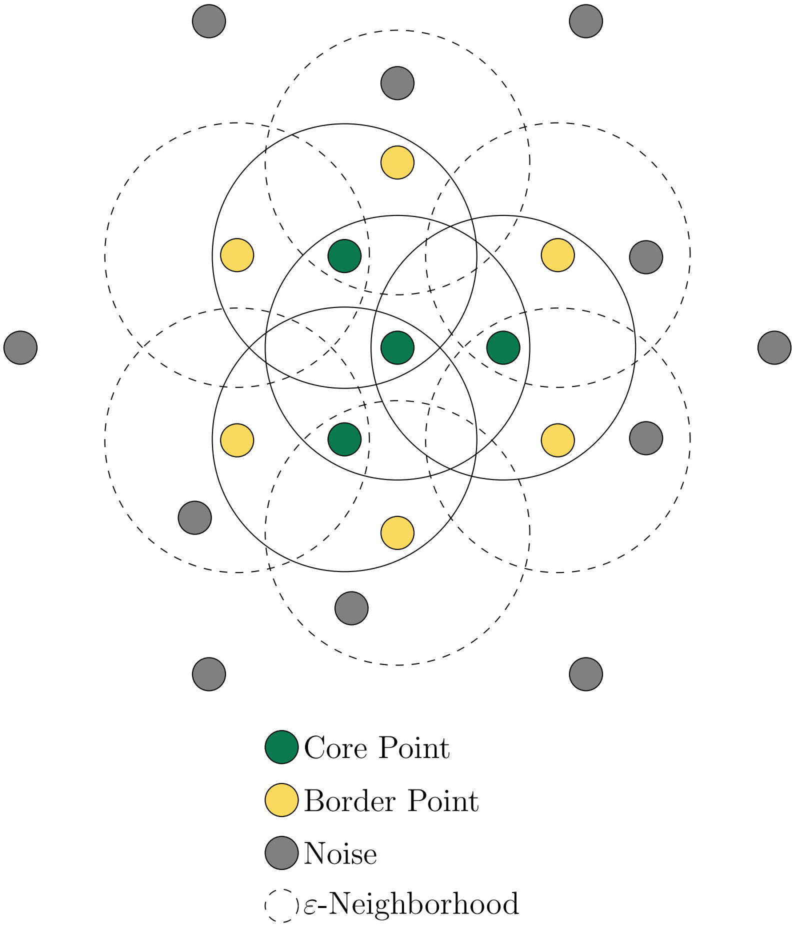


Figure 2.11: A basic density-based clustering with `MinPts` $= 3$ points and a unit ε represented by the circles — solid for the core neighborhoods and dashed for border (recreated from [31]). We see that the green points each contain 3 neighbors, but while the border points do not, they are within ε of a core point and we thus cluster it along with the core points. The (single) cluster that results from this set of datapoints are the green and gold points depicted.

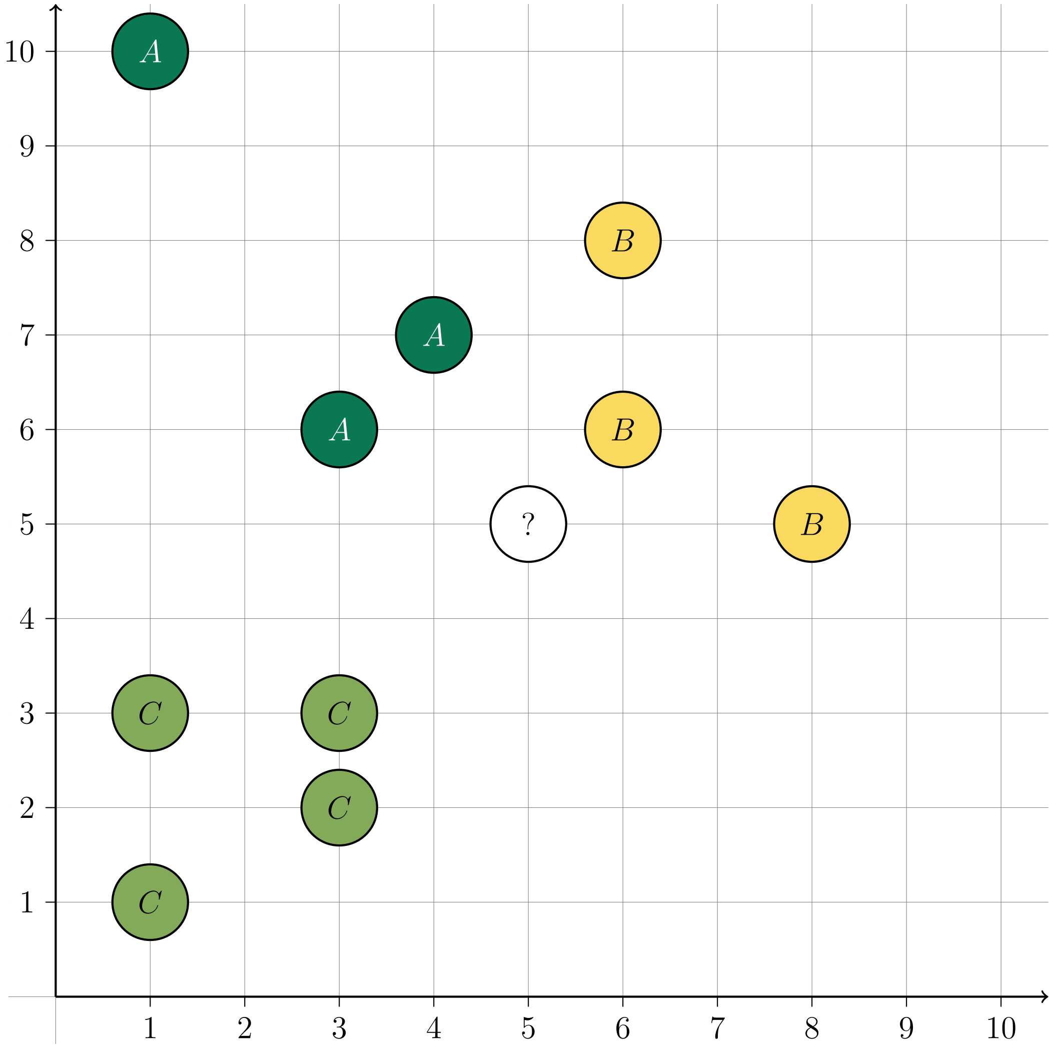


Figure 2.12: Example graph of datapoints in a coordinate space. All but one of the datapoints have a class associated with them, A, B, or C. The class of one point, denoted by '?', requires determination. Comparison functions like Euclidean distance or Manhattan distance may be the most appropriate way to compare these datapoints and Figure 2.13 shows how Euclidean distance and various k can classify this point.

k	Class	Position	Similarity
	?	(5,5)	0.000
1	B	(6,6)	1.414
2	A	(3,6)	2.236
3	A	(4,7)	2.236
4	C	(3,3)	2.828

(a) For $k = 4$, the classification for the unknown-class datapoint is A.

k	Class	Position	Similarity
	?	(5,5)	0.000
1	B	(6,6)	1.414
2	A	(3,6)	2.236
3	A	(4,7)	2.236
4	C	(3,3)	2.828
5	B	(8,5)	3.000
6	B	(6,8)	3.162

(b) For $k = 6$, the classification for the unknown-class datapoint is B.

k	Class	Position	Similarity
	?	(5,5)	0.000
1	B	(6,6)	1.414
2	A	(3,6)	2.236
3	A	(4,7)	2.236
4	C	(3,3)	2.828
5	B	(8,5)	3.000
6	B	(6,8)	3.162
7	C	(3,2)	3.606
8	C	(1,3)	4.472
9	C	(1,1)	5.657

(c) For $k = 9$, the classification for the unknown-class datapoint is C.

Figure 2.13: A simple k-Nearest Neighbors example, showing how the resultant classification of an unknown-class datapoint can change simply by adjusting the value of k. This example uses the data in Figure 2.12 with similarity values calculated by the Euclidean distance of the two points.

Chapter 3

RELATED WORK

Existing Microbial Source Tracking (MST) methodologies require a fecal indicator bacteria (FIB) fingerprinting method that allows for strain discrimination and a method of classification. Early MST methods [8] worked by measuring the ratio of fecal coliform bacteria to streptococci ratios. These methods fell out of use due to the “widely varying survival rates of the bacterial groups in the environment” [61]. In order to effectively use FIB for MST, researchers had to develop new methods to fingerprint and classify them with the appropriate host species or find related strains.

Fingerprinting FIB usually falls into two categories: phenotyping and genotyping. Phenotypic methods of fingerprinting usually involve “morphology of colonies on various culture media, biochemical tests, serology, killer toxin susceptibility, pathogenicity, and antibiotic susceptibility,” none of which allows researchers to reliably distinguish between closely related strains [36]. Genetic fingerprinting — genotyping — “has become widely used . . . due to its high resolution” [36] and many methods exist that allow for effective discrimination [61, 62].

Classification methods use a variety of statistical measures to make determinations to either related strains or host species, but most fall into library-dependent and library-independent. Library-independent MST searches for the presence of certain microbes in fecal matter or contaminated water. The presence of certain microbes can indicate what host species may have deposited the fecal matter. Unfortunately, this method relies on prior knowledge of the types of microbes that may occur in the types of potential host species[1], limiting the effectiveness of host species determination [61].

[1] Often times, these methods can only detect whether the fecal content came from a human and maybe some domestic animal species [61].

Library dependent techniques work by building a database of FIB fingerprints that come from known host species. These techniques usually differ in the fingerprinting process, which the classification technique depends upon. Using these libraries, researchers can handle common FIB from a variety of host species, making it incredibly agile. Disadvantages of this technique come from: the need to build a large library size which can become cost-prohibitive; the transient nature of some *E. coli* strains, assuming *E. coli* is the FIB of choice; and the fact that the applicability of the database is limited to the region in which the database was built from [61]. CPLOP is a library-based MST technique that uses *E. coli* as FIB and pyroprints as cost-effective fingerprints and researchers use it to understand *E. coli* strains and determine the host species of fecal matter.

3.1 Pyroprinting

A key component of strain-based Microbial Source Tracking is the representation of strains of fecal indicator bacteria (FIB). Numerous genotypic methods exist for differentiation between strains of *E. coli*. One can find a detailed discussion of why pyroprints perform better than these options in [33].

Cal Poly researchers introduce the concept and construction of pyroprints in [9, 33], describing the process through which they construct pyroprints from the multiple loci of isolated *E. coli* DNA. It discusses the work done in [65], which confirmed the reproducibility of pyroprinting and determined that a Pearson correlation coefficient correlation above 0.99 "could be a good threshold to minimize false separation of isolates from the same strain." These works explain in detail the advantages of using the pyroprinting methodology with respect to cost ("[p]yroprinting could reduce the cost of a library-based MST investigation by up to 60%" [9]), reproducibility (much of which can be found in [65]), and discrimination between (known) strains of *E. coli*, compared to existing state of the art methods. It asserts that while the researchers used *E. coli*, the pyroprinting process applies to a broad range of bacteria whose genome contain multiple loci.

In-silico simulations done in [10] delved into the sensitivity of using the Pearson correlation coefficient ρ to compare constructed pyroprints of known *E. coli* alleles gathered from the National Center for Biotechnology Information database. CUDA programming on a GPU sped up the ρ computation considerably, allowing the researchers to understand, given all possible combinations of seven known alleles to form a simulated isolate, how many isolates are "*hard to differentiate*," i.e. have a $\rho_{\textit{ITS-1}}$ and $\rho_{\textit{ITS-2}}$ above 0.99 [10]. The work in [10] supplements *in-vitro* work performed in [43].

Senior projects and master's theses [57], [67], and [73] discuss the development of many of the tools in CPLOP, from the backend database construction, to the frontend web view and usage for investigation. Cal Poly researchers have placed a large emphasis on validation of the methodologies included in building pyroprints from *E. coli* isolates. Biology students investigated how *E. coli* strains change in response to a variety of factors. Computer science students at Cal Poly have developed many tools to aid the biologists in both validating their methodologies and performing *E. coli* strain research on various host animals and host species

3.2 Empirical Strain Research

CPLOP has enabled numerous research projects in the field of biology. The following is a list of empirical strain research performed using CPLOP:

- Using Hadoop to Identify False Positives in Bacterial Strain Typing from DNA Fingerprints [2]

- Demographics and Transfer of *E. coli* Within Bos taurus Populations [17]
- *E. coli* Strain Demographics and Transmission in Cattle [18]
- Application of Pyroprinting for Source Tracking of E. coli in Pennington Creek [45]
- Demographics of *E. coli* Strains in the Human Gut Using Pyroprints: A Novel MST Method [46]
- *Escherichia coli* Strain Diversity in Humans: Effects of Sampling Effort and Methodology [47]
- Investigating the Dominant *Escherichia coli* Strain in Lambs and Ewes Using Pyroprinting: A Novel Method for Strain Identification [49]
- Source Tracking of Fecal Contamination Along San Luis Obispo (SLO) Creek [63]
- Short Communication: Typing and Tracking *Bacillaceae* in Raw Milk and Milk Powder Using Pyroprinting [71]

These studies provide significant insight into the evolution and transmission of *E. coli* strains and demonstrate the effectiveness of using pyroprints and CPLOP as a MST method. Many of the above studies provided a culminating experience for undergraduates and graduates in biology and computer science. What we aim to provide with k-RAP is a set of tools that students and researchers have at their disposal to make it easier to make reproducible discoveries and assertions about strains in CPLOP.

3.3 Clustering

Presented in [43, 44] is the comparison of two hierarchical clustering techniques. *Primer5* [12] and a chronology-sensitive hierarchical clustering algorithm. Using metadata about when researchers collected the samples used to build the isolates, the hierarchical clustering proceeds to first cluster isolates from samples collected on the same day and continues to cluster by increasing days away from the initial collection date. They found that the clusters built by the chronology-sensitive hierarchical clustering algorithm resembled the *Primer5* clusters, but were unsure of whether these clusters were appropriate.

The work in [43, 44] went on to become a part of *OhClust!* (**O**ntology-Based **h**ierarchical **Clust**ering**!**) [41, 42, 68], a metadata-aware hierarchical clustering algorithm that allows CPLOP researchers to provide a metadata ontology to guide the order of hierarchical clustering. Hierarchical clustering in general is a very calculation-intensive process, making it a problematic tool for servers with limited computational power. The computational crux comes with the number of comparisons needed between clusters — clusters of isolates in CPLOP's case.

Most hierarchical clustering algorithms compute the distances between clusters and agglomerate by picking clusters to combine into a cluster (made of clusters) for the next hierarchy. Cluster distances are merely the distance between some representative member — possibly an average of the actual members — of one cluster with a representative from the other cluster. The representatives used to compute distance may be the members in each cluster that are, for example, closest to each other, farthest from each other, or the centroid of each cluster.

Computationally intense distance metrics make implementing a performant hierarchical clustering algorithm problematic for programmers. The way *OhClust!* gets around this difficulty is by precomputing the distances — Pearson correlation coefficients in CPLOP — beforehand and storing them in memory. This greatly speeds up the clustering, but requires at least 4GB of memory for the distance lookup table alone. Since the servers that host CPLOP only have 4GB of RAM in total, directly incorporating *OhClust!* into CPLOP is not possible.

In [31], Eric Johnson presents a density-based clustering algorithm for pyroprints in order to build an intuitive clustering method that uses density and nearness and an efficient range query algorithm to find nearby isolates. DBSCAN [21], short for Density-Based Spatial Clustering of Applications with Noise, can be efficient if the distance metric used satisfies the triangle inequality. Unfortunately, Pearson correlation coefficient does not satisfy the triangle inequality, but work in [31] adjusts the comparison metric to use the Euclidean distance of z-score normalizations and optimizes further by organizing the pyroprints into a tree, making DBSCAN a viable method of clustering for the servers that host CPLOP.

An attempt to use DBSCAN for CPLOP isolates [31] as a naïve MST method in [38] revealed that for the isolates that actually clustered (i.e. were not determined to be noise), the accuracy was fairly high. Essentially, the MST method in [38] clusters an unknown-host species isolate along with the rest of the known-host species isolates in CPLOP and classifies it as the most plural host species in the resulting cluster. However, [31] clustered only about half of the isolates in CPLOP, while the rest remained unclustered and thus unclassified. The investigation in [38] was useful to confirm suspicions of so-called “transient” strains of *E. coli* bacteria. Section 2.3.1 discusses the details of [31] relevant to this thesis, Section 5.1 describes a methodology to use it as a MST method, and Section 5.3 discusses the investigation in [38].

The clustering methods presented in [41, 42, 43, 44, 68] and [31] are examples of typical investigations into bacterial strain research. On their own, they do not constitute an actual MST methodology[2]. Ultimately, the goal of CPLOP is to be able to objectively classify the host species of an *E. coli* isolate. Thus, merely clustering isolates is insufficient for MST. This thesis presents k-RAP, an MST technique that works with the pyroprints of isolates in CPLOP as a solution to MST.

3.4 k-NN Techniques

A plethora of k-Nearest Neighbors (k-NN)[3] methods exist, but most are various attempts to optimize the search space — either with efficient range queries or by leveraging information about the space to improve search speed — or modifications to the neighbor list structure to improve classification. Surveys on k-NN techniques [7, 30] show that each variation builds data structures for efficient query, or abstracts the notion of the usually Euclidean distance metric to build a more accurate classifier, while others may weight the neighbors or remove neighbors from consideration based off of some criteria. An exception to the typically euclidean distance metric comes in the way of recommender systems [15, 50], which use a notion of similarity based off of scores. While efficient range query interests us, we have a solution for it in [31].

The method that most closely resembles what we are after comes from techniques that build multiple k-NN classifiers by generating feature subsets and polling the classifier to determine the class of the unknown datapoint [4, 5, 72]. Similar to bagging and bootstrapping techniques used to train other classification algorithms, the feature-set of known datapoints is either reasonably partitioned into feature-subsets [4, 5], or clustered into subsets [72]. Some even perturb the data and group features to create multiple k-NN classifiers [32]. The resulting classification from the k-NN subset is then aggregated and the final classification is determined by majority voting. This approach will not apply to CPLOP, since we do not merely have a single

[2] The work in [38] attempts to use clustering as a MST technique.

[3] See Section 2.3.2.

comparison metric that we want to partition into multiple to improve classification. Isolates in CPLOP always have two entirely separate metrics that we must make a reasonable decision from.

The primary goal of the k-NN Resolution Algorithms for Pyroprints (k-RAP) is to resolve the two comparison metrics that CPLOP has for comparing isolates. That is, given an isolate, in order to find nearby isolates, one must separately compute the Pearson correlation coefficient ρ for each ITS, giving us two comparison metrics, $\rho_{\mathit{ITS\text{-}1}}$ and $\rho_{\mathit{ITS\text{-}2}}$. Typically, the vectors in k-NN techniques represent the entire set of features for a particular datapoint. Many k-NN algorithms assume that the distance metric used — usually Euclidean distance — will encode a useful notion of distance.

k-RAP can apply to other datasets with separate comparison metrics, especially those that contain types of features that Euclidean distance does not apply to. For example, in a demographic study for, say, a political study, subjects may have a multitude of features with different metrics of comparison. Location of residence may be one and favorite color another[4], with a goal of classifying a subject's political party. Euclidean distance may not be appropriate for the location metric, since great-circle distance on a globe may encode closeness more accurately. For color, while it may be straightforward to represent red, green, and blue values as a vector, Euclidean distance may not be the best choice to gauge similarity in color, certainly not in the same way as the great-circle distance, especially for the reasons put forth in [39], which discusses the tremendous difficulties in building a uniform perceptive color space. Simply combining these two features into a single vector and performing Euclidean distance may not produce the most appropriate results. Nevertheless, these metrics on their own are perfectly amenable to their own, accordant distance metric that cannot necessarily be used on other features, making it easy to create a k-NN

[4] Certainly, many other demographic and psychological metrics can exist, but for simplicity's sake, let us consider only these two.

for each feature separately. As such, when using k-NN on datasets with a complex set of features there is a need for the ability to resolve separate k-NN lists in order to usefully classify datapoints.

Chapter 4

MICROBIAL SOURCE TRACKING METHODOLOGY

An effective, automated microbial source tracking (MST) methodology is key to CPLOP's success as a tool to aid fecal contamination investigators. Until recently, Cal Poly researchers performed most MST by hand [45, 63]. The goal for any MST method, when used to source fecal contamination, is to take fecal matter or a substance contaminated with fecal matter and determine, or classify, the host species that provided the fecal matter; library-dependent MST methods leverage the known-host species information stored in their library, usually digital representations of fecal indicator bacteria (FIB) stored in a database. CPLOP is such a library-dependent technique that aims to support MST using pyroprints of both *ITS* regions of the FIB *E. coli*. Given a FIB isolate from an unknown-host species[1], a library-dependent MST technique determines the host species of the unknown using the information in the library. Towards this end, we built and investigated two MST techniques, one that approaches host species classification from the perspective of strains in the database and another that directly uses isolates present in the database. This Chapter outlines the abstract approaches we chose to take, while Chapter 5 and Chapter 6 detail the specific techniques we used for the strain-based and isolate-based approaches respectively.

4.1 Strain-Based

Strain typing is central to library-dependent MST methods and building a host species classification technique that uses strains directly is an intuitive approach to take. If an unknown-host species matches a strain in the library, then we can make

[1] often referred to as the "the unknown" or "the unknown isolate"

a reasonable assertion as to its host species if the strain has a dominant host species. Thus, given an unknown-host species isolate:

1. Incorporate the unknown-host species FIB isolate into CPLOP
2. Build strains of FIB using the isolates in CPLOP
3. Classify the source host species of the isolate as the dominant host species of the strain it ended up in

Strain construction can happen in many ways, but from a computational perspective, it is very amenable to clustering. If one can imagine a coordinate space that encapsulates mathematical representations (vectors) of FIB isolates, then strains are the close groupings (clusters) of these isolate representations. For CPLOP, this means constructing strains from the pyroprints of collected *E. coli* isolates.

As explained in Chapter 5, we use DBSCAN, a density-based clustering algorithm that we introduce in Section 2.3.1, which builds clusters using a similarity metric, allowing for the concept of noise — datapoints that remain unclustered. Dense groupings of similar isolates fits our intuition of bacterial isolate strains because closely related "families" of isolates will appear in the same cluster. A primary limitation of DBSCAN is that it may not cluster some isolates, which still aligns with our notion of strains; sometimes, isolates are not part of any strain present in CPLOP. As a result, we need a fallback MST method that can work for every isolate.

4.2 Isolate-Based

Using isolates directly can give us a level of flexibility that strict strain typing may not allow. An unknown-host species may not fit within the sometimes strict definition of a strain that a particular library may have, but we may still be able to make a reasonable assertion as to its host species based on the isolates that are most similar to the unknown. Thus, given an unknown-host species isolate:

1. Find k known-host species isolates from CPLOP most similar to it, called the k-nearest neighbors

2. Classify the source host species of the isolate as the dominant host species of the k-nearest neighbors

The described approach is the k-Nearest Neighbors classification algorithm (k-NN) introduced in Section 2.3.2, however, due to the multiple comparison functions needed between isolates, k-NN, in its natural form, will not work with CPLOP. Section 2.2.3 details why the comparison of two isolates to each other requires multiple comparison functions. Chapter 6 explains the comparison function resolution strategy employed by the k-Nearest Neighbors Resolution Algorithms for Pyroprints, the isolate-based MST method we built and investigated.

Chapter 5

CLUSTERING FOR BACTERIAL STRAINS

5.1 Methodology

In order to understand the make up of bacterial strains in CPLOP, we chose to investigate how a density-based clustering algorithm might group the isolates we have collected so far and how that might affect a rudimentary MST technique based off of clustering: cluster an unknown isolate along with the isolates in CPLOP and classify it as the most plural species of the cluster. Density-based clustering ties in well with our notion of closely-related strains of *E. coli* (the relation being the separate Pearson correlation coefficient comparison of each *ITS* region we use to compare isolates). From the computer science point of view, a bacterial strain is essentially a cluster of *E. coli* isolate representations stored in CPLOP. Our MST method, thus, works as follows:

1. **Strain Identification.** Identify bacterial strains in CPLOP by clustering all CPLOP isolates.

2. **MST.** Given an isolate of unknown origin, find the cluster it belongs to. Return the host species of the plurality of isolates in the cluster.

Our clustering algorithm is the density-based clustering algorithm developed by Johnson [31]. It extends DBSCAN for the case of two comparison functions between data points (our isolates are compared based on the two *ITS* regions) and implements an efficient spatial data structure to manage the retrieval of the data points.

DBSCAN can easily use an efficient range query technique to find nearby points and speed up clustering time considerably by taking advantage of the triangle inequality that proper metric spaces have. Unfortunately, Pearson correlation coefficient does not encode a metric space, because it fails the triangle inequality[1]. This complicates range queries, discussed in Section 2.3.1, because spatial indexes tend to rely on the triangle inequality, usually with Euclidean distance, to argue that certain points can be ignored during a spatial index tree traversal.

To allow us the use of a spatial data structure with the Pearson correlation coefficient to store data points during the clustering procedure we use instead the Euclidean distance pyroprint z-score normalizations, which we derive by recognizing in Equation 2.1 that Pearson correlation coefficient is made up of z-score normalizations of $\vec{x}$ and $\vec{y}$. The z-score normalization of $\vec{x}$ is:

$$z(x_i) = \frac{x_i - \mu_{\vec{x}}}{\sigma_{\vec{x}}}$$

where $\mu_{\vec{x}}$ and $\sigma_{\vec{x}}$ are the mean and standard deviation of the values in a single pyroprint respectively. Thus, for clustering, we compare pyroprints using the Euclidean distance d of z-scores.

$$d(\vec{z}_{\vec{x}}, \vec{z}_{\vec{y}}) = \sqrt{\sum_{i=1}^{D}(z(x_i) - z(y_i))^2} \tag{5.1}$$

where D is the number of dimensions. This allows us to use spatial indexes and $O(\log n)$ lookup in DBSCAN.

Each isolate is represented in CPLOP by a pair of pyroprints: one each from each *ITS-1* and *ITS-2* region, complicating the use of DBSCAN and the meaning of α threshold. We handle this in DBSCAN by performing two range queries, one each for *ITS-1* and *ITS-2*, taking the intersection of the two results. We must, however, pick a suitable ε for each *ITS* region.

[1] $d(x, z) \leq d(x, y) + d(y, z)$

Table 5.1: Converted α threshold to fit the new metric space defined by (5.1).

***ITS* Region**	*ITS-1* α	*ITS-2* α
$\rho(\vec{x}, \vec{y})$	0.995	0.995
D	96	94
$d(\vec{z}_{\vec{x}}, \vec{z}_{\vec{y}})$	0.9747	0.9644

CPLOP uses a threshold value of $\alpha = 0.995$ to compare two pyroprints. Pyroprints with Pearson correlation coefficient above α are considered to represent the same DNA material, while pyroprints with a Pearson correlation coefficient below α are considered to represent different DNA material [65, 67, 68].

The number of dispensations D used to build a pyroprint differ for the *ITS-1* and *ITS-2* regions. Because $D_{\textit{ITS-1}} \neq D_{\textit{ITS-2}}$, the original α under the space defined by (5.1) no longer applies in the same way to both regions. An alternative formulation of (5.1), with respect to the Pearson correlation coefficient ρ, is:

$$d(\vec{z}_{\vec{x}}, \vec{z}_{\vec{y}}) = \sqrt{2 \cdot D \cdot \rho(\vec{x}, \vec{y})} \tag{5.2}$$

where D is the number of dimensions and $D_{\vec{x}} = D_{\vec{y}} = D$. Using Equation 5.2, we can convert α to the values in Table 5.1. We use these converted α values as the ε for each *ITS* region's `RangeQuery`.

When clustering CPLOP isolates using our density-based clustering algorithm, we need to set up the two parameters at our disposal: `MinPts` and ε. For ε we choose the two values shown in Table 5.1 converted from the 0.995 Pearson correlation coefficient threshold of pyroprint similarity. Essentially, we only want to consider the ε-neighborhood of a pyroprint that contains the other pyroprints that we consider to represent the same DNA material.

For the `MinPts` parameter, we use *grid search* running our clustering with `MinPts` set to $1, 2, 3, 4, 5, 6$, and 7. The `MinPts` value adjusts how strict our definition of a cluster is. That is, the higher the value of `MinPts`, the more neighbors a core point must have with ε of it and its neighbors to become a cluster. Balancing this value with the coverage of our algorithm is crucial to its success, because for too low of a value, we may not have a clear plurality in a cluster, while too high of a value may miss some smaller clusters that might classify our unknown isolate into something other than noise.

5.2 Evaluation

Clustering for bacterial strains has two aspects that we must evaluate: how pure the bacterial strains (clusters) are and how many of the isolates in CPLOP end up in a cluster (as opposed to noise). The former tests whether the *E. coli* strains stay relatively unique to the host species from which they come from, the core theory of library-based MST. The latter tells us how effective Section 5.1 describes how we can use clustering as a MST method — cluster an unknown isolate along with the isolates in CPLOP and classify it as the most plural species of the cluster. Cluster purity can easily gauge how effective this technique is at MST, since the concept readily summarizes to how pure, from a host species perspective, a cluster is. Density-based clustering algorithms such as DBSCAN may not cluster every datapoint, as mentioned in Section 2.3.1, labeling some datapoints (isolates) as noise and thus leaving them unclustered.

5.2.1 Cluster and Clustering Purity

In this paper we look at the results of clustering CPLOP data using this algorithm from the perspective of cluster purity. We call a cluster (bacterial strain) *100% pure* if all isolates that belong to it come from the same host species.

Of interest to us is the following information:

1. The number of 100% pure clusters and the percentage of bacterial isolates from CPLOP clustered into pure clusters.
2. The structure of impure clusters: specifically, whether a dominant host species can be clearly identified in each cluster.
3. Coverage: the total number of CPLOP isolates found to belong to a strain.
4. MST Accuracy: the percentage of isolates for which the strain-based MST procedure produces the correct response.

Thus, our core measure is *cluster purity*, the proportion of a cluster that comes from the most plural host species of that particular cluster. A *100% pure cluster* is a cluster which only contains data points (isolates) with the same class label (same host species of origin).

Consider a cluster $C = \{c_1, \ldots, c_K\}$. Let $s(c)$ refer to the species of isolate c. Let m be the plurality species label for data points in C, and let the total number of points in C with $s(c) = m$ be s_m. Then the *individual cluster purity* ν of cluster C is:

$$\nu(C) = \frac{s_m}{K}$$

In addition to computing the purity of individual clusters we want to have an understanding of the overall purity on the entire dataset. Given a *clustering* $\mathcal{C} = \{C_1, \ldots, C_n\}$ on a dataset, we define the size $\mathcal{M}$ of the set of clusters:

$$\mathcal{M} = \sum_{i=1}^{n} |C_i| \tag{5.3}$$

The *overall clustering purity* is:

$$\sum_{i=1}^{n} \frac{|C_i|}{\mathcal{M}} \cdot \nu(C_i) \tag{5.4}$$

One can think of (5.4) as a form of weighted arithmetic mean of the purities, where the size of the cluster adds more weight to the value.

5.2.2 Clustering Coverage

Coverage of the dataset is important to an effective MST method. The density-based clustering method we use has one key disadvantage: a clustering run with the parameter `MinPts`, treats all points that do not fit into a cluster of size of at least `MinPts` as noise. This means that as the value of `MinPts` grows, so will the number of isolates that do not cluster into a strain.

Given the parameter `MinPts` of the clustering algorithm, we collect the following four measures, that collectively represent the breakdown of all data points (isolates) in CPLOP:

1. *Noise.* Number/percentage of isolates clustered as noise points.
2. *Misses.* Number/percentage of isolates from minority species in impure clusters.
3. *Hits.* Number/percentage of isolates from plurality species in impure clusters.
4. *Pure points.* Number/percentage of isolates in 100% pure clusters.

5.3 Results

In gauging how effective our clustering method is against CPLOP, we looked at the distribution of cluster sizes, the number of isolates that fell into high purity clusters, the number of unique species in each cluster and how that affected the size and purity, and overall coverage and accuracy metrics. From these data, we gained some insight into the clustering algorithm and were able to visualize some predictions we had about the biological aspects of strains.

5.3.1 Cluster Size Distribution

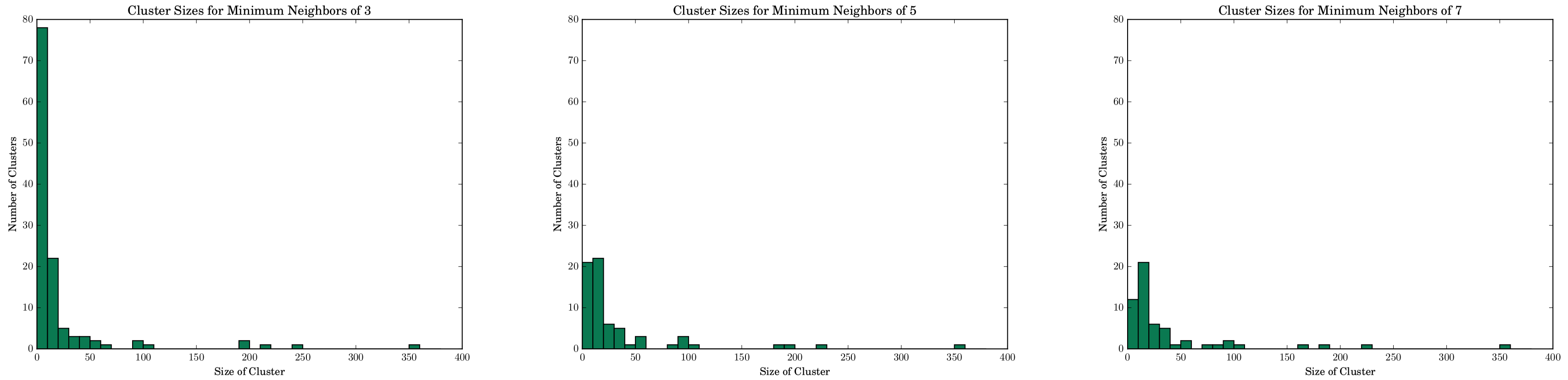
Figure 5.1 shows the distribution of cluster sizes as `MinPts` increase from 3, to 5, to 7. We see that at all three `MinPts` values, the number of small clusters (fewer than 10) dominates the overall makeup of clusters. Figure 5.1 shows a propensity towards small clusters at low `MinPts` values. This creates a high number of 100% or almost 100% pure clusters. Most clusters are tiny, with a few larger clusters for small `MinPts` values.

As we approach higher `MinPts` values, the smaller clusters disappear. As `MinPts` increases from 3 to 5, we lose over half of clusters of size smaller than 10; while as `MinPts` increases to 7, we lose only a few more. Furthermore, while the number of clusters with 10-20 isolates stays relatively stable across `MinPts` values, the number of clusters with 50-100 isolates increase for `MinPts` of 5 and 7.

5.3.2 Cluster Purity Distribution

Of interest to our investigation is the number of isolates that fall within clusters of a particular purity. Figure 5.2 shows the number of isolates that fall within a cluster of a particular purity as a histogram. We notice that as `MinPts` increases, the purity skews towards purer clusters and that a portion of isolates remain in an impure cluster regardless of the `MinPts` value.

From `MinPts` of 3 all the way to 7, there are about 400 isolates that land in a cluster of purity between 0.4 and 0.5. This group of isolates remains largely unchanged

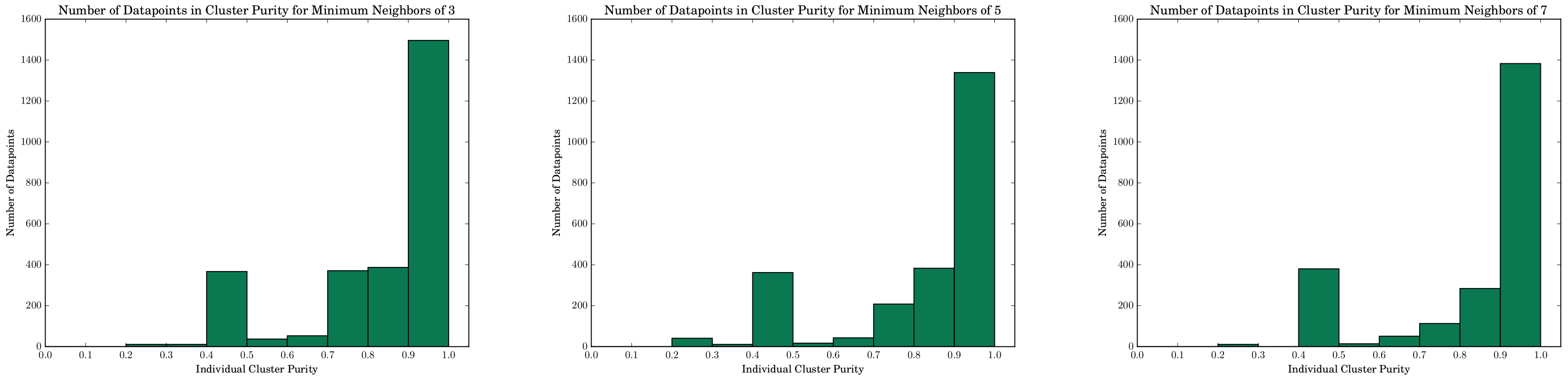


(a) Cluster Size Distribution for `MinPts` of 3

(b) Cluster Size Distribution for `MinPts` of 5

(c) Cluster Size Distribution for `MinPts` of 7

Figure 5.1: The size distribution of clusters skews heavily towards smaller clusters.



(a) Cluster Purity Distribution for `MinPts` of 3

(b) Cluster Purity Distribution for `MinPts` of 5

(c) Cluster Purity Distribution for `MinPts` of 7

Figure 5.2: The number of isolates that fall into a cluster of a given purity. We notice that that the number of isolates that fall into the 0.90 to 1.00 cluster purity range decreases as we increase `MinPts` from 3 to 5, but increases from 5 to 7.

as we restrict the MinPts value. We suspect (and discuss in Section 5.4) that certain *E. coli* strains find themselves in many host species fecal matter.

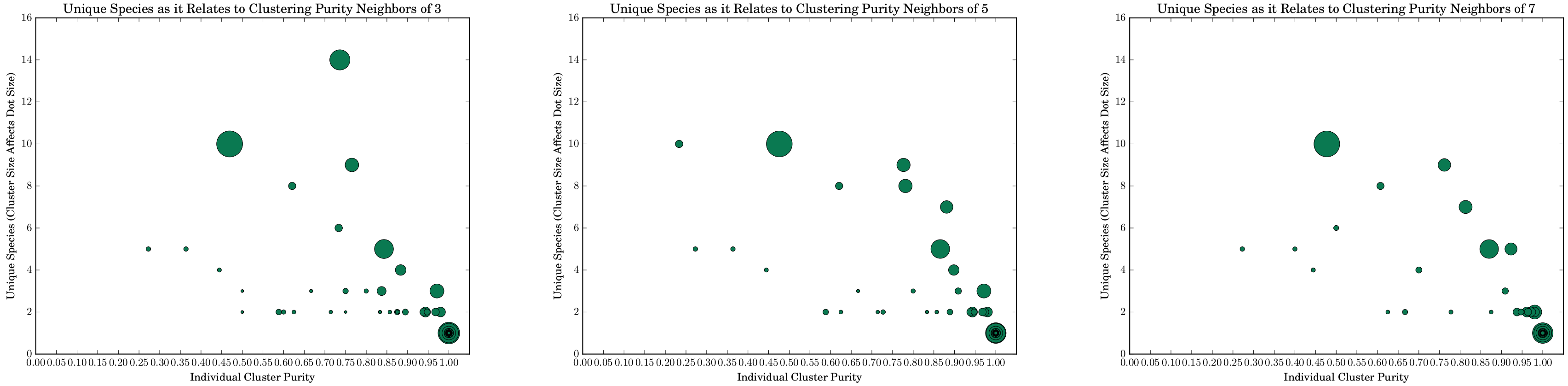
5.3.3 Unique Species in Each Cluster

Knowing the number of unique host species in our clusters is key to understanding how our strain-based MST algorithm performs. Figure 5.3 plots the number of unique species in each cluster (vertical axis) against individual cluster purity (horizontal axis) representing each cluster as a circle of diameter proportional to cluster size[2]. The points at the lower right represent many clusters of various size of 100% (or near so) purity and are stacked from largest behind cluster to smallest in front.

As MinPts values increase we see one cluster at the top right (MinPts=3) with 14 unique species disappear as MinPts becomes 5. One very low purity cluster at a MinPts value of 5 disappears when we increase MinPts to 7 in Figure 5.3c.

A particularly large cluster at around 0.45 purity with 11 unique host species, remains relatively intact (and is clearly recognizable) as MinPts changes from 3, to 5, to 7. This can account for the large amount of isolates clustered into impure clusters in Figure 5.2.

As we restrict the cluster size with MinPts, we see that this appears to break up some clusters and cause others to become bigger. It is difficult to track exactly how a cluster changes without making some simplifying assumptions or without tracking all 4,610 isolates as they move from cluster to cluster.



(a) Cluster Purity for MinPts of 3

(b) Cluster Purity for MinPts of 5

(c) Cluster Purity for MinPts of 7

Figure 5.3: These three dimensional graphs show the individual cluster purity in the horizontal axis, the number of unique species in the vertical axis, and the relative size of the cluster in the diameter of the dots. Each individual dot is its own cluster. We find that as we restrict cluster to needing more neighbors (increasing the MinPts value), we lose some clusters and gain more pure clusters.

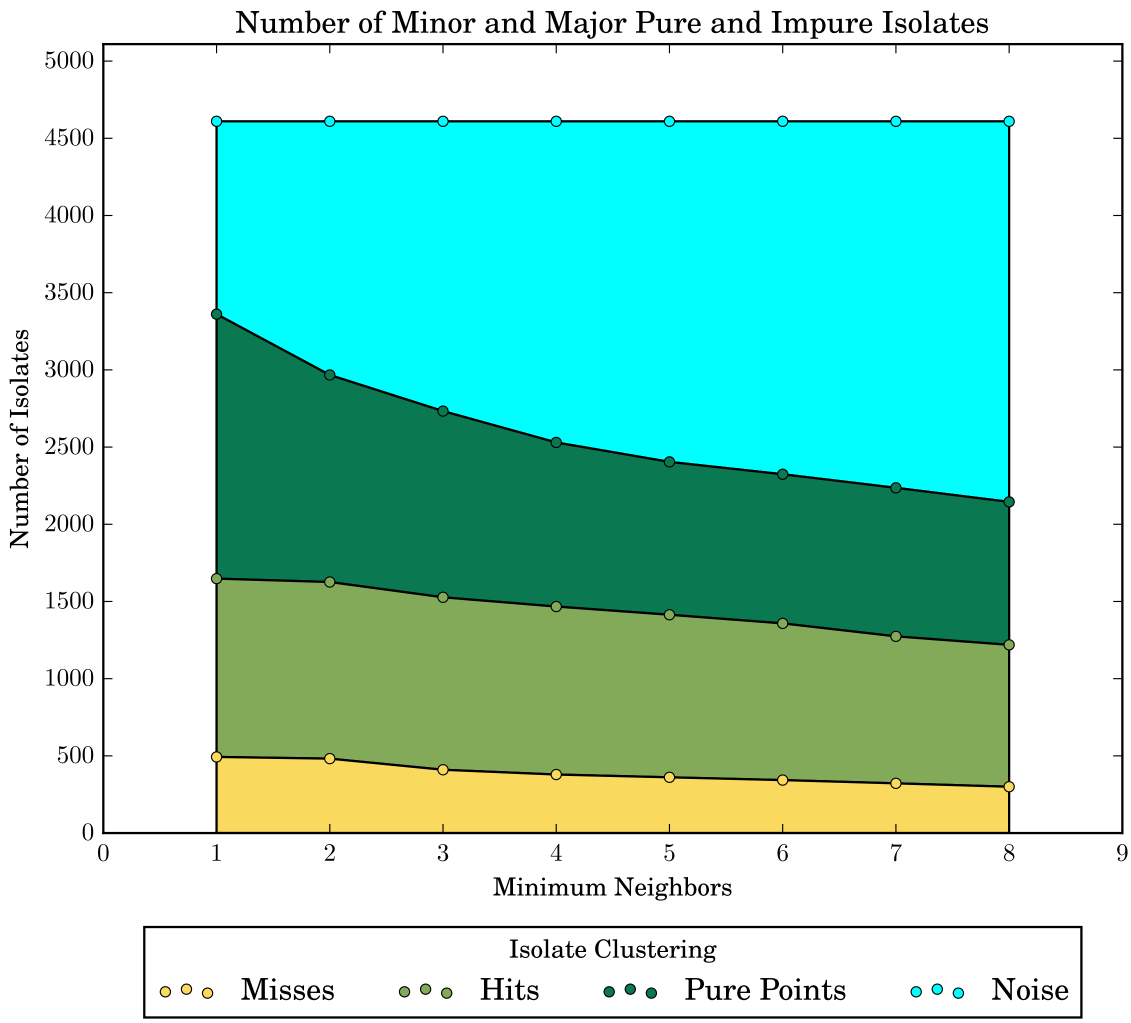


Figure 5.4: As `MinPts` increases, we see that we cluster fewer isolates. throughout, the number of major pure isolates stays relatively equal to the number of major impure isolates.

5.3.4 Clustering Coverage

Clustering coverage is important to consider, since we want our clustering algorithm to apply to as many isolates as possible. Towards this end we investigated the four metrics introduced in Section 5.2.2 — noise, misses, hits, and pure points — for each `MinPts` clustering investigated. We hope to find the `MinPts` value that gives us the most pure points, but will also settle for the fewest misses, shown in Figure 5.4.

[2] A linear scaling of the diameter with respect to *the largest cluster amongst all the clusterings* defines the diameters of the dots.

The cyan area is noise — isolates that were not clustered. The dark green area is the proportion of pure points. Light green is the number of hits. Gold is the number of misses.

Table 5.2 displays the number of isolates that fall into the categories in Figure 5.4. Table 5.3 shows the percentage of *all* isolates that ended up in a cluster where their host species was the minority, the most dominant, and that make up the entirety of the cluster and those categorized as noise. Table 5.3 shows the percentage of *clustered* isolates that ended up in a cluster where their host species was the minority, the most dominant, and that make up the entirety of the cluster. Table 5.5 compares the

MinPts	**Misses**	**Hits**	**Pure Points**	**Noise**	**Total**
1	493	1155	1713	1249	4610
2	482	1144	1341	1643	4610
3	410	1117	1206	1877	4610
4	379	1088	1063	2080	4610
5	361	1053	990	2206	4610
6	343	1015	966	2286	4610
7	322	952	962	2374	4610
8	300	919	926	2465	4610

Table 5.2: The number of isolates that ended up in clusters where their host species is the minority (Misses), the dominant (Hits), and make up the entirety of the cluster (Pure Points) or were categorized as Noise.

MinPts	**Misses (%)**	**Hits (%)**	**Pure Points (%)**	**Noise (%)**	**All Isolates**
1	10.7	25.1	37.2	27.1	4610
2	10.5	24.8	29.1	35.6	4610
3	8.9	24.2	26.2	40.7	4610
4	8.2	23.6	23.1	45.1	4610
5	7.8	22.8	21.5	47.9	4610
6	7.4	22.0	21.0	49.6	4610
7	7.0	20.7	20.9	51.5	4610
8	6.5	19.9	20.1	53.5	4610

Table 5.3: The percentage of all isolates that ended up in clusters where their host species is the minority (Misses), the dominant (Hits), and make up the entirety of the cluster (Pure Points) or were categorized as Noise.

MinPts	Misses (%)	Hits (%)	Pure Points (%)	Clustered Isolates
1	14.7	34.4	51.0	3361
2	16.2	38.6	45.2	2967
3	15.0	40.9	44.1	2733
4	15.0	43.0	42.0	2530
5	15.0	43.8	41.2	2404
6	14.8	43.7	41.6	2324
7	14.4	42.6	43.0	2236
8	14.0	42.8	43.2	2145

Table 5.4: The percent of isolates clustered that ended up in a cluster where their host species is the minority (Misses), the dominant (Hits), and make up the entirety of the cluster (Pure Points).

percent of all isolates that DBSCAN placed in a cluster and determined to be noise, while Table 5.6 shows the actual values.

MinPts	Clustered (%)	Noise (%)	All Isolates
1	72.9	27.1	4610
2	64.4	35.6	4610
3	59.3	40.7	4610
4	54.9	45.1	4610
5	52.1	47.9	4610
6	50.4	49.6	4610
7	48.5	51.5	4610
8	46.5	53.5	4610

Table 5.5: The percent of isolates that DBSCAN placed in a cluster or determined to be noise.

MinPts	Clustered	Noise	All Isolates
1	3361	1249	4610
2	2967	1643	4610
3	2733	1877	4610
4	2530	2080	4610
5	2404	2206	4610
6	2324	2286	4610
7	2236	2374	4610
8	2145	2465	4610

Table 5.6: The number of isolates that DBSCAN placed in a cluster or determined to be noise.

It is good to note that the number of misses are low and flatten out as we increase `MinPts` from a value of 3, giving us good reason not to investigate clustering where `MinPts` is greater than we have already investigated. The number of pure points stays relatively equal to the number of hits. Important in Figure 5.4 is the amount of isolates that the algorithm does cluster. The combination of the gold and two green areas show the total number clustered, while the cyan shows the number of isolates that were *not* clustered. It is unfortunate that the number of noise isolates is high, but we plan to mitigate that in future work.

5.3.5 Overall Clustering Purity

Overall clustering purity, defined in Equation 5.4, is the number of isolates clustered that end up in a cluster where their host species is the most plural host species. The overall accuracy is the proportion of correctly classified isolates out of all the isolates under consideration. We want to maximize both values, but would prefer the former over the latter. Coverage is an issue we are concerned about, but we plan to mitigate this issue by leveraging [37] against clusters of isolates.

Figure 5.5 shows the overall accuracy compared to the overall clustering purity. A `MinPts` value equal to 3 is the last `MinPts` value where the overall accuracy stays above 0.50. It is not for a lack of correctness, as Figure 5.4 shows, but more that isolates simply are not being clustered as we restrict the `MinPts` value. In fact, the overall clustering purity in Figure 5.5 stays relatively constant. This means that if an isolate is clustered by our algorithm, it will likely be clustered with other isolates of the same host species.

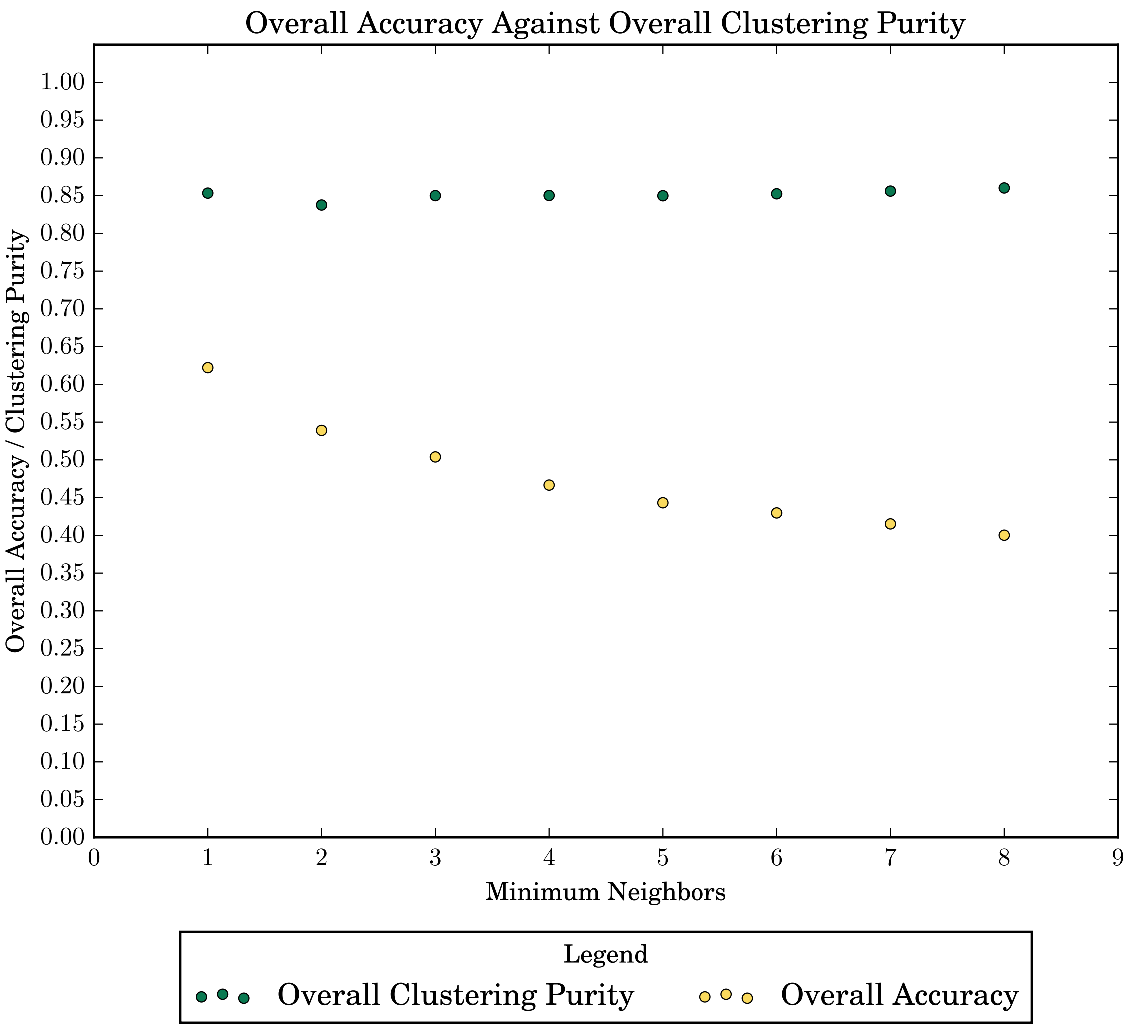


Figure 5.5: The overall accuracy decreases as we restrict `MinPts`. The overall clustering purity stays relatively the same as we increase the value for `MinPts`. That is, for clustered isolates, the classification algorithm stays relatively the same relative to the number of isolates accurately clustered.

5.4 Discussion

In general, we observe two trends in our data. For the isolates that get clustered into strains, our approach correctly identifies the host species with over 80-85% accuracy. This accuracy is sufficient to conduct sophisticated MST studies. Most of the strains discovered in the CPLOP data show high degree of purity, and even considering the presence of a few large impure clusters, most of the clustered isolates fall into strains of high purity.

At the same time, the pure strain-based approach suffers from a drop in the coverage as the size of a cluster grows. This means that in general CPLOP isolates tend to be very diverse and come from strains for which not enough DNA material has been collected and pyrosequenced. Identifying the host species for isolates that do not fall into strains/clusters using the pure strain-based method is impossible. In future work, our goal is to combine the k-NN-based MST method of [37] with the strain-based approach discussed in this paper to increase coverage while preserving the high MST accuracy.

One factor explaining the large impure clusters is the possibility that these clusters represent what the biologists call “transient” strains, i.e., strains that persist in more than one host species. Such a characteristic can compound MST by making certain strains of *E. coli* less reliable as FIB for identifying host species. In Figure 5.2, we see evidence of that and it is revealed in Figure 5.3. One mitigation strategy may be to reduce the presence of these strains in the library holding the FIB. Another may be to fall back to an alternative MST technique that works with CPLOP when an unknown isolate falls into an impure cluster. Finally, if a true transient strain is indeed discovered, and an isolate is mapped to it, our MST procedure can simply acknowledge that the query isolate belongs to a transient strain and provide information about the host species that show high frequency of *E. coli* incidence from this strain. In order to handle the lack of complete clustering coverage — when DBSCAN marks an isolate as noise — we propose a fallback method : the k-Nearest Neighbors Resolution Algorithms for Pyroprints, described in Chapter 6.

Chapter 6

THE k-NEAREST NEIGHBORS RESOLUTION ALGORITHMS FOR PYROPRINTS

Since each *E. coli* isolate in CPLOP has a pyroprint of two separate *ITS* regions[1], we effectively have two comparison functions between datapoints (isolates), complicating our use of k-NN. k-NN provides CPLOP biologists a transparent and intuitive way of understanding the host species classification it asserts, so we find that it will be a usefully insightful . Applying k-NN to CPLOP isolates gives us two lists, one for each *ITS* region, that we must make a single host species classification from. In order to accommodate multiple comparison functions, we need a strategy to resolve multiple k-nearest neighbors lists. Rather than create a new similarity metric out of a pair of similarity scores, which may deviate from the inherent nature of the comparison function, we choose to update the k-NN method with four different ways of selecting the resultant category label: the k-Nearest Neighbors Resolution Algorithms for Pyroprints (k-RAP). We describe these four methods, the evaluation criteria we judge them by, and their results below.

6.1 Methodology

In what follows, we generalize our problem. Given u and v, two library objects (isolates), and a collection of comparison functions, $\mathbb{C}= (\mathbb{C}_1, \ldots \mathbb{C}_m)$, with $m > 1$, comparing u to v gives us a collection of values: $\mathbb{C}(u, v) = (\mathbb{C}_1(u, v), \ldots, \mathbb{C}_m(u, v))$. All four resolution procedures described in this section work with such a generalized representation of isolates and comparison functions between them.

[1] see Section 2.2.3

Given an unknown isolate u, a library of classified[2] isolates $\mathbb{L}$, and a set of comparison functions $\mathbb{C}$, we compare u to each object in $\mathbb{L}$ using each comparison function in $\mathbb{C}$. To resolve these comparison functions, we propose four algorithms:

6.1.1 Comparing Isolates

Comparing isolates to each other is of primary interest to biologists using CPLOP. CPLOP represents each isolate by a pair of mutually incomparable pyroprints: one for each of the two *ITS* regions. As a result, given isolates I_1, I_2, we can represent each as a pair of pyroprint vectors

$$I_1 = (\vec{q_1}, \vec{q_2}) \text{ and } I_2 = (\vec{r}_1, \vec{r}_2),$$

where $\vec{q_1}$ and $\vec{r_1}$ are respectively I_1 and I_2's *ITS-1* pyroprint and $\vec{q_2}$ and $\vec{r}_2$ are respectively I_1 and I_2's *ITS-2* pyroprint [9]. Since pyroprints from different regions are incomparable, comparing isolates must be done as follows:

$$\mathbb{C}(I_1, I_2) = (\rho(\vec{q_1}, \vec{r_1}), \rho(\vec{q_2}, \vec{r_2})),$$

where $\rho(\cdot, \cdot)$ is between pyroprints of the same *ITS* region and is the Pearson correlation coefficient. Thus, when comparing isolates, we effectively have two different similarity metrics, one for each *ITS* region:

$$\mathbb{C}(I_1, I_2) = (\mathbb{C}_1(I_1, I_2), \mathbb{C}_2(I_1, I_2)).$$

[2] A "classified isolate" is an isolate for which the host species has been identified in the database.

Algorithm 2 Isolate Comparison Metric

Input:

- $u \in \mathcal{I}$: an isolate
- $v \in \mathcal{I}$: an isolate

Output:

- $\mathbb{R}$ value, indicating similarity

Requires:

- Each isolate has a pyroprint in the i^{th} ITS region

```
procedure $C_i(u, v)$
    $\vec{p_u} \leftarrow \textsc{GetPyroprint}_i(u)$
    $\vec{p_v} \leftarrow \textsc{GetPyroprint}_i(v)$
    return $\textsc{PearsonCorrelation}(\vec{p_u}, \vec{p_v})$
end procedure
```

6.1.2 α Filtering

Our first modification to k-NN is an additional condition at step 4, after finding the k-nearest neighbors:

4. Consider only the top k entries in N above threshold α

The α threshold allows biologists to filter out neighbors that are among the k closest, but too dissimilar to compare. When comparing multiple pyroprints of the same region of a single isolate for quality control, the Pearson correlation coefficient between them is strictly above 0.995. As a result, for many other studies — not necessarily MST-focused — CPLOP researchers use a Pearson correlation coefficient of 0.990 or above to define a strain of *E. coli*. Filtering by some value near this may give more accurate results and provides an intuitive way to relate these lists to other studies.

Using the example from Figures 2.12 and 2.13, Figure 6.1 shows how when using k-NN where $k = 9$, we can further restrict the k-nearest neighbors list if we believe

k	Class	Position	Similarity
	?	(5,5)	0.000
1	B	(6,6)	1.414
2	A	(3,6)	2.236
3	A	(4,7)	2.236
4	C	(3,3)	2.828
5	B	(8,5)	3.000
6	B	(6,8)	3.162
7	C	(3,2)	3.606
8	C	(1,3)	4.472
9	C	(1,1)	5.657

Figure 6.1: For $k = 9$, filtering the k-nearest neighbors by $\alpha = 4.000$ ends up dropping the last two datapoints (marked in gray) off of the list, changing the resultant classification from C to B. In the case of Euclidean distance, using α means we filter any Euclidean distance *above* α.

k-NN should ignore any datapoints outside of α. For Figure 6.1, Euclidean distance is the chosen comparison function and using an α threshold to filter means we filter any datapoints from the k-nearest neighbors list if the similarity value is *above* α. CPLOP uses pyroprints as datapoints and the Pearson correlation coefficient to compare pyroprints; the Pearson correlation coefficient similarity value ranges from -1 to 1, where a Pearson correlation coefficient close to 1 denotes highly similar datapoints. Thus, filtering using an α threshold on Pearson correlation coefficient values requires filtering datapoints that are *below* α. Algorithm 3 describes this process in pseudocode.

Algorithm 3 k-NN with α Threshold

Input:

- *list*: sorted list

Output:

- *nearest*: the top k elements above the threshold α

Requires:

- k and α are predetermined
- each element in *list* has a similarity field *sim*

```
procedure FILTER$_{k,\alpha}$($list$)
    $nearest \leftarrow \emptyset$
    $i \leftarrow 0$
    while $i < k$ and $list[i].sim < \alpha$ do
        add $list[i]$ to $nearest$
        $i \leftarrow i + 1$
    end while
    return $nearest$
end procedure
```

6.1.3 Meanwise Resolution

Meanwise Resolution combines the result of the two comparisons between two isolates in order to build a single k-nearest neighbors list, from which simple k-NN proceeds to classify. Combining the similarity values from the two comparison functions is simply a mapping from the two similarity values to a single value. Arithmetic mean, geometric mean, and Euclidean distance are examples, but we chose to use Euclidean distance. Figure 6.2 depicts an example classifying an unknown isolate u using the arithmetic mean as the mean function.

Formally: For u and a $p \in \mathbb{L}$, we take the mean of the result of all of the comparison functions and build a single k-nearest neighbors list from it. The mean can be any metric mapping $\mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ and in the investigated implementation, we use the Euclidean distance, also known as the L^2 norm. A single k-nearest neighbors list

k	Isolate x	Host species	$\rho_{\textit{ITS-1}}(u, x)$	$\rho_{\textit{ITS-2}}(u, x)$	$mean(\rho_{\textit{ITS-1}}, \rho_{\textit{ITS-2}})$
1	a	Cat	0.994	0.991	0.993
2	b	Dog	0.990	0.994	0.992
3	c	Dog	0.995	0.989	0.992
4	d	Chicken	0.985	0.987	0.986
5	e	Cat	0.980	0.990	0.985
6	f	Cat	0.978	0.990	0.984
7	g	Cat	0.980	0.984	0.982
8	h	Chicken	0.952	0.960	0.956

Figure 6.2: An example classifying an unknown isolate u using Meanwise Resolution with $k = 8$, where the mean function is the arithmetic average of the two ρ values.

results from this algorithm that we filter by k and α and use to classify the unknown. Algorithm 4 describes this process in pseudocode.

Algorithm 4 Meanwise Resolution

Input:

- $u \in \mathbb{U}$: an unknown isolate
- $\mathbb{L} \subseteq \mathcal{I}$: a library of known isolates

Output:

- $\mathcal{S}$ value, classifying the unknown isolate u

Requires:

- k, α, and the set of comparison metrics $\mathbb{C}$ are predetermined

```
procedure ClassifyMean($u, \mathbb{L}$)
    $N \leftarrow \emptyset$                      /* Make nearest neighbors list.          */
    for $p \in \mathbb{L}$ do                  /* For each library element:             */
        $\mathbb{A} \leftarrow \{\emptyset\}$                  /* Make empty set for results.       */
        for $C_i \in \mathbb{C}$ do              /* For each comparison metric:       */
            $sim \leftarrow C_i(u, p)$                           /* Compare.            */
            add $sim$ to $\mathbb{A}$                            /* Track result.       */
        end for
        $mean \leftarrow$Mean($\mathbb{A}$)                /* Mean the results.             */
        add $(mean, p)$ to $N$               /* Add mean to neighbors.        */
    end for
    sort $N$ by $sim$                      /* Sort by most similar.            */
    $N \leftarrow$Filter$_{k,\alpha}(N)$               /* Keep the nearest.                */
    $s \leftarrow$ FindMostPluralSpecies($M$)
    return $s$
end procedure
```

6.1.4 Resolution by Winner

Resolution by Winner performs k-NN classification on the two k-nearest neighbors lists resulting from the two comparison functions, ρ on *ITS-1* and ρ on *ITS-2*, picking the host species with the highest number of instances in its original list. In this way, the strategy picks the "winning" classification from the two lists. Figure 6.3 shows how this works.

Formally: For each comparison function, we make a k-nearest neighbors list and

k (*ITS-1*)	Host species
1	Bat
2	Bat
3	Cat
4	Pigeon
5	Pigeon
6	Human
7	Human
8	Bat

(a) A $k = 8$ nearest neighbors list for the *ITS-1* region of an unknown isolate.

k (*ITS-2*)	Host species
1	Pigeon
2	Bat
3	Cat
4	Pigeon
5	Bat
6	Pigeon
7	Cat
8	Pigeon

(b) A $k = 8$ nearest neighbors list for the *ITS-2* region of an unknown isolate.

Figure 6.3: In this example of Resolution by Winner for $k = 8$, k-NN on the *ITS-1* region results in a classification of Bat, since there are 3 bats in its k-nearest neighbors list and k-NN on *ITS-2* results in Pigeon, since there are 4 Pigeons in its k-nearest neighbors list. The classification resulting from these two lists is Pigeon, since Pigeon shows up more in its original list.

filter by k and α accordingly. Once we finish building each comparison function's k-nearest neighbors list, we find the most plural classification from each list and track the number of times that classification shows up in that list. Then, we classify u based off the classification that has the highest number in its corresponding list. Algorithm 5 describes this process in pseudocode.

Algorithm 5 Resolution by Winner

Input:

- $u \in \mathbb{U}$: an unknown isolate
- $\mathbb{L} \subseteq \mathcal{I}$: a library of known isolates

Output:

- $\mathcal{S}$ value, classifying the unknown isolate u

Requires:

- k, α, and the set of comparison metrics $\mathbb{C}$ are predetermined

```
procedure ClassifyWinner($u, \mathbb{L}$)
    $N \leftarrow \emptyset$                       /* New list to track neighbor lists.          */
    for $C_i \in \mathbb{C}$ do                   /* For each comparison metric:                */
        $N_i \leftarrow \emptyset$                     /* Make nearest neighbors list.           */
        for $p \in \mathbb{L}$ do                 /* For each library element:              */
            $sim \leftarrow C_i(u, p)$                     /* Compare.                       */
            add $(sim, p)$ to $N_i$                   /* Add to neighbors.              */
        end for
        sort $N_i$ by $sim$                        /* Sort by most similar.              */
        $N_i \leftarrow$Filter$_{k,\alpha}(N_i)$                 /* Keep the nearest.                  */
        add $N_i$ to $N$
    end for
    $S \leftarrow \{\emptyset\}$                     /* To track each list's most plural.          */
    for $N_i \in N$ do
        $s \leftarrow$ FindMostPluralSpecies($N_i$)
        add $s$ to $S$
    end for
    return Max($S$)                 /* The most plural overall.                   */
end procedure
```

6.1.5 Resolution by Union

Resolution by Union builds a list that is the union of both k-nearest neighbors lists, classifying the unknown as the most dominant species in the resulting union It may be that the most dominant host species in each list is different, but a second or third most dominant host species may be more appropriate since it shows up in both lists. Figure 6.4 shows an example of this.

In Figure 6.4, Figures 6.4a and 6.4b show the k-nearest neighbors lists for *ITS-1* and *ITS-2* respectively. The dominant host species in each are Human, with 5 instances, and Turkey, with 4 instances. Combining these into a union of the two in Figure 6.4d[3], we see that the dominant host species is Dog with 6 instances in the union, due to the 3 instances of Dog in each *ITS* k-nearest neighbors list.

Formally: For each comparison function, we make a k-nearest neighbors list and filter by k and α accordingly. After building each k-nearest neighbors list, we combine the lists into a set, keeping track of the original list position for tie-breaking. From this set, which we dub the union, we count the classifications present in the union and classify u as the most plural in the union of the lists, compared to the other lists. Algorithm 6 describes this process in pseudocode.

[3] The union depicted in Figure 6.4d contains the original k for clarity of entry source. These values can help with tie breaking as well, if breaking ties based on average position, or lowest k.

k (*ITS-1*)	Host species
1	Dog
2	Human
3	Human
4	Human
5	Dog
6	Dog
7	Human
8	Human

(a) A $k = 8$ nearest neighbors list for the *ITS-1* region of an unknown isolate.

k (*ITS-2*)	Host species
1	Turkey
2	Dog
3	Dog
4	Turkey
5	Turkey
6	Dog
7	Turkey
8	Cow

(b) A $k = 8$ nearest neighbors list for the *ITS-2* region of an unknown isolate.

Region	k	Host species
ITS-1	1	Dog
ITS-2	1	Turkey
ITS-1	2	Human
ITS-2	2	Dog
ITS-1	3	Human
ITS-2	3	Dog
ITS-1	4	Human
ITS-2	4	Turkey
ITS-1	5	Dog
ITS-2	5	Turkey
ITS-1	6	Dog
ITS-2	6	Dog
ITS-1	7	Human
ITS-2	7	Turkey
ITS-1	8	Human
ITS-2	8	Cow

(c) Resulting union of the k-nearest neighbors lists from each *ITS* region.

List	Host species	Count
ITS-1	Human	5
ITS-1	Dog	3
ITS-2	Turkey	4
ITS-2	Dog	3
ITS-2	Cow	1
Union	Dog	6
Union	Human	5
Union	Turkey	4
Union	Cow	2

(d) Host species counts of all three lists, *ITS-1*, *ITS-2*, and the union of the two.

Figure 6.4: In this example of Resolution by Union for $k = 8$, we see that when considering each *ITS* k-nearest neighbors list separately, there are two different dominant host species.

Algorithm 6 Resolution by Union

Input:

- $u \in \mathbb{U}$: an unknown isolate
- $\mathbb{L} \subseteq \mathcal{I}$: a library of known isolates

Output:

- $\mathcal{S}$ value, classifying the unknown isolate u

Requires:

- k, α, and the set of comparison metrics $\mathbb{C}$ are predetermined

```
procedure ClassifySetwise($u, \mathbb{L}$)
    $\mathbb{M} \leftarrow \{\emptyset\}$                /* Create new common set.              */
    for $C_i \in \mathbb{C}$ do             /* For each comparison metric:         */
        $N_i \leftarrow \emptyset$                  /* Make nearest neighbors list.        */
        for $p \in \mathbb{L}$ do              /* For each library element:           */
            $sim \leftarrow C_i(u, p)$                 /* Compare.                    */
            add $(sim, p)$ to $N_i$                /* Add to neighbors.           */
        end for
        sort $N_i$ by $sim$             /* Sort by most similar.               */
        $N_i \leftarrow$Filter$_{k,\alpha}(N_i)$              /* Keep the nearest.            */
        for $n \in N_i$ do              /* For each nearest neighbor:          */
            add $n$ to $\mathbb{M}$                    /* Add to common set.          */
        end for
    end for
    $s \leftarrow$ FindMostPluralSpecies($M$)
    return $s$
end procedure
```

6.1.6 Resolution by Intersection

Resolution by Intersection works to build a single set of size k, called the intersection, consisting only of isolates that appear in both *ITS* k-nearest neighbors lists. It does this by initially querying for each *ITS* region, expanding the list until the intersection is the required size — k. This is the most complicated and restrictive of the resolution strategies, possibly requiring α filtering in order to restrict the search. Filtering by α means the construction on the intersection stops if expanding the k-nearest neighbors means going outsidethe α threshold, which may result in an intersection that is smaller than k[4].

Figure 6.5 shows how this strategy proceeds. In Figure 6.5a, we see that there are many Cows in the *ITS-1* k-nearest neighbors list, but few of those isolates show up in the *ITS-2* k-nearest neighbors list in Figure 6.5b. We then build the intersection, shown in Figure 6.5c, but there are not enough common isolates to make a k=8 sized intersection, so we must extend the initial lists to 13 in order to find enough common isolates. Once we have found enough common isolates, k-NN can proceed normally on the intersection.

For each comparison function, we make a k-nearest neighbors list and filter by k and α accordingly, but ensure that we do not lose track of the entire sorted list of results. After building each k-nearest neighbors list, we inspect each list for common isolates. We add isolates that appear in every list into a set that we call the intersection. If the size of the intersection is k, then we are done. Otherwise, we increase the length of our individual lists by δ and search for common isolate. This process repeats until the size of the intersection is k, or all of the isolates in the individual lists are below threshold α. Algorithm 7 describes this function in pseudocode.

[4] Since the same is true for natural k-NN, we find this acceptable.

k (*ITS-1*)	Isolate ID	Host species
1	**5823**	**Pig**
2	**2833**	**Pig**
3	8873	Cow
4	**5939**	**Cow**
5	6156	Cow
6	3676	Human
7	7853	Cow
8	**5331**	**Cow**
9	2189	Cow
10	**2053**	**Pig**
11	**8962**	**Pig**
12	**3813**	**Human**
13	**8173**	**Pig**

(a) A $k = 8$ nearest neighbors list including (fabricated) isolate IDs for the *ITS-1* region of an unknown isolate, extended in order to find isolates in common with the *ITS-2* list, which are bold.

k (*ITS-1*)	Isolate ID	Host species
1	**2833**	**Pig**
2	2916	Cow
3	**3813**	**Human**
4	**5939**	**Cow**
5	6854	Dog
6	**5823**	**Pig**
7	**2053**	**Pig**
8	8485	Dog
9	**8173**	**Pig**
10	6497	Human
11	9208	Cow
12	**5331**	**Cow**
13	**8962**	**Pig**

(b) A $k = 8$ nearest neighbors list including (fabricated) isolate IDs for the *ITS-2* region of an unknown isolate, extended in order to find isolates in common with the *ITS-1* list, which are bold.

k (*ITS-1*)	k (*ITS-2*)	Isolate ID	Host species
2	1	2833	Pig
1	6	5823	Pig
4	4	5939	Cow
12	3	3813	Human
10	7	2053	Pig
8	12	5331	Cow
13	9	8173	Pig
11	13	8962	Pig

(c) Common isolates in the *ITS-1* and *ITS-2* k-nearest neighbors lists, sorted by average k.

Figure 6.5: An example Resolution by Intersection for $k = 8$ showing how the original k-nearest neighbors lists did not have enough isolates — 8 are needed — in common, but when extended out to 15, there are 8 common isolates. From here, classification can proceed to classify the unknown as the most dominant host species in the intersection, Pig.

Algorithm 7 Resolution by Intersection

Input:

- $u \in \mathbb{U}$: an unknown isolate
- $\mathbb{L} \subseteq \mathcal{I}$: a library of known isolates

Output:

- $\mathcal{S}$ value, classifying the unknown isolate u

Requires:

- k, α, δ, and the set of comparison metrics $\mathbb{C}$ are predetermined

```
procedure ClassifyIntersection($u$, $\mathbb{L}$)
    $N \leftarrow \emptyset$                      /* New list to track neighbor lists.    */
    for $C_i \in \mathbb{C}$ do                   /* For each comparison metric:          */
        $N_i \leftarrow \emptyset$                 /* Make nearest neighbors list.         */
        for $p \in \mathbb{L}$ do                  /* For each library element:            */
            $sim \leftarrow C_i(u, p)$                      /* Compare.                     */
            add $(sim, p)$ to $N_i$                         /* Add to neighbors.            */
        end for
        sort $N_i$ by $sim$                          /* Sort by most similar.               */
        add $N_i$ to $N$
    end for
    $done \leftarrow$ false
    while $\neg done$ do
        for $N_i \in N$ do
            $N'_i \leftarrow \emptyset$
            $N'_i \leftarrow$ Filter$_{k,\alpha}(N_i)$
        end for
        if $|N'_1 \cap \cdots \cap N'_n| < k$ then
            $k \leftarrow k + \delta$
        else
            $N_\cap \leftarrow N'_1 \cap \cdots \cap N'_n$
            $done \leftarrow$ true
        end if
    end while
    return FindMostPluralSpecies($N_\cap$)
end procedure
```

6.2 Evaluation

Evaluating k-RAP requires an understanding of how well it classifies the host species of an isolate. There are a few areas of focus that we have when interpreting the results of k-RAP:

- What size k achieves the best results?
- What size α achieves the best results?
- Which metric resolution algorithm achieves the best results?

Indeed we can define "best" in many ways, but we choose to look at two metrics, recall and precision, and a combination of the two, the F-measure. The metrics look at the accuracy of the classification on the object and the object on the classification respectively, while F_1-measure hopes to represent a balance between the two. We test k-RAP by performing cross validation with holdout.

6.2.1 Cross Validation with Holdout

To gauge the effectiveness of k-RAP at classifying the host species of an isolate, we cross-validated against the library by separately holding out each isolate in CPLOP from CPLOP, classifying it against CPLOP, and verifying whether it is correct. Since each isolate in CPLOP has the correct host species, we know whether a classification is correct or not.

6.2.2 Recall

In our study, recall tracks how well we are able to discover all isolates from a given category, i.e. with a given host species. Given a category (host species name), the recall for that host species is the percentage of isolates taken from this host species that have been properly identified. For example, if our database had 100 cat isolates, and 74 of them were classified by our method as having come from a cat, the recall would be 74%. In this study, we compute both overall recall (what percentage of isolates were classified as their proper host species label) as well as host species-level recall (what percentage of isolates that came from dogs/humans/sheep/etc. were classified as their proper label).

6.2.3 Precision

Precision tracks how well our method avoids misclassification errors. Given a category and a list of isolates our method classified as belonging to it, the precision of the method on the category is the percent of isolates from the list that has the correct label. For example, if our method returned 100 isolates labelled "Dog" of which 77 isolates really did come from dogs, the precision of the method is 77%. As with recall, we compute both overall precision, as well as the precision for each category/species label.

6.2.4 *F*-Measure

The F_1-measure, F_1, is the *harmonic mean* of the precision, P and the recall, R:

$$F_1 = \frac{2}{\frac{1}{P} + \frac{1}{R}} = 2 \cdot \frac{P \cdot R}{P + R}$$

While we prefer maximizing this value, a value near 0.5 means we are doing well.

6.3 Results

Our results focused on adjusting three parameters: the number k of nearest neighbors to consider, the α threshold value, and the resolution algorithm.

6.3.1 Adjusting k

Adjusting k is an important first step. We investigate k values ranging from 1 to 17, but focus primarily on $k \leq 12$. At this point, we do not filter the results in order to focus primarily on the affect of the size of the k-nn list. Thus, α is 0, allowing for the full k list to factor into classification.

Overall, for $k \geq 5$, the accuracy does not improve, but instead levels off. Depending on the resolution algorithm, this value is between 65% and 75% accuracy, as shown in Figure 6.6. By "overall," we mean that for every classification, we validated if it was correct and calculated what proportion to all classifications made that represents to determine accuracy. When looking at all classifications, precision and recall are identical values, as is F-measure.

One good example is the Cow. As Figure 6.7 shows, Cow follows a trend similar to the overall accuracy, staying roughly between 70% and 95% accurate. Certain algorithms get worse for $k > 5$, while other improve.

Figure 6.8 examines the relationship between R and P. This can help us understand the trade offs of choosing one k over another. We will later build a meaningful strategy for how confident we are at recalling a species versus our confidence in a classification of a species.

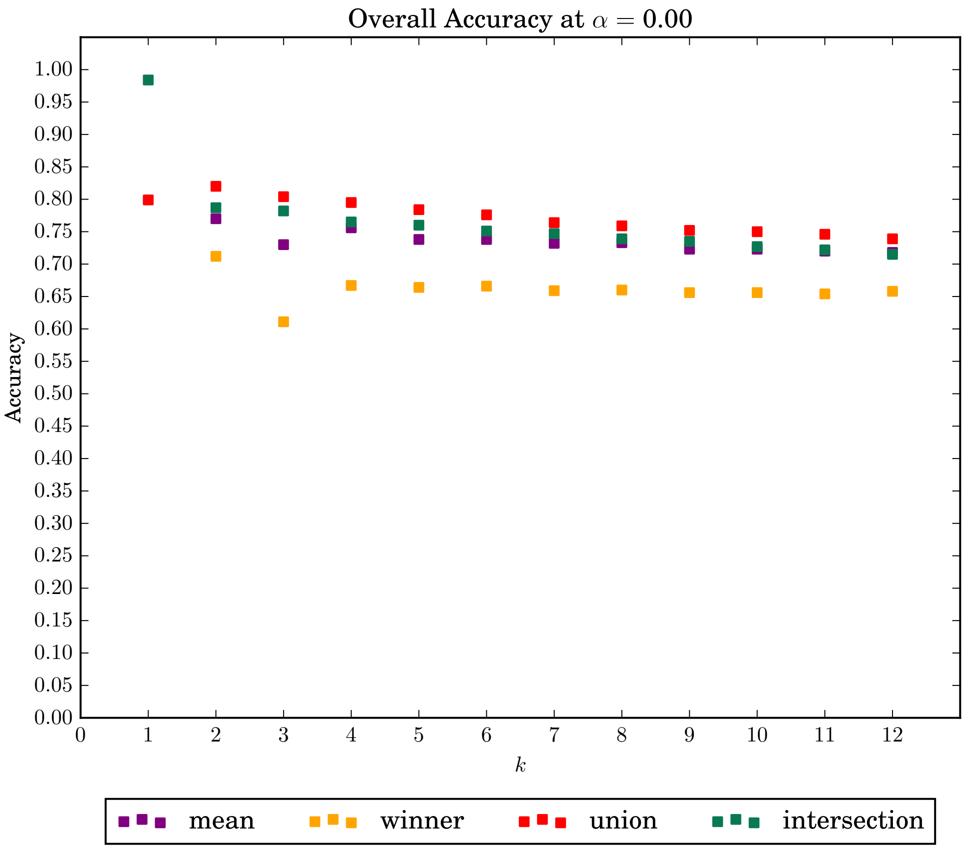


Figure 6.6: The accuracy of all classifications performed with CPLOP across the four different algorithms with $\alpha = 0.00$ shows little improvement for $k > 5$. We look at only the percentage of correct classifications, since that value is equivalent to the precision and the recall.

6.3.2 Adjusting α

By adding a threshold value, we investigated whether this further limitation improves the accuracy by restricting outliers from populating a k-nn list. We investigate $\alpha = \{0.00, 0.98, 0.99\}$. Outside of this study, $\alpha = 0.99$ defines the boundary between strains. One reason we investigate 0.98 is to see whether loosening our definition of strain differentiation gives us a better accuracy.

Overall, we observe that the accuracy slightly improves as we increase the α threshold. Figure 6.9 shows that overall, the accuracy increases as we increase α.

Adding the α made minimal changes to the accuracy of Cow classifications, so

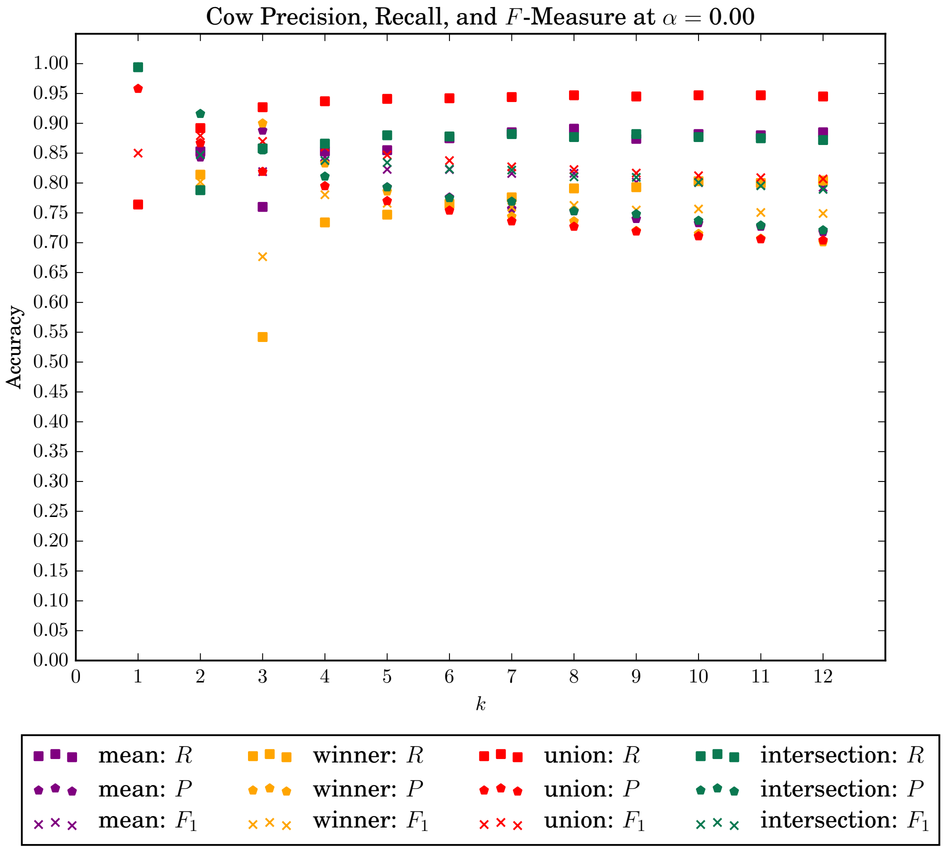


Figure 6.7: There are 1838 Cow isolates in CPLOP. For most resolution algorithms, we observe little improvement when $k > 5$.

only the recall versus precision is shown in Figure 6.10. More details into how α affect the classification accuracy can be seen in Tables 6.1, 6.2, and 6.3.

6.3.3 Adjusting the Algorithm

Choosing which algorithm to resolve the two different regions of each isolate is an important step. We investigate the differences between the aforementioned four algorithms as they relate to k and α values and how each differ among species of different representation. With library-based-MST, it is important to realize the representation of a species in the library may heavily skew the accuracy of the library.

While interpreting the data, we state that there may be some "%" increase or

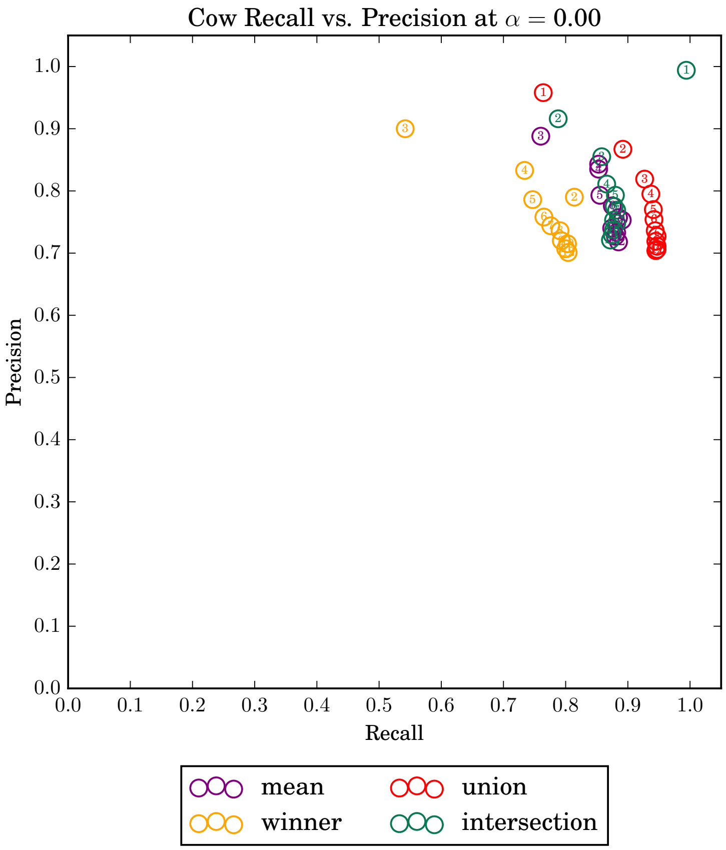


Figure 6.8: There are 1838 Cow isolates in CPLOP. Looking at the Recall as it compares to the Precision for $\alpha = 0.99$ allows us to visualize the tradeoffs we make when picking a k value. Labeled within each datapoint is the k value at that point

decrease which we intend to mean the increase in the raw value of the percentage. Additionally, values in the tables represent the proportion of the three metrics, but are easily interpreted as percentages. Section 6.2 explains the meaning of each precision (P), recall (R), and F-measure (F_1).

Overall, with $\alpha = 0.00$, Figure 6.6 illustrates that the resolution by union algorithm consistently performs better. For $k = 7$ and $\alpha = 0.00$, Table 6.1 shows that using the resolution by unions algorithm performs with 76.4% accuracy with meanwise and resolution by winner and intersection respectively achieving 73.2%, 65.9%, and 74.7 accuracy%. Poorly represented species, like the Cat, Chicken, and Seagull did not benefit from the resolution by union algorithm, each achieving no classifications,

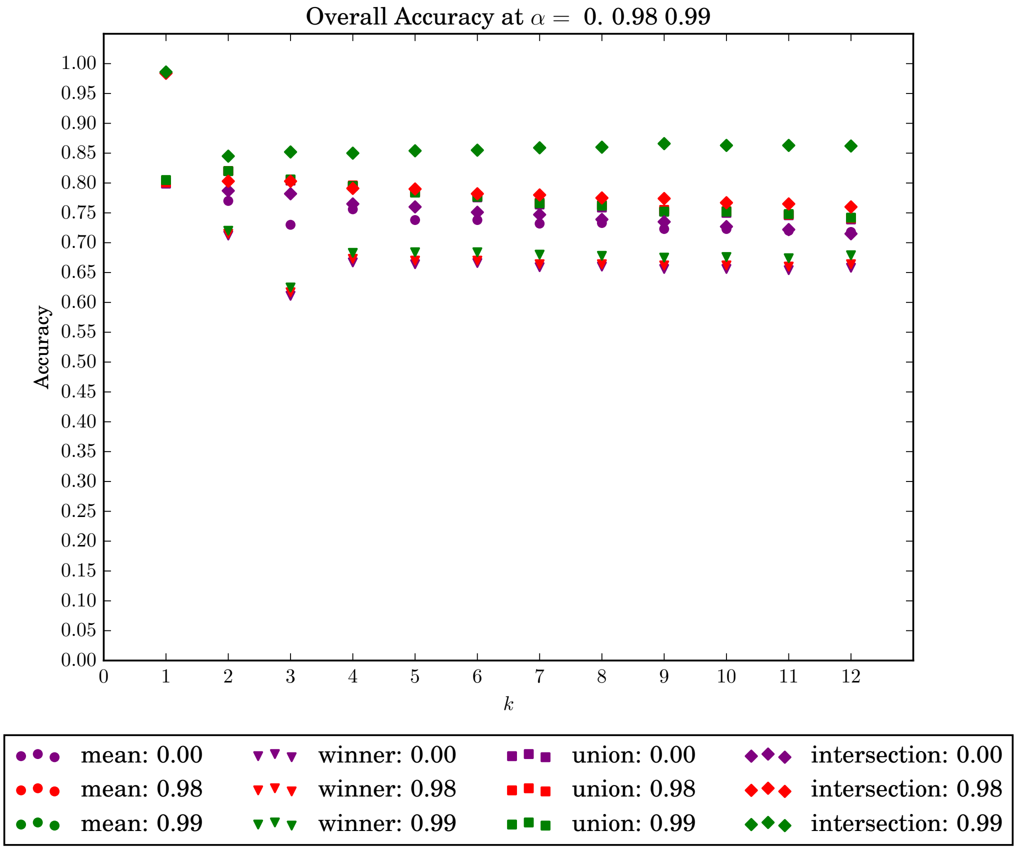


Figure 6.9: Shown is the accuracy of all classifications performed with CPLOP across the four different algorithms. We find that the accuracy of certain resolution algorithms perform better with higher α values.

correct or otherwise.

Once we restrict with a somewhat loose threshold of 0.98, overall we see that the intersection method provides the best accuracy, improving on non-thresholded values. For $k = 7$ and $\alpha = 0.98$, the intersection algorithm achieves 78.0% accuracy, while resolution by winner and union respectively achieve 66.4% and 76.7% accuracy.

Table 6.2 shows that a handful of poorly represented species achieved slightly better results when $\alpha = 0.98$. Notably, the intersection algorithm F-measure increased slightly for Wild Turkey, Cat, and Chicken on the order of 3%.

Unfortunately, the meanwise algorithm fails to classify when we use a large enough α and thus we have ommited the results in Tables 6.2 and 6.3. In certain cells of the

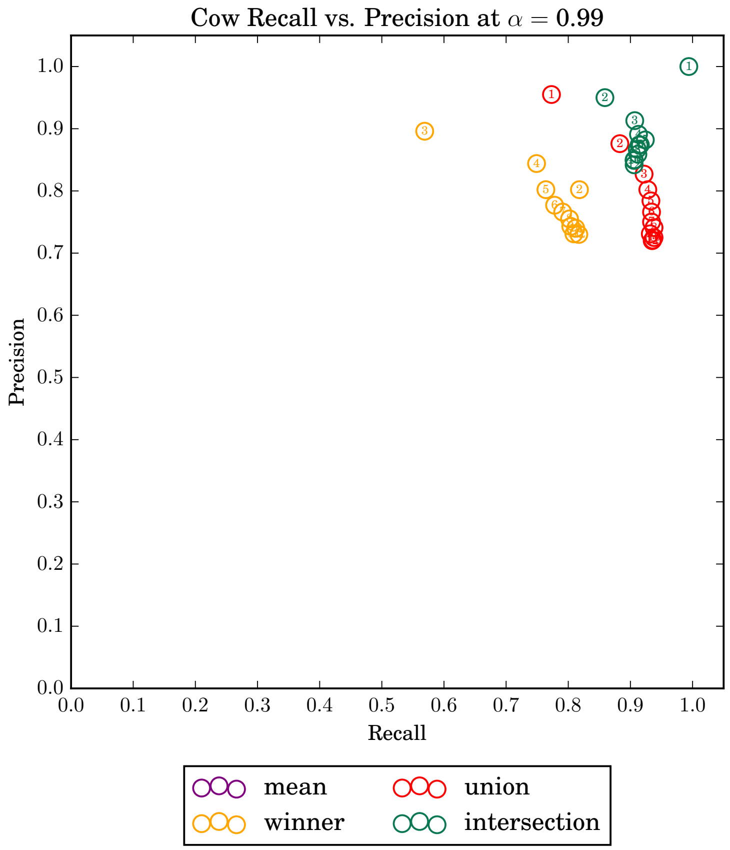


Figure 6.10: There are 1838 Cow isolates in CPLOP. Increasing the α for a species with this many isolates made minimal improvements to the accuracy on all but the resolution by intersection algorithm, which, when compared to Figure 6.8 noticeably improved.

tables, including Table 6.1, empty values in either P or F_1 mean no classifications were made of that species.

Restricting with $\alpha = 0.99$, our definition of strain differentiation, overall accuracy improves more with resolution by intersection and less so with resolutions by winner and union, garnering 85.9%, 68.0%, and 76.6% accuracy respectively. Again, meanwise resolution fails to produce any classifications.

For poorly represented species, we see some similar improvements for P, R, and F_1, but also some exceptions. Wild Turkey for example, improves by about 2%-3% for resolutions by winner and union and 11% for resolution by intersection, while Cat

Table 6.1: Precision (P), Recall (R), and F-Measure (F_1) overall and for particular species at k=7, $\alpha = 0.00$.

		Meanwise			Winner		
Host species	|Isolates|	P	R	F_1	P	R	F_1
Overall	4682	0.732	0.732	0.732	0.659	0.659	0.659
Human	1471	0.857	0.922	0.888	0.771	0.861	0.814
Cow	1718	0.757	0.885	0.816	0.744	0.776	0.760
Pigeon	194	0.420	0.242	0.307	0.280	0.253	0.266
Dog	149	0.596	0.436	0.504	0.449	0.356	0.397
Wild Turkey	72	0.383	0.250	0.303	0.277	0.181	0.219
Chicken	40	0.182	0.050	0.078	0.143	0.075	0.098
Cat	39	0.571	0.308	0.400	0.438	0.359	0.395
Bat	37	0.857	0.973	0.911	0.692	0.973	0.809
Seagull	11	0.000	0.000	0.000	0.000	0.000	0.000

		Union			Intersection		
Host species	|Isolates|	P	R	F_1	P	R	F_1
Overall	4682	0.764	0.764	0.764	0.747	0.747	0.747
Human	1471	0.843	0.930	0.884	0.839	0.925	0.880
Cow	1718	0.736	0.944	0.827	0.769	0.882	0.822
Pigeon	194	0.569	0.170	0.262	0.470	0.284	0.354
Dog	149	0.761	0.450	0.566	0.649	0.497	0.563
Wild Turkey	72	0.688	0.306	0.424	0.510	0.361	0.423
Chicken	40	0.000	0.000	0.000	0.250	0.100	0.143
Cat	39	0.889	0.410	0.561	0.571	0.308	0.400
Bat	37	0.857	0.973	0.911	0.857	0.973	0.911
Seagull	11		0.000			0.000	

decreases by 3% for resolution by winner, but improves by 6% and 47% for resolution by union and intersection.

6.3.4 Underrepresented Species

Some species had worse accuracy than the overall accuracy. In particular, species such as Chicken with only 40 isolates representing it showed similar leveling of accuracy for $k > 5$, but had far poorer accuracy, as shown in Figure 6.11. For $k > 5$, the accuracy of classifying chicken ranges from as low as 10% to a peak of 26%. The classification accuracy for many species in CPLOP heavily relies on its representation

Table 6.2: Precision (P), Recall (R), and F-Measure (F_1) overall and for particular species at k=7, $\alpha = 0.98$.

		Winner		
Host species	|Isolates|	P	R	F_1
Overall	4682	0.664	0.664	0.664
Human	1471	0.773	0.865	0.816
Cow	1718	0.749	0.777	0.763
Pigeon	194	0.287	0.254	0.269
Dog	149	0.448	0.349	0.392
Wild Turkey	72	0.308	0.222	0.258
Chicken	40	0.150	0.075	0.100
Cat	39	0.467	0.359	0.406
Bat	37	0.692	0.973	0.809
Seagull	11	0.000	0.000	0.000

		Union			Intersection		
Host species	|Isolates|	P	R	F_1	P	R	F_1
Overall	4682	0.767	0.767	0.767	0.780	0.780	0.780
Human	1471	0.845	0.930	0.885	0.876	0.950	0.912
Cow	1718	0.742	0.943	0.831	0.799	0.894	0.844
Pigeon	194	0.538	0.181	0.271	0.521	0.333	0.406
Dog	149	0.756	0.456	0.569	0.698	0.536	0.606
Wild Turkey	72	0.697	0.319	0.438	0.571	0.387	0.461
Chicken	40	0.000	0.000	0.000	0.308	0.121	0.174
Cat	39	0.889	0.410	0.561	0.632	0.353	0.453
Bat	37	0.857	0.973	0.911	0.878	0.973	0.923
Seagull	11		0.000		0.000	0.000	0.000

in CPLOP.

One notable exception is the Bat. In everyone application of our k-NN algorithms, Bat has above 95% accuracy. It is possible that due to their small size and relative dietary segregation from the surrounding environment that the strains of *E. coli* stay particularly unique. It may also be a quirk of the fact that each isolate comes from a single host animal, making it difficult to draw conclusions from such results.

Table 6.3: Precision (P), Recall (R), and F-Measure (F_1) overall and for particular species at k=7, $\alpha = 0.99$.

		Winner		
Host species	|Isolates|	P	R	F_1
Overall	4682	0.680	0.680	0.680
Human	1471	0.780	0.872	0.823
Cow	1718	0.766	0.791	0.778
Pigeon	194	0.314	0.263	0.286
Dog	149	0.527	0.401	0.455
Wild Turkey	72	0.320	0.222	0.262
Chicken	40	0.167	0.100	0.125
Cat	39	0.433	0.333	0.376
Bat	37	0.720	0.973	0.828
Seagull	11	0.429	0.273	0.334

		Union			Intersection		
Host species	|Isolates|	P	R	F_1	P	R	F_1
Overall	4682	0.766	0.766	0.766	0.859	0.859	0.859
Human	1471	0.843	0.925	0.882	0.926	0.979	0.952
Cow	1718	0.750	0.934	0.832	0.874	0.914	0.894
Pigeon	194	0.476	0.205	0.287	0.611	0.468	0.530
Dog	149	0.739	0.456	0.564	0.838	0.738	0.785
Wild Turkey	72	0.719	0.319	0.442	0.667	0.455	0.541
Chicken	40	0.000	0.000	0.000	0.000	0.000	0.000
Cat	39	0.938	0.385	0.546	0.909	0.588	0.714
Bat	37	0.837	0.973	0.900	0.973	1.000	0.986
Seagull	11	0.000	0.000	0.000		0.000	

6.4 Discussion

As a MST method, k-RAP performs well and offers many options for Cal Poly researchers to improve classification accuracy. Compared to the DBSCAN-based MST technique in Chapter 5, the accuracy of k-RAP performs slightly worse, but can classify the host species of every isolate in CPLOP, unlike the DBSCAN-based method, which marks nearly half of CPLOP as noise. Furthermore, researchers have an automatic and transparent method of host species classification, which can help them make more reasonable assertions when performing MST. Future work may be able to merge these two techniques together to improve classification speed and possi-

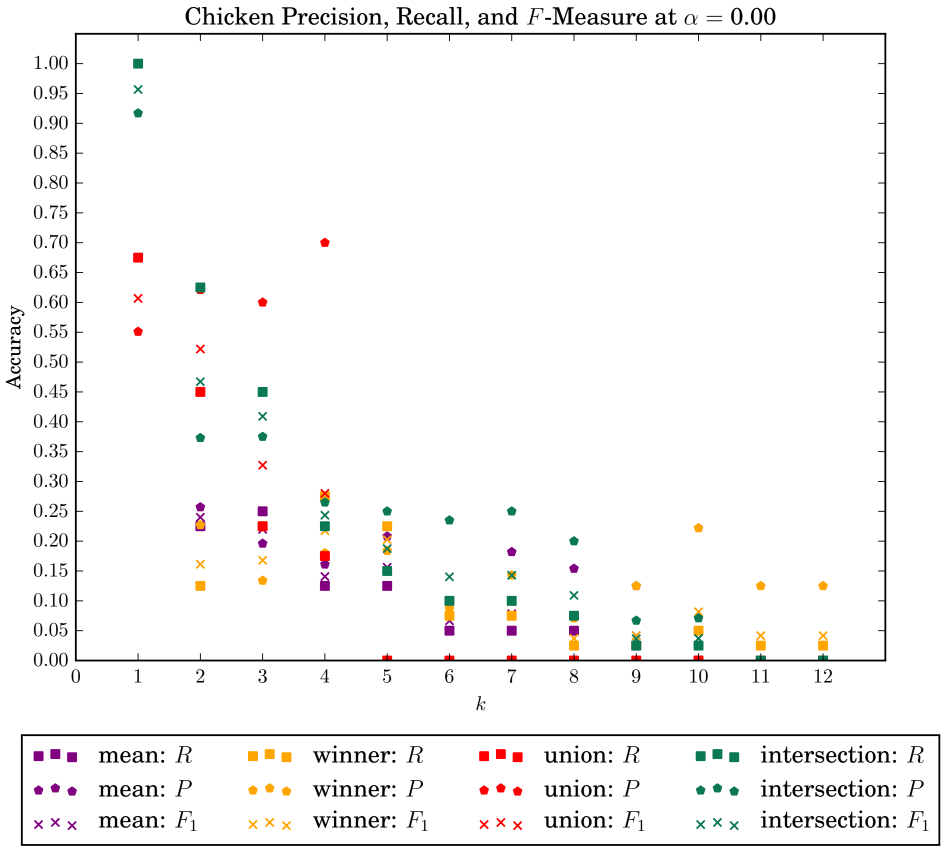


Figure 6.11: There are 40 Chicken isolates in CPLOP. Unfortunately, due to their low representation in CPLOP, classification accuracy is low.

bly improve MST accuracy. Such a "k-Nearest Clusters" classification methodology might make range queries somewhat faster, but it is unclear how it might affect the classification accuracy.

Chapter 7

IMPLEMENTATION

There are two software components that comprise this work, one for clustering with DBSCAN, which also creates the graphs of cluster metrics, and another for the k-Nearest Neighbors Resolution Algorithms for Pyroprints (k-RAP). Future work will incorporate k-RAP into CPLOP for microbial source tracking.

7.1 Graphing Cluster Metrics

The clustering code is comprised of two main parts: the clustering library from [31] and a series of scripts to graph the cluster metrics shown in Chapter 5. The clustering library queries a local instance of the CPLOP database, performs the clustering, and saves the resulting clusters in a series of `pickle`, `json`, and `csv` files. The graphing code then parses the clusters, counting the species in each cluster and calculating the purity in order to build the series of graphs. All of this code is written in Python 3 and uses `matplotlib` for graphing. The clustering library is described in [31]and found at `https://github.com/ejohns32/Thesis-Code`. The code written for this work does not change any of it, but instead adds code to control relevant pieces and graph the metrics.

7.2 The k-Nearest Neighbors Resolution Algorithms for Pyroprints

The code for the k-Nearest Neighbors Resolution Algorithms for Pyroprints (k-RAP) is a Java library that performs the evaluation described in Section 6.2. It queries a local instance of the CPLOP database and for each CPLOP isolate, it performs cross-validation with holdout for a range of k values, α values, and with each resolution strategy. It saves the classification results as a `csv`, which a Python 3 script graphs using `matplotlib`. The code resides at `https://github.com/jmcgover/k-rap` as well as relevant documentation.

Chapter 8

CONCLUSION

In order to combat the issue of contamination of publicly accessible water supplies, namely fecal contamination, the Cal Poly Biological Sciences Department teamed up with the Cal Poly Computer Science Department to build a library-based MST method, called CPLOP. Using the *E. coli* isolated from fecal samples as fecal indicator bacteria, Cal Poly students pyrosequence the PCR-amplified internal transcribed spacer regions of the *E. coli* and store the resulting vector, called a pyroprint, in the CPLOP database for later retrieval, analysis, and comparison. This thesis investigates two MST methodologies: a density-based clustering method built for CPLOP that clusters for bacterial strains in order to classify an unknown isolate and the k-Nearest Neighbors Resolution Algorithms for Pyroprints (k-RAP), a set of four k-nearest neighbors list resolution strategies for data with multiple comparison functions.

8.1 Clustering for Bacterial Strains

In this paper, we study the accuracy of a clustering-based MST approach which scales significantly better: the bacterial isolate information stored in CPLOP is clustered using an efficient density-based clustering technique. It clusters data by taking two parameters — the minimum number of neighbors and an ε range that those neighbors must be within to form a cluster — and performs range queries in a spatial index that performs $O(\log n)$ look-up on neighbors using a comparison metric. Compared to previous work [37, 42], it requires fewer comparisons to other isolates and computational resources, by being able to perform reasonably fast clusterings on a consumer laptop in minutes.

To ascertain how well this technique classifies, we build a notion of cluster purity. By calculating the proportion of the entire cluster that the most-plural host species makes up, we hope to understand how a density-based clustering algorithm clusters the CPLOP data. Furthermore, we inspect coverage and overall accuracy.

Results are that for `MinPts` between 1 and 5, respectively, we are able to cluster between 72.9% and 52.1% of the isolates in CPLOP, with between 51.0% and 41.2% falling into pure clusters and another 34.4% to 43.8% falling into clusters where their host species is the most dominant. Most clusters have high purity and low number of unique species, which is promising for using this for MST. Transient strains are also visible in the clustering technique, which will further aid the biologists working on CPLOP in researching transient strains and improving MST techniques. Future work will leverage other MST techniques designed for CPLOP against this to make up for the lack of coverage and transient strains.

8.2 k-RAP Effectiveness

Generally, when using k-NN, it is preferred to use single digit k values. Through our investigation of these various k-NN classification algorithms, we find that that general advice holds true. For our dataset, using $k \geq 5$ does not produce significantly different results. Choosing $k < 5$ is a dangerous notion, since it is likely that an outlier may make its way into the k-nearest neighbors list, confounding the results. Staying with $5 \leq k \leq 9$ appears to be a safe and reasonable option, providing a good balance between accuracy and filtering of isolates.

Outside of this study, we choose to differentiate between strains of *E. coli* using $\alpha = 0.99$. It appears that using this value is advantageous. There were, however, some exceptions to those results, motivating us to consider non-thresholded k-nearest neighbors lists when classifying an unknown isolate.

The four resolution algorithms — meanwise, winner, union, and intersection — each have their own quirks and behaviors as we alter k and α.

Meanwise, which currently uses the Euclidean norm to resolve different metrics, did not respond to the α threshold and completely stopped classifying anything for α near 1. This is very likely due to Euclidean norm mapping $([0, 1], \ldots, [0, 1]) \to [0, \sqrt{1 + \cdots + 1}]$. To get around this, we multiplied the resulting norm by a factor of $\sqrt{2}$, which may have unexpected results. We may investigate this further, or choose a more natural norming method, like arithmetic or geometric mean. With no α filtering, it performed third best with an overall 73.2% classification accuracy.

Winner performs worst, classifying accurately between 65% and 68% of the time. Some alterations to this algorithm may make it more reliable, such as only counting the species that appear in all lists.

Unionwise performs very well. Without filtering the k-nearest neighbors lists by α, we find that the unionwise method classifies best, with an overall accuracy of 76.4%. However, once we add in α filtering, the unionwise does not improve, staying relatively close to 76%.

Intersection performs best when we use α. This is likely due to the "list" actually being a set of common isolates. Overall, the accuracy was 74.7%, 78%, and 85.9% for α =0.00 (no filtering), 0.98, and 0.99 respectively.

Overall, we find that the intersection algorithm performs the best and recommend moving forward with it. While unionwise did perform well, it did not respond well to thresholding and still did not perform as well as the intersection algorithm overall. Meanwise and winner may be more useful with previously mentioned modifications and we may investigate these in the future.

Poorly distributed representation of species and environmental incomparabilities are issues endemic to library-based MST. CPLOP has an overabundance of Cow and

Human isolates, and an underrepresentation of many of the species in the database. This dilutes the k-nearest neighbors list considerably for species like the Chicken and Cat.

Library population issues aside, environmental limitations are another concern for accuracy. Nearly every sample in the library comes from a 30 mile radius around Cal Poly, making the collected pyroprints potentially incomparable to pyroprints collected from a different region.

8.3 Future Work

Future work should incorporate k-RAP into CPLOP, study the DBSCAN clustering method using the overall clustering entropy, and investigate whether combining the strain-based and isolate-based approach improves MST and is more efficient on the CPLOP database. Incorporating k-RAP into CPLOP will provide researchers the ability to make transparent, repeatable assertions as to the host species of an unknown isolate. Investigating clustering entropy can give us more insight into the makeup of clusters and extending this concept to k-RAP classifications, to measure how "close the competition is" between host species in the k-nearest neighbors lists. Merging strains into the k-RAP methods will reduce the number of comparisons, since querying against the database will involve querying against clusters of isolates, as opposed to isolates themselves. Whether this benefits the accuracy of MST with CPLOP needs to be investigated.

8.3.1 CPLOP Incorporation of k-RAP

Future work needs to incorporate k-RAP into CPLOP directly, so Cal Poly researchers have direct access to the MST methodologies. CPLOP researchers showed interest in viewing the k-nearest neighbors lists when using k-RAP on an isolate.

Building an interactive MST workflow can give researchers a better insight into host species determinations and help them avoid having to painstakingly perform MST by hand. Future work needs to at least make k-RAP available to researchers in CPLOP and should consider building an visually interactive way of using it.

8.3.2 Entropy

Entropy is another validity measure that for clusterings, represents the "degree to which each cluster consists of objects of a single class" [70] and for k-RAP can tell us how contentious the host species determination was. Similar to the cluster and clustering purity, cluster and clustering entropy can give us an idea of the nature of bacterial strains created by a clustering method. Ideally, a good clustering method minimizes the clustering entropy.

Consider a cluster C, consisting of datapoints with class labels from $\mathcal{L}$. Given a class label $L \in \mathcal{L}$, the value $P_C(L)$ calculates the proportion of datapoints in C that have class label L. The individual *cluster entropy* is:

$$e(C) = \sum_{L \in \mathcal{L}} P_C(L) \log_2 P_C(L) \tag{8.1}$$

In addition to computing the entropy of individual clusters we want to have an understanding of the overall entropy on the entire dataset for a given clustering. As before, given a *clustering* $\mathcal{C} = \{C_1, \ldots, C_n\}$ on a dataset, we define the size $\mathcal{M}$ of the set of clusters:

$$\mathcal{M} = \sum_{i=1}^{n} |C_i| \tag{8.2}$$

The overall *clustering entropy* is:

$$\sum_{i=1}^{n} \frac{|C_i|}{\mathcal{M}} \cdot \nu(C_i) \tag{8.3}$$

One can think of (8.3) as a form of weighted arithmetic mean of the individual entropies. Larger clusters, as a result, affect this overall value more. Future work should use this metric to compare clusterings with different parameters or methodologies, seeking to minimize the entropy, and extend the analysis to k-RAP classifications.

8.3.3 k-Nearest Clusters

Combining the strain-based and isolate-based method of classification is a natural next step. Using k-RAP as a simple fallback method when the strain-based method fails to classify an unknown host species isolate is one approach. Incorporating strains directly into k-RAP is yet another, wherein the neighbors in the k-nearest neighbors list may be either isolates or strains — which we can represent as clusters. As a result, we would instead have a k-Nearest Clusters classification algorithm, through which, we can still apply the resolution strategies from k-RAP. Future work should consider different cluster weighting methods for the host species in clusters that appear in the k-nearest neighbors list and determine whether any of them are useful for MST.

8.3.4 Efficiency Study

The efficiency of the classification methodologies in this work and any combinations thereof needs to be investigated as well as their performance on the hardware that supports the CPLOP database. Offline determination of strains, be it through clustering or otherwise, can occur offline, speeding up the determination of strain membership during MST. Combining the two approaches in this work into a k-Nearest Clusters methodology may also speed up MST. k-NN, and k-RAP as a result, compares the unknown isolate to every datapoint in the database, but if instead the datapoints might be clusters, fewer comparisons may need to be made.

INDEX

BIBLIOGRAPHY

[1] 2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops, BIBMW 2012, Philadelphia, USA, October 4-7, 2012. IEEE, 2012.

[2] C. C. Adams. Using Hadoop to Identify False Positives in Bacterial Strain Typing from DNA Fingerprints. California Polytechnic State University, San Luis Obispo, 2016.

[3] J. M. Albert, J. Munakata-Marr, L. Tenorio, and R. L. Siegrist. Statistical evaluation of bacterial source tracking data obtained by rep-PCR DNA fingerprinting of Escherichia coli. Environmental science & technology, 37(20):4554–4560, 2003.

[4] S. D. Bay. Combining nearest neighbor classifiers through multiple feature subsets. In Shavlik [64], pages 37–45.

[5] S. D. Bay. Nearest neighbor classification from multiple feature subsets. Intell. Data Anal., 3(3):191–209, 1999.

[6] L. Belanche-Muñoz and A. R. Blanch. Machine learning methods for microbial source tracking. Environmental Modelling & Software, 23(6):741–750, 2008.

[7] N. Bhatia and Vandana. Survey of nearest neighbor techniques. CoRR, abs/1007.0085, 2010.

[8] G. Bitton. Microbial indicators of fecal contamination. Wastewater Microbiology, Third Edition, pages 153–171, 2005.

[9] M. W. Black, J. VanderKelen, A. Montana, A. Dekhtyar, E. Neal, A. Goodman, and C. L. Kitts. Pyroprinting: A rapid and flexible genotypic

fingerprinting method for typing bacterial strains. Journal of Microbiological Methods, 105:121 – 129, 2014.

[10] D. Brandt, A. Montana, B. Somers, M. Black, A. Goodman, and C. Kitts. Pyroprinting sensitivity analysis on the GPU. In 2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops, BIBMW 2012, Philadelphia, USA, October 4-7, 2012 [1], pages 951–953.

[11] Cal Poly. Cal Poly Github, 2016. `http://www.github.com/CalPoly`.

[12] K. R. CLARKE. Non-parametric multivariate analyses of changes in community structure. Australian journal of ecology, 18(1):117–143, 1993.

[13] T. M. Cover and P. E. Hart. Nearest neighbor pattern classification. IEEE Trans. Information Theory, 13(1):21–27, 1967.

[14] T. R. Desmarais, H. M. Solo-Gabriele, and C. J. Palmer. Influence of soil on fecal indicator organisms in a tidally influenced subtropical environment. Applied and environmental microbiology, 68(3):1165–1172, 2002.

[15] C. Desrosiers and G. Karypis. A comprehensive survey of neighborhood-based recommendation methods. In Ricci et al. [56], pages 107–144.

[16] J. W. Dickerson Jr. Evaluation, Development and Improvement of Genotypic, Phenotypic and Chemical Microbia PhD thesis, Virginia Polytechnic Institute and State University, 2008.

[17] J. R. Dillard. Demographics and Transfer of Escherichia coli Within Bos taurus Populations. Master's thesis, California Polytechnic State University, San Luis Obispo, 2015.

[18] J. R. Dillard, J. J. VanderKelen, J. D. Kent, A. D. Frey, P. J. McCreesh,

D. Britton, T. Branck, M. W. Black, and C. L. Kitts. E. coli Strain Demographics and Transmission in Cattle. Strain, 10(1):11, 2013.

[19] W. Ding, T. Washio, H. Xiong, G. Karypis, B. M. Thuraisingham, D. J. Cook, and X. Wu, editors. 13th IEEE International Conference on Data Mining Workshops, ICDM Workshops, TX, USA, December 7-10, 2013. IEEE Computer Society, 2013.

[20] P. B. Eckburg, E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill, K. E. Nelson, and D. A. Relman. Diversity of the human intestinal microbial flora. science, 308(5728):1635–1638, 2005.

[21] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In Kdd, volume 96, pages 226–231, 1996.

[22] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data MiningI, pages 226–231. AAAI Press, 1996.

[23] R. A. Finkel and J. L. Bentley. Quad trees a data structure for retrieval on composite keys. Acta informatica, 4(1):1–9, 1974.

[24] E. Fix and J. L. Hodges Jr. Discriminatory analysis-nonparametric discrimination: consistency properties. Technical report, DTIC Document, 1951.

[25] E. Fix and J. L. Hodges Jr. Discriminatory analysis-nonparametric discrimination: Small sample performance. Technical report, DTIC Document, 1952.

[26] Q. Hua, A. Ji, and Q. He. Multiple real-valued K nearest neighbor classifiers system by feature grouping. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10-13 October 2010 [52], pages 3922–3925.

[27] J. Huan, S. Miyano, A. Shehu, X. T. Hu, B. Ma, S. Rajasekaran, V. K. Gombar, M. Schapranow, I. Yoo, J. Zhou, B. Chen, V. Pai, and B. G. Pierce, editors. 2015 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2015, Washington, DC, USA, November 9-12, 2015. IEEE Computer Society, 2015.

[28] R. Hubbard, G. Newton, and G. Hill. Water quality and the grazing animal. Journal of animal science, 82(13_suppl):E255–E263, 2004.

[29] International Conference on Machine Learning and Cybernetics, ICMLC 2010, Qingdao, China, July 11-14, 2010, Proceedings. IEEE, 2010.

[30] L. Jiang, Z. Cai, D. Wang, and S. Jiang. Survey of improving k-nearest-neighbor for classification. In Lei [35], pages 679–683.

[31] E. Johnson. Density-Based Clustering of High-Dimensional DNA Fingerprints for Library-Dependent Microbial Source Tracking. Master's thesis, California Polytechnic State University, San Luis Obispo, 2015.

[32] W. Juan. Multiple nearest neighbor classifiers system based on feature perturbation by mutual information. In International Conference on Machine Learning and Cybernetics, ICMLC 2010, Qingdao, China, July 11-14, 2010, Proceedings [29], pages 247–251.

[33] J. Kent, M. Alvarado, J. VanderKelen, A. Montana, J. Soliman, A. Dekhtyar, A. Goodman, C. Kitts, and M. Black. Pyroprinting: Novel

Pyrosequencing-Based Method for Studying E. coli Diversity and Microbial Source Tracking (779.8). The FASEB Journal, 28(1 Supplement):779–8, 2014.

[34] D. T. Larose. Discovering knowledge in data: an introduction to data mining. John Wiley & Sons, 2005.

[35] J. Lei, editor. Fourth International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2007, 24-27 August 2007, Haikou, Hainan, China, Proceedings, Volume 1. IEEE Computer Society, 2007.

[36] W. Li, D. Raoult, and P.-E. Fournier. Bacterial strain typing in the genomic era. FEMS Microbiology Reviews, 33(5):892–916, 2009.

[37] J. D. McGovern, A. Dekhtyar, C. Kitts, M. Black, J. Vanderkelen, and A. Goodman. Leveraging the k-nearest neighbors classification algorithm for microbial source tracking using a bacterial DNA fingerprint library. In Huan et al. [27], pages 1694–1701.

[38] J. D. McGovern, E. Johnson, A. Dekhtyar, M. Black, C. Kitts, and J. Vanderkelen. Library-based microbial source tracking via strain identification. In Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB 2016, Seattle, WA, USA, October 2-5, 2016 [51], pages 364–373.

[39] R. N. McLeod. A Proof of Concept for Crowdsourcing Color Perception Experiments. California Polytechnic State University, San Luis Obispo, 2014.

[40] D. J. Meagher. Octree encoding: A new technique for the representation, manipulation and display of arbitrary 3-d objects by computer. Electrical and Systems Engineering Department Rensseiaer Polytechnic Institute Image Processing Laboratory, 1980.

[41] A. Montana. Algorithms for Library-Based Microbial Source Tracking. Master's thesis, California Polytechnic State University San Luis Obispo, 2013.

[42] A. Montana, A. Dekhtyar, M. Black, C. Kitts, and A. Goodman. Ontological hierarchical clustering for library-based microbial source tracking. In Ding et al. [19], pages 568–576.

[43] A. Montana, A. Dekhtyar, E. Neal, M. Black, and C. Kitts. Chronology-sensitive hierarchical clustering of pyrosequenced DNA samples of e. coli: A case study. In Wu et al. [74], pages 155–159.

[44] A. Montana, A. Dekhtyar, E. Neal, M. Black, and C. Kitts. Investigating temporal strain diversity in human e. coli populations using pyroprinting: A novel strain identification method. Technical report, Technical report, California Polytechnic State University, San Luis Obispo, CA, 2012.

[45] C. Moritz, D. Shapiro, and C. Pann. Application of Pyroprinting for Source Tracking of E. coli in Pennington Creek. California Polytechnic State University, San Luis Obispo, 2015.

[46] E. Neal, C. Sabatini, W. Tang, M. Black, and C. Kitts. Demographics of E. coli Strains in the Human Gut Using Pyroprints: A Novel MST Method. In CSUPERB, Poster. Jan, 2012.

[47] E. R. Neal. Escherichia coli Strain Diversity in Humans: Effects of Sampling Effort and Methodology. Master's thesis, California Polytechnic State University, San Luis Obispo, 2013.

[48] M. Neave, H. Luter, A. Padovan, S. Townsend, X. Schobben, and K. Gibb. Multiple approaches to microbial source tracking in tropical northern australia. MicrobiologyOpen, 3(6):860–874, 2014.

[49] J. Nguyen, J. Vanderkelen, M. Black, and C. Kitts. Investigating the Dominant Escherichia coli Strain in Lambs and Ewes Using Pyroprinting: A Novel Method for Strain Identification. California Polytechnic State University, San Luis Obispo, 2015.

[50] X. Ning, C. Desrosiers, and G. Karypis. A comprehensive survey of neighborhood-based recommendation methods. In Ricci et al. [55], pages 37–76.

[51] Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB 2016, Seattle, WA, USA, October 2-5, 2016. ACM, 2016.

[52] Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10-13 October 2010. IEEE, 2010.

[53] V. Ramachandran, editor. Proceedings of the Fourth Annual ACM/SIGACT-SIAM Symposium on Discrete Algorithms, 25-27 January 1993, Austin, Texas. ACM/SIAM, 1993.

[54] S. Ranka, iTamer Kahveci, and M. Singh, editors. ACM International Conference on Bioinformatics, Computational Biology and Biomedicine, BCB' 12, Orlando, FL, USA - October 08 - 10, 2012. ACM, 2012.

[55] F. Ricci, L. Rokach, and B. Shapira, editors. Recommender Systems Handbook. Springer, 2015.

[56] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors. Recommender Systems Handbook. Springer, 2011.

[57] C. Ricketts. Cal Poly Library of Pyroprints: Quality Control Analysis and Web Development. Master's thesis, California Polytechnic State University, San Luis Obispo, 2014.

[58] K. Ritter, E. Carruthers, C. Carson, R. Ellender, V. Harwood, K. Kingsley, C. Nakatsu, M. Sadowsky, B. Shear, B. West, et al. Assessment of statistical methods used in library-based approaches to microbial source tracking. J Water Health, 1:209–223, 2003.

[59] S. Rogers and J. Haines. Detecting and mitigating the environmental impact of fecal pathogens originating from confined animal feeding operations: review. United States Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory, 2005.

[60] M. Ronaghi, M. Uhlén, and P. Nyrén. A sequencing method based on real-time pyrophosphate. Science, 281(5375):363–365, 1998.

[61] D. Sargeant, W. R. Kammin, and S. Collyard. Review and critique of current microbial source tracking (mst) techniques. Environmental Assessment Program, Washington State Department of Ecology, 2011.

[62] T. M. Scott, J. B. Rose, T. M. Jenkins, S. R. Farrah, and J. Lukasik. Microbial source tracking: Current methodology and future directions. APPLIED AND ENVIRONMENTAL MICROBIOLOGY, pages 5796–5803, 2002.

[63] D. Shapiro, J. Kent, M. Zuleta, C. Kitts, M. Black, and J. VanderKelen. Source Tracking of Fecal Contamination Along San Luis Obispo (SLO) Creek. The FASEB Journal, 29(1 Supplement):575–12, 2015.

[64] J. W. Shavlik, editor. Proceedings of the Fifteenth International Conference on Machine Learning (ICML 1998), Madison, Wisconsin, USA, July 24-27, 1998. Morgan Kaufmann, 1998.

[65] D. Shealy. Exploration of PyroPrinting for Environmental Forensics. Technical report, California Polytechnic State University, San Luis Obispo, California, June 2012.

[66] J. M. Simpson, J. W. Santo Domingo, and D. J. Reasoner. Microbial source tracking: state of the science. Environmental science & technology, 36(24):5279–5288, 2002.

[67] J. L. Soliman. CPLOP: The Cal Poly Library of Pyroprints. Master's thesis, California Polytechnic State University San Luis Obispo, 2013.

[68] J. L. Soliman, A. Dekhtyar, J. Vanderkellen, A. Montana, M. Black, E. Neal, K. Webb, C. Kitts, and A. Goodman. Microbial source tracking by molecular fingerprinting. In Ranka et al. [54], pages 617–619.

[69] J. Stewart, R. Ellender, J. Gooch, S. Jiang, S. Myoda, and S. Weisberg. Recommendations for microbial source tracking: lessons from a methods comparison study. J Water Health, 1:225–231, 2003.

[70] P.-N. Tan et al. Introduction to Data Mining. Pearson Education India, 2006.

[71] J. J. VanderKelen, R. D. Mitchell, A. Laubscher, M. W. Black, A. L. Goodman, A. K. Montana, A. M. Dekhtyar, R. Jimenez-Flores, and C. L. Kitts. Short Communication: Typing and Tracking Bacillaceae in Raw Milk and Milk Powder Using Pyroprinting. Journal of Dairy Science, 99(1):146–151, 2016.

[72] L. Wang, Q. Hua, X. Wang, and Q. Chen. Combination of multiple nearest neighbor classifiers based on feature subset clustering method. In Yeung et al. [75], pages 538–547.

[73] K. Webb. Cplop-cal poly's library of pyroprints. California Polytechnic State University, San Luis Obispo, 2011.

[74] F. Wu, M. J. Zaki, S. Morishita, Y. Pan, S. Wong, A. Christianson, and X. Hu, editors. IEEE International Conference on Bioinformatics and Biomedicine,

BIBM 2011, Atlanta, GA, USA, November 12-15, , 2011. IEEE Computer Society, 2011.

[75] D. S. Yeung, Z. Liu, X. Wang, and H. Yan, editors. Advances in Machine Learning and Cybernetics, 4th International Conference, ICMLC 2005, Guangzhou, China, August 18-21, 2005, Revised Selected Papers, volume 3930 of Lecture Notes in Computer Science. Springer, 2006.

[76] P. N. Yianilos. Data structures and algorithms for nearest neighbor search in general metric spaces. In Ramachandran [53], pages 311–321.

APPENDICES

Appendix A

CLUSTER COUNTS

Below are the host species counts of each cluster for the clustering performed.

A.1 Cluster Counts for `MinPts` = 1

Host species	**Count**	**Proportion**
Human	168	0.468
Cow	166	0.462
Pig	7	0.019
Pigeon	5	0.014
Sheep	3	0.008
Ground Squirrel	3	0.008
Chicken	3	0.008
Dog	2	0.006
Horse	1	0.003
Coyote	1	0.003
Total	**359**	**1.000**

Table A.1: Cluster 1 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	248	1.000
Total	**248**	**1.000**

Table A.2: Cluster 2 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	163	0.734
Pigeon	12	0.054
Dog	10	0.045
Human	8	0.036
Wild Turkey	8	0.036
Ground Squirrel	5	0.023
Cliff Sparrow	4	0.018
Horse	3	0.014
Chicken	3	0.014
Sheep	2	0.009
Turkey Vulture	1	0.005
Pelican	1	0.005
Deer Mouse	1	0.005
Pig	1	0.005
Total	**222**	**1.000**

Table A.3: Cluster 3 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	199	1.000
Total	**199**	**1.000**

Table A.4: Cluster 4 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	161	0.843
Cow	15	0.079
Mountain Lion	8	0.042
Deer	6	0.031
Horse	1	0.005
Total	**191**	**1.000**

Table A.5: Cluster 5 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	101	0.962
Coyote	2	0.019
Mountain Lion	2	0.019
Total	**105**	**1.000**

Table A.6: Cluster 6 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	76	0.760
Dog	8	0.080
Ground Squirrel	5	0.050
Grey Fox	4	0.040
Mountain Lion	2	0.020
Pigeon	2	0.020
Cat	1	0.010
Coyote	1	0.010
Orangutan	1	0.010
Total	**100**	**1.000**

Table A.7: Cluster 7 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	99	1.000
Total	**99**	**1.000**

Table A.8: Cluster 8 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	50	0.526
Cow	41	0.432
Cliff Sparrow	3	0.032
Pigeon	1	0.011
Total	**95**	**1.000**

Table A.9: Cluster 9 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	53	0.883
Sheep	3	0.050
Cliff Sparrow	2	0.033
Cat	2	0.033
Total	**60**	**1.000**

Table A.10: Cluster 10 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	49	0.980
Cow	1	0.020
Total	**50**	**1.000**

Table A.11: Cluster 11 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Dog	48	1.000
Total	**48**	**1.000**

Table A.12: Cluster 12 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Bat	36	0.837
Cow	6	0.140
Human	1	0.023
Total	**43**	**1.000**

Table A.13: Cluster 13 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	41	1.000
Total	**41**	**1.000**

Table A.14: Cluster 14 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	23	0.575
Sheep	11	0.275
Pig	3	0.075
Seagull	1	0.025
Mallard Duck	1	0.025
Western Kingbird	1	0.025
Total	**40**	**1.000**

Table A.15: Cluster 15 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Dog	31	0.939
Wild Turkey	2	0.061
Total	**33**	**1.000**

Table A.16: Cluster 16 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	31	0.969
Wild Turkey	1	0.031
Total	**32**	**1.000**

Table A.17: Cluster 17 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Ground Squirrel	18	0.621
Human	2	0.069
Red-shoulder Hawk	2	0.069
Great Horned Owl	2	0.069
Horse	2	0.069
Cow	1	0.034
Red Shoulder Hawk	1	0.034
Cat	1	0.034
Total	**29**	**1.000**

Table A.18: Cluster 18 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Sheep	21	1.000
Total	**21**	**1.000**

Table A.19: Cluster 19 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	21	1.000
Total	**21**	**1.000**

Table A.20: Cluster 20 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	21	1.000
Total	**21**	**1.000**

Table A.21: Cluster 21 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	19	1.000
Total	**19**	**1.000**

Table A.22: Cluster 22 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	18	0.947
Pigeon	1	0.053
Total	**19**	**1.000**

Table A.23: Cluster 23 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Sheep	17	0.895
Chicken	2	0.105
Total	**19**	**1.000**

Table A.24: Cluster 24 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	18	1.000
Total	**18**	**1.000**

Table A.25: Cluster 25 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	10	0.588
Ground Squirrel	7	0.412
Total	**17**	**1.000**

Table A.26: Cluster 26 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	16	1.000
Total	**16**	**1.000**

Table A.27: Cluster 27 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	14	0.875
Ground Squirrel	2	0.125
Total	**16**	**1.000**

Table A.28: Cluster 28 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	15	1.000
Total	**15**	**1.000**

Table A.29: Cluster 29 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	15	1.000
Total	**15**	**1.000**

Table A.30: Cluster 30 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	14	1.000
Total	**14**	**1.000**

Table A.31: Cluster 31 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	8	0.571
Wild Turkey	6	0.429
Total	**14**	**1.000**

Table A.32: Cluster 32 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	13	1.000
Total	**13**	**1.000**

Table A.33: Cluster 33 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.34: Cluster 34 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.35: Cluster 35 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	12	1.000
Total	**12**	**1.000**

Table A.36: Cluster 36 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Ground Squirrel	12	1.000
Total	**12**	**1.000**

Table A.37: Cluster 37 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	10	0.909
Ground Squirrel	1	0.091
Total	**11**	**1.000**

Table A.38: Cluster 38 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	4	0.364
Wild Turkey	3	0.273
Red-Winged Blackbird	2	0.182
Pigeon	1	0.091
Pig	1	0.091
Total	**11**	**1.000**

Table A.39: Cluster 39 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	3	0.273
Wild Turkey	2	0.182
Elephant Seal	2	0.182
California Sea Lion	2	0.182
Sea Lion	2	0.182
Total	**11**	**1.000**

Table A.40: Cluster 40 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	11	1.000
Total	**11**	**1.000**

Table A.41: Cluster 41 of 380 for `MinPts`=1

Host species	Count	Proportion
Deer	10	1.000
Total	**10**	**1.000**

Table A.42: Cluster 42 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	9	1.000
Total	**9**	**1.000**

Table A.43: Cluster 43 of 380 for `MinPts`=1

Host species	Count	Proportion
Pelican	4	0.444
Red Throated Loon	2	0.222
Common Loon	2	0.222
Common Murre	1	0.111
Total	**9**	**1.000**

Table A.44: Cluster 44 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	9	1.000
Total	**9**	**1.000**

Table A.45: Cluster 45 of 380 for `MinPts`=1

Host species	Count	Proportion
Ground Squirrel	9	1.000
Total	**9**	**1.000**

Table A.46: Cluster 46 of 380 for `MinPts`=1

Host species	Count	Proportion
Pigeon	5	0.625
Cliff Sparrow	3	0.375
Total	**8**	**1.000**

Table A.47: Cluster 47 of 380 for `MinPts`=1

Host species	Count	Proportion
Cat	8	1.000
Total	**8**	**1.000**

Table A.48: Cluster 48 of 380 for `MinPts`=1

Host species	Count	Proportion
Sea Otter	8	1.000
Total	**8**	**1.000**

Table A.49: Cluster 49 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	7	0.875
Cow	1	0.125
Total	**8**	**1.000**

Table A.50: Cluster 50 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	8	1.000
Total	**8**	**1.000**

Table A.51: Cluster 51 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	7	1.000
Total	**7**	**1.000**

Table A.52: Cluster 52 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Ground Squirrel	7	1.000
Total	**7**	**1.000**

Table A.53: Cluster 53 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	7	1.000
Total	**7**	**1.000**

Table A.54: Cluster 54 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	5	0.714
Pig	2	0.286
Total	**7**	**1.000**

Table A.55: Cluster 55 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	6	0.857
Wild Turkey	1	0.143
Total	**7**	**1.000**

Table A.56: Cluster 56 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	6	0.857
Chicken	1	0.143
Total	**7**	**1.000**

Table A.57: Cluster 57 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	5	0.714
Ground Squirrel	1	0.143
Pelican	1	0.143
Total	**7**	**1.000**

Table A.58: Cluster 58 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Ground Squirrel	7	1.000
Total	**7**	**1.000**

Table A.59: Cluster 59 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	7	1.000
Total	**7**	**1.000**

Table A.60: Cluster 60 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	5	0.833
Chicken	1	0.167
Total	**6**	**1.000**

Table A.61: Cluster 61 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	6	1.000
Total	**6**	**1.000**

Table A.62: Cluster 62 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Bear	5	0.833
Ground Squirrel	1	0.167
Total	**6**	**1.000**

Table A.63: Cluster 63 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	5	0.833
Wild Turkey	1	0.167
Total	**6**	**1.000**

Table A.64: Cluster 64 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	6	1.000
Total	**6**	**1.000**

Table A.65: Cluster 65 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Ground Squirrel	6	1.000
Total	**6**	**1.000**

Table A.66: Cluster 66 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	6	1.000
Total	**6**	**1.000**

Table A.67: Cluster 67 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	6	1.000
Total	**6**	**1.000**

Table A.68: Cluster 68 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	6	1.000
Total	**6**	**1.000**

Table A.69: Cluster 69 of 380 for `MinPts`=1

Host species	Count	Proportion
Mountain Lion	6	1.000
Total	**6**	**1.000**

Table A.70: Cluster 70 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.71: Cluster 71 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.72: Cluster 72 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.73: Cluster 73 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.74: Cluster 74 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.75: Cluster 75 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pig	2	0.400
Mountain Lion	2	0.400
Human	1	0.200
Total	**5**	**1.000**

Table A.76: Cluster 76 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	5	1.000
Total	**5**	**1.000**

Table A.77: Cluster 77 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	5	1.000
Total	**5**	**1.000**

Table A.78: Cluster 78 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	5	1.000
Total	**5**	**1.000**

Table A.79: Cluster 79 of 380 for `MinPts`=1

Host species	Count	Proportion
Pigeon	5	1.000
Total	**5**	**1.000**

Table A.80: Cluster 80 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.81: Cluster 81 of 380 for `MinPts`=1

Host species	Count	Proportion
Pig	5	1.000
Total	**5**	**1.000**

Table A.82: Cluster 82 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.83: Cluster 83 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.84: Cluster 84 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.85: Cluster 85 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	5	1.000
Total	**5**	**1.000**

Table A.86: Cluster 86 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.87: Cluster 87 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.88: Cluster 88 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.89: Cluster 89 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Opossum	4	1.000
Total	**4**	**1.000**

Table A.90: Cluster 90 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.91: Cluster 91 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	2	0.500
Tree Swallow	2	0.500
Total	**4**	**1.000**

Table A.92: Cluster 92 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.93: Cluster 93 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.94: Cluster 94 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.95: Cluster 95 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.96: Cluster 96 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Mountain Lion	4	1.000
Total	**4**	**1.000**

Table A.97: Cluster 97 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Rabbit	4	1.000
Total	**4**	**1.000**

Table A.98: Cluster 98 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.99: Cluster 99 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.100: Cluster 100 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	3	0.750
Dog	1	0.250
Total	**4**	**1.000**

Table A.101: Cluster 101 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	4	1.000
Total	**4**	**1.000**

Table A.102: Cluster 102 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	0.750
Cliff Sparrow	1	0.250
Total	**4**	**1.000**

Table A.103: Cluster 103 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	2	0.500
Chicken	1	0.250
Cow	1	0.250
Total	**4**	**1.000**

Table A.104: Cluster 104 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.105: Cluster 105 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	4	1.000
Total	**4**	**1.000**

Table A.106: Cluster 106 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Eared Grebe	2	0.500
Red Tailed Hawk	1	0.250
Mallard Duck	1	0.250
Total	**4**	**1.000**

Table A.107: Cluster 107 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cliff Sparrow	2	0.500
Cow	2	0.500
Total	**4**	**1.000**

Table A.108: Cluster 108 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.109: Cluster 109 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.110: Cluster 110 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	4	1.000
Total	**4**	**1.000**

Table A.111: Cluster 111 of 380 for `MinPts`=1

Host species	Count	Proportion
Pigeon	4	1.000
Total	**4**	**1.000**

Table A.112: Cluster 112 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	3	0.750
Ground Squirrel	1	0.250
Total	**4**	**1.000**

Table A.113: Cluster 113 of 380 for `MinPts`=1

Host species	Count	Proportion
Seagull	2	0.500
California Sea Lion	2	0.500
Total	**4**	**1.000**

Table A.114: Cluster 114 of 380 for `MinPts`=1

Host species	Count	Proportion
Horse	2	0.500
Pigeon	1	0.250
Cow	1	0.250
Total	**4**	**1.000**

Table A.115: Cluster 115 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.116: Cluster 116 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	4	1.000
Total	**4**	**1.000**

Table A.117: Cluster 117 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.118: Cluster 118 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pig	4	1.000
Total	**4**	**1.000**

Table A.119: Cluster 119 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.120: Cluster 120 of 380 for `MinPts`=1

Host species	Count	Proportion
Wild Turkey	4	1.000
Total	**4**	**1.000**

Table A.121: Cluster 121 of 380 for `MinPts`=1

Host species	Count	Proportion
Pig	4	1.000
Total	**4**	**1.000**

Table A.122: Cluster 122 of 380 for `MinPts`=1

Host species	Count	Proportion
White Crowned Sparrow	4	1.000
Total	**4**	**1.000**

Table A.123: Cluster 123 of 380 for `MinPts`=1

Host species	Count	Proportion
Pigeon	4	1.000
Total	**4**	**1.000**

Table A.124: Cluster 124 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.125: Cluster 125 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.126: Cluster 126 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	2	0.500
Human	1	0.250
Cow	1	0.250
Total	**4**	**1.000**

Table A.127: Cluster 127 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	4	1.000
Total	**4**	**1.000**

Table A.128: Cluster 128 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cat	3	1.000
Total	**3**	**1.000**

Table A.129: Cluster 129 of 380 for `MinPts`=1

Host species	Count	Proportion
Chicken	3	1.000
Total	**3**	**1.000**

Table A.130: Cluster 130 of 380 for `MinPts`=1

Host species	Count	Proportion
Sheep	3	1.000
Total	**3**	**1.000**

Table A.131: Cluster 131 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	3	1.000
Total	**3**	**1.000**

Table A.132: Cluster 132 of 380 for `MinPts`=1

Host species	Count	Proportion
Dog	3	1.000
Total	**3**	**1.000**

Table A.133: Cluster 133 of 380 for `MinPts`=1

Host species	Count	Proportion
Sea Otter	2	0.667
Pigeon	1	0.333
Total	**3**	**1.000**

Table A.134: Cluster 134 of 380 for `MinPts`=1

Host species	Count	Proportion
Cliff Sparrow	2	0.667
Pigeon	1	0.333
Total	**3**	**1.000**

Table A.135: Cluster 135 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.136: Cluster 136 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.137: Cluster 137 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.138: Cluster 138 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	3	1.000
Total	**3**	**1.000**

Table A.139: Cluster 139 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.140: Cluster 140 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	0.667
Sheep	1	0.333
Total	**3**	**1.000**

Table A.141: Cluster 141 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Common Loon	2	0.667
Red Tailed Hawk	1	0.333
Total	**3**	**1.000**

Table A.142: Cluster 142 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pig	2	0.667
Cow	1	0.333
Total	**3**	**1.000**

Table A.143: Cluster 143 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Wild Turkey	3	1.000
Total	**3**	**1.000**

Table A.144: Cluster 144 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	3	1.000
Total	**3**	**1.000**

Table A.145: Cluster 145 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.146: Cluster 146 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Seagull	3	1.000
Total	**3**	**1.000**

Table A.147: Cluster 147 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
American Kestrel	2	0.667
Red Tailed Hawk	1	0.333
Total	**3**	**1.000**

Table A.148: Cluster 148 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	0.667
Red-shoulder Hawk	1	0.333
Total	**3**	**1.000**

Table A.149: Cluster 149 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	3	1.000
Total	**3**	**1.000**

Table A.150: Cluster 150 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	3	1.000
Total	**3**	**1.000**

Table A.151: Cluster 151 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	3	1.000
Total	**3**	**1.000**

Table A.152: Cluster 152 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	3	1.000
Total	**3**	**1.000**

Table A.153: Cluster 153 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	3	1.000
Total	**3**	**1.000**

Table A.154: Cluster 154 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	3	1.000
Total	**3**	**1.000**

Table A.155: Cluster 155 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.156: Cluster 156 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	0.667
Human	1	0.333
Total	**3**	**1.000**

Table A.157: Cluster 157 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.158: Cluster 158 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.159: Cluster 159 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.160: Cluster 160 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	3	1.000
Total	**3**	**1.000**

Table A.161: Cluster 161 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.162: Cluster 162 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Horse	3	1.000
Total	**3**	**1.000**

Table A.163: Cluster 163 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.164: Cluster 164 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.165: Cluster 165 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.166: Cluster 166 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cat	3	1.000
Total	**3**	**1.000**

Table A.167: Cluster 167 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pig	2	0.667
Cow	1	0.333
Total	**3**	**1.000**

Table A.168: Cluster 168 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.169: Cluster 169 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.170: Cluster 170 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.171: Cluster 171 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.172: Cluster 172 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Wild Turkey	3	1.000
Total	**3**	**1.000**

Table A.173: Cluster 173 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Dog	3	1.000
Total	**3**	**1.000**

Table A.174: Cluster 174 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	3	1.000
Total	**3**	**1.000**

Table A.175: Cluster 175 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	3	1.000
Total	**3**	**1.000**

Table A.176: Cluster 176 of 380 for `MinPts`=1

Host species	Count	Proportion
Wild Turkey	3	1.000
Total	**3**	**1.000**

Table A.177: Cluster 177 of 380 for `MinPts`=1

Host species	Count	Proportion
Dog	2	0.667
Cat	1	0.333
Total	**3**	**1.000**

Table A.178: Cluster 178 of 380 for `MinPts`=1

Host species	Count	Proportion
Horse	2	0.667
Cow	1	0.333
Total	**3**	**1.000**

Table A.179: Cluster 179 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pig	3	1.000
Total	**3**	**1.000**

Table A.180: Cluster 180 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Ground Squirrel	3	1.000
Total	**3**	**1.000**

Table A.181: Cluster 181 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Sheep	3	1.000
Total	**3**	**1.000**

Table A.182: Cluster 182 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Dog	3	1.000
Total	**3**	**1.000**

Table A.183: Cluster 183 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Ground Squirrel	2	1.000
Total	**2**	**1.000**

Table A.184: Cluster 184 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.185: Cluster 185 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cliff Sparrow	2	1.000
Total	**2**	**1.000**

Table A.186: Cluster 186 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cat	2	1.000
Total	**2**	**1.000**

Table A.187: Cluster 187 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	1	0.500
Pigeon	1	0.500
Total	**2**	**1.000**

Table A.188: Cluster 188 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.189: Cluster 189 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Sheep	2	1.000
Total	**2**	**1.000**

Table A.190: Cluster 190 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.191: Cluster 191 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.192: Cluster 192 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.193: Cluster 193 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.194: Cluster 194 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.195: Cluster 195 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.196: Cluster 196 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.197: Cluster 197 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.198: Cluster 198 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.199: Cluster 199 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.200: Cluster 200 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Golden Eagle	2	1.000
Total	**2**	**1.000**

Table A.201: Cluster 201 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Wild Turkey	2	1.000
Total	**2**	**1.000**

Table A.202: Cluster 202 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.203: Cluster 203 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.204: Cluster 204 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cliff Sparrow	2	1.000
Total	**2**	**1.000**

Table A.205: Cluster 205 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.206: Cluster 206 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Dog	2	1.000
Total	**2**	**1.000**

Table A.207: Cluster 207 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.208: Cluster 208 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.209: Cluster 209 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.210: Cluster 210 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Sheep	2	1.000
Total	**2**	**1.000**

Table A.211: Cluster 211 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.212: Cluster 212 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.213: Cluster 213 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cat	2	1.000
Total	**2**	**1.000**

Table A.214: Cluster 214 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Dog	2	1.000
Total	**2**	**1.000**

Table A.215: Cluster 215 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
White Crowned Sparrow	2	1.000
Total	**2**	**1.000**

Table A.216: Cluster 216 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.217: Cluster 217 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.218: Cluster 218 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.219: Cluster 219 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.220: Cluster 220 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.221: Cluster 221 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.222: Cluster 222 of 380 for `MinPts`=1

Host species	Count	Proportion
Dog	2	1.000
Total	**2**	**1.000**

Table A.223: Cluster 223 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.224: Cluster 224 of 380 for `MinPts`=1

Host species	Count	Proportion
Cat	2	1.000
Total	**2**	**1.000**

Table A.225: Cluster 225 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.226: Cluster 226 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.227: Cluster 227 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.228: Cluster 228 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	1	0.500
Wild Turkey	1	0.500
Total	**2**	**1.000**

Table A.229: Cluster 229 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.230: Cluster 230 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.231: Cluster 231 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	2	1.000
Total	**2**	**1.000**

Table A.232: Cluster 232 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Dog	2	1.000
Total	**2**	**1.000**

Table A.233: Cluster 233 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.234: Cluster 234 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.235: Cluster 235 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.236: Cluster 236 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.237: Cluster 237 of 380 for `MinPts`=1

Host species	Count	Proportion
Cat	2	1.000
Total	**2**	**1.000**

Table A.238: Cluster 238 of 380 for `MinPts`=1

Host species	Count	Proportion
Cat	2	1.000
Total	**2**	**1.000**

Table A.239: Cluster 239 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.240: Cluster 240 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.241: Cluster 241 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Chicken	2	1.000
Total	**2**	**1.000**

Table A.242: Cluster 242 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Opossum	2	1.000
Total	**2**	**1.000**

Table A.243: Cluster 243 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Ground Squirrel	2	1.000
Total	**2**	**1.000**

Table A.244: Cluster 244 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.245: Cluster 245 of 380 for `MinPts`=1

Host species	Count	Proportion
Ground Squirrel	2	1.000
Total	**2**	**1.000**

Table A.246: Cluster 246 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.247: Cluster 247 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.248: Cluster 248 of 380 for `MinPts`=1

Host species	Count	Proportion
White Crowned Sparrow	2	1.000
Total	**2**	**1.000**

Table A.249: Cluster 249 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.250: Cluster 250 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.251: Cluster 251 of 380 for `MinPts`=1

Host species	Count	Proportion
Horse	2	1.000
Total	**2**	**1.000**

Table A.252: Cluster 252 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.253: Cluster 253 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.254: Cluster 254 of 380 for `MinPts`=1

Host species	Count	Proportion
Pigeon	2	1.000
Total	**2**	**1.000**

Table A.255: Cluster 255 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Ground Squirrel	2	1.000
Total	**2**	**1.000**

Table A.256: Cluster 256 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.257: Cluster 257 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	2	1.000
Total	**2**	**1.000**

Table A.258: Cluster 258 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Dog	2	1.000
Total	**2**	**1.000**

Table A.259: Cluster 259 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Mountain Lion	2	1.000
Total	**2**	**1.000**

Table A.260: Cluster 260 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.261: Cluster 261 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.262: Cluster 262 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Elephant Seal	2	1.000
Total	**2**	**1.000**

Table A.263: Cluster 263 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Clark Grebe	2	1.000
Total	**2**	**1.000**

Table A.264: Cluster 264 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Dog	2	1.000
Total	**2**	**1.000**

Table A.265: Cluster 265 of 380 for `MinPts`=1

Host species	Count	Proportion
Pigeon	2	1.000
Total	**2**	**1.000**

Table A.266: Cluster 266 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.267: Cluster 267 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.268: Cluster 268 of 380 for `MinPts`=1

Host species	Count	Proportion
Cat	2	1.000
Total	**2**	**1.000**

Table A.269: Cluster 269 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.270: Cluster 270 of 380 for `MinPts`=1

Host species	Count	Proportion
Wild Turkey	2	1.000
Total	**2**	**1.000**

Table A.271: Cluster 271 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.272: Cluster 272 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.273: Cluster 273 of 380 for `MinPts`=1

Host species	Count	Proportion
Dog	2	1.000
Total	**2**	**1.000**

Table A.274: Cluster 274 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.275: Cluster 275 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.276: Cluster 276 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Bobcat	2	1.000
Total	**2**	**1.000**

Table A.277: Cluster 277 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.278: Cluster 278 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.279: Cluster 279 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.280: Cluster 280 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.281: Cluster 281 of 380 for `MinPts`=1

Host species	Count	Proportion
Chicken	2	1.000
Total	**2**	**1.000**

Table A.282: Cluster 282 of 380 for `MinPts`=1

Host species	Count	Proportion
Pigeon	1	0.500
Cow	1	0.500
Total	**2**	**1.000**

Table A.283: Cluster 283 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.284: Cluster 284 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.285: Cluster 285 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	2	1.000
Total	**2**	**1.000**

Table A.286: Cluster 286 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.287: Cluster 287 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Chicken	2	1.000
Total	**2**	**1.000**

Table A.288: Cluster 288 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.289: Cluster 289 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.290: Cluster 290 of 380 for `MinPts`=1

Host species	Count	Proportion
Opossum	2	1.000
Total	**2**	**1.000**

Table A.291: Cluster 291 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.292: Cluster 292 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.293: Cluster 293 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.294: Cluster 294 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	1	0.500
Wild Turkey	1	0.500
Total	**2**	**1.000**

Table A.295: Cluster 295 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Wild Turkey	2	1.000
Total	**2**	**1.000**

Table A.296: Cluster 296 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.297: Cluster 297 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cat	2	1.000
Total	**2**	**1.000**

Table A.298: Cluster 298 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Chicken	2	1.000
Total	**2**	**1.000**

Table A.299: Cluster 299 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.300: Cluster 300 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.301: Cluster 301 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.302: Cluster 302 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Crow	1	0.500
Horse	1	0.500
Total	**2**	**1.000**

Table A.303: Cluster 303 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Dog	2	1.000
Total	**2**	**1.000**

Table A.304: Cluster 304 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Opossum	2	1.000
Total	**2**	**1.000**

Table A.305: Cluster 305 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Wild Turkey	1	0.500
Cow	1	0.500
Total	**2**	**1.000**

Table A.306: Cluster 306 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.307: Cluster 307 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	2	1.000
Total	**2**	**1.000**

Table A.308: Cluster 308 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.309: Cluster 309 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.310: Cluster 310 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.311: Cluster 311 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	1	0.500
Horse	1	0.500
Total	**2**	**1.000**

Table A.312: Cluster 312 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.313: Cluster 313 of 380 for `MinPts`=1

Host species	Count	Proportion
Ground Squirrel	2	1.000
Total	**2**	**1.000**

Table A.314: Cluster 314 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.315: Cluster 315 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.316: Cluster 316 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Surf Scoter	2	1.000
Total	**2**	**1.000**

Table A.317: Cluster 317 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Ground Squirrel	2	1.000
Total	**2**	**1.000**

Table A.318: Cluster 318 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Dog	2	1.000
Total	**2**	**1.000**

Table A.319: Cluster 319 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Sheep	2	1.000
Total	**2**	**1.000**

Table A.320: Cluster 320 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Dog	2	1.000
Total	**2**	**1.000**

Table A.321: Cluster 321 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.322: Cluster 322 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	2	1.000
Total	**2**	**1.000**

Table A.323: Cluster 323 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Racoon	2	1.000
Total	**2**	**1.000**

Table A.324: Cluster 324 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.325: Cluster 325 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.326: Cluster 326 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.327: Cluster 327 of 380 for `MinPts`=1

Host species	Count	Proportion
Pig	2	1.000
Total	**2**	**1.000**

Table A.328: Cluster 328 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.329: Cluster 329 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.330: Cluster 330 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.331: Cluster 331 of 380 for `MinPts`=1

Host species	Count	Proportion
White Crowned Sparrow	2	1.000
Total	**2**	**1.000**

Table A.332: Cluster 332 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.333: Cluster 333 of 380 for `MinPts`=1

Host species	Count	Proportion
Ground Squirrel	2	1.000
Total	**2**	**1.000**

Table A.334: Cluster 334 of 380 for `MinPts`=1

Host species	Count	Proportion
Pigeon	2	1.000
Total	**2**	**1.000**

Table A.335: Cluster 335 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.336: Cluster 336 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	1	0.500
Horse	1	0.500
Total	**2**	**1.000**

Table A.337: Cluster 337 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cat	2	1.000
Total	**2**	**1.000**

Table A.338: Cluster 338 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.339: Cluster 339 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Ground Squirrel	2	1.000
Total	**2**	**1.000**

Table A.340: Cluster 340 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.341: Cluster 341 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.342: Cluster 342 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	1	0.500
Cow	1	0.500
Total	**2**	**1.000**

Table A.343: Cluster 343 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cliff Sparrow	2	1.000
Total	**2**	**1.000**

Table A.344: Cluster 344 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Horse	2	1.000
Total	**2**	**1.000**

Table A.345: Cluster 345 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.346: Cluster 346 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Ground Squirrel	2	1.000
Total	**2**	**1.000**

Table A.347: Cluster 347 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	2	1.000
Total	**2**	**1.000**

Table A.348: Cluster 348 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.349: Cluster 349 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.350: Cluster 350 of 380 for `MinPts`=1

Host species	Count	Proportion
Ground Squirrel	2	1.000
Total	**2**	**1.000**

Table A.351: Cluster 351 of 380 for `MinPts`=1

Host species	Count	Proportion
Pigeon	2	1.000
Total	**2**	**1.000**

Table A.352: Cluster 352 of 380 for `MinPts`=1

Host species	Count	Proportion
Horse	2	1.000
Total	**2**	**1.000**

Table A.353: Cluster 353 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.354: Cluster 354 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.355: Cluster 355 of 380 for `MinPts`=1

Host species	Count	Proportion
White Crowned Sparrow	2	1.000
Total	**2**	**1.000**

Table A.356: Cluster 356 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.357: Cluster 357 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.358: Cluster 358 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.359: Cluster 359 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.360: Cluster 360 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	2	1.000
Total	**2**	**1.000**

Table A.361: Cluster 361 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Wild Turkey	2	1.000
Total	**2**	**1.000**

Table A.362: Cluster 362 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.363: Cluster 363 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Sheep	2	1.000
Total	**2**	**1.000**

Table A.364: Cluster 364 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.365: Cluster 365 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.366: Cluster 366 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.367: Cluster 367 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.368: Cluster 368 of 380 for `MinPts`=1

Host species	Count	Proportion
Pigeon	2	1.000
Total	**2**	**1.000**

Table A.369: Cluster 369 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.370: Cluster 370 of 380 for `MinPts`=1

Host species	Count	Proportion
Human	2	1.000
Total	**2**	**1.000**

Table A.371: Cluster 371 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.372: Cluster 372 of 380 for `MinPts`=1

Host species	Count	Proportion
Ground Squirrel	1	0.500
Cow	1	0.500
Total	**2**	**1.000**

Table A.373: Cluster 373 of 380 for `MinPts`=1

Host species	Count	Proportion
Deer	1	0.500
Pigeon	1	0.500
Total	**2**	**1.000**

Table A.374: Cluster 374 of 380 for `MinPts`=1

Host species	Count	Proportion
Cow	2	1.000
Total	**2**	**1.000**

Table A.375: Cluster 375 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Human	2	1.000
Total	**2**	**1.000**

Table A.376: Cluster 376 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Cow	2	1.000
Total	**2**	**1.000**

Table A.377: Cluster 377 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Dog	2	1.000
Total	**2**	**1.000**

Table A.378: Cluster 378 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
California Sea Lion	2	1.000
Total	**2**	**1.000**

Table A.379: Cluster 379 of 380 for `MinPts`=1

Host species	**Count**	**Proportion**
Pigeon	2	1.000
Total	**2**	**1.000**

Table A.380: Cluster 380 of 380 for `MinPts`=1

A.2 Cluster Counts for `MinPts` = 2

Host species	Count	Proportion
Human	168	0.468
Cow	166	0.462
Pig	7	0.019
Pigeon	5	0.014
Sheep	3	0.008
Ground Squirrel	3	0.008
Chicken	3	0.008
Dog	2	0.006
Horse	1	0.003
Coyote	1	0.003
Total	**359**	**1.000**

Table A.381: Cluster 1 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	248	1.000
Total	**248**	**1.000**

Table A.382: Cluster 2 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	163	0.734
Pigeon	12	0.054
Dog	10	0.045
Human	8	0.036
Wild Turkey	8	0.036
Ground Squirrel	5	0.023
Cliff Sparrow	4	0.018
Horse	3	0.014
Chicken	3	0.014
Sheep	2	0.009
Turkey Vulture	1	0.005
Pelican	1	0.005
Deer Mouse	1	0.005
Pig	1	0.005
Total	**222**	**1.000**

Table A.383: Cluster 3 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	199	1.000
Total	**199**	**1.000**

Table A.384: Cluster 4 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	161	0.843
Cow	15	0.079
Mountain Lion	8	0.042
Deer	6	0.031
Horse	1	0.005
Total	**191**	**1.000**

Table A.385: Cluster 5 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	101	0.962
Coyote	2	0.019
Mountain Lion	2	0.019
Total	**105**	**1.000**

Table A.386: Cluster 6 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	76	0.760
Dog	8	0.080
Ground Squirrel	5	0.050
Grey Fox	4	0.040
Mountain Lion	2	0.020
Pigeon	2	0.020
Cat	1	0.010
Coyote	1	0.010
Orangutan	1	0.010
Total	**100**	**1.000**

Table A.387: Cluster 7 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	99	1.000
Total	**99**	**1.000**

Table A.388: Cluster 8 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Human	50	0.526
Cow	41	0.432
Cliff Sparrow	3	0.032
Pigeon	1	0.011
Total	**95**	**1.000**

Table A.389: Cluster 9 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	53	0.883
Sheep	3	0.050
Cliff Sparrow	2	0.033
Cat	2	0.033
Total	**60**	**1.000**

Table A.390: Cluster 10 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	49	0.980
Cow	1	0.020
Total	**50**	**1.000**

Table A.391: Cluster 11 of 183 for `MinPts`=2

Host species	Count	Proportion
Dog	48	1.000
Total	**48**	**1.000**

Table A.392: Cluster 12 of 183 for `MinPts`=2

Host species	Count	Proportion
Bat	36	0.837
Cow	6	0.140
Human	1	0.023
Total	**43**	**1.000**

Table A.393: Cluster 13 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	41	1.000
Total	**41**	**1.000**

Table A.394: Cluster 14 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	23	0.575
Sheep	11	0.275
Pig	3	0.075
Seagull	1	0.025
Mallard Duck	1	0.025
Western Kingbird	1	0.025
Total	**40**	**1.000**

Table A.395: Cluster 15 of 183 for `MinPts`=2

Host species	Count	Proportion
Dog	31	0.939
Wild Turkey	2	0.061
Total	**33**	**1.000**

Table A.396: Cluster 16 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	31	0.969
Wild Turkey	1	0.031
Total	**32**	**1.000**

Table A.397: Cluster 17 of 183 for `MinPts`=2

Host species	Count	Proportion
Ground Squirrel	18	0.621
Human	2	0.069
Red-shoulder Hawk	2	0.069
Great Horned Owl	2	0.069
Horse	2	0.069
Cow	1	0.034
Red Shoulder Hawk	1	0.034
Cat	1	0.034
Total	**29**	**1.000**

Table A.398: Cluster 18 of 183 for `MinPts`=2

Host species	Count	Proportion
Sheep	21	1.000
Total	**21**	**1.000**

Table A.399: Cluster 19 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	21	1.000
Total	**21**	**1.000**

Table A.400: Cluster 20 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	21	1.000
Total	**21**	**1.000**

Table A.401: Cluster 21 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	19	1.000
Total	**19**	**1.000**

Table A.402: Cluster 22 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	18	0.947
Pigeon	1	0.053
Total	**19**	**1.000**

Table A.403: Cluster 23 of 183 for `MinPts`=2

Host species	Count	Proportion
Sheep	17	0.895
Chicken	2	0.105
Total	**19**	**1.000**

Table A.404: Cluster 24 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	18	1.000
Total	**18**	**1.000**

Table A.405: Cluster 25 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	10	0.588
Ground Squirrel	7	0.412
Total	**17**	**1.000**

Table A.406: Cluster 26 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	16	1.000
Total	**16**	**1.000**

Table A.407: Cluster 27 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	14	0.875
Ground Squirrel	2	0.125
Total	**16**	**1.000**

Table A.408: Cluster 28 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	15	1.000
Total	**15**	**1.000**

Table A.409: Cluster 29 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	15	1.000
Total	**15**	**1.000**

Table A.410: Cluster 30 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	14	1.000
Total	**14**	**1.000**

Table A.411: Cluster 31 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	8	0.571
Wild Turkey	6	0.429
Total	**14**	**1.000**

Table A.412: Cluster 32 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	13	1.000
Total	**13**	**1.000**

Table A.413: Cluster 33 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.414: Cluster 34 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.415: Cluster 35 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	12	1.000
Total	**12**	**1.000**

Table A.416: Cluster 36 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Ground Squirrel	12	1.000
Total	**12**	**1.000**

Table A.417: Cluster 37 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	10	0.909
Ground Squirrel	1	0.091
Total	**11**	**1.000**

Table A.418: Cluster 38 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	4	0.364
Wild Turkey	3	0.273
Red-Winged Blackbird	2	0.182
Pigeon	1	0.091
Pig	1	0.091
Total	**11**	**1.000**

Table A.419: Cluster 39 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	3	0.273
Wild Turkey	2	0.182
Elephant Seal	2	0.182
California Sea Lion	2	0.182
Sea Lion	2	0.182
Total	**11**	**1.000**

Table A.420: Cluster 40 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	11	1.000
Total	**11**	**1.000**

Table A.421: Cluster 41 of 183 for `MinPts`=2

Host species	Count	Proportion
Deer	10	1.000
Total	**10**	**1.000**

Table A.422: Cluster 42 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	9	1.000
Total	**9**	**1.000**

Table A.423: Cluster 43 of 183 for `MinPts`=2

Host species	Count	Proportion
Pelican	4	0.444
Red Throated Loon	2	0.222
Common Loon	2	0.222
Common Murre	1	0.111
Total	**9**	**1.000**

Table A.424: Cluster 44 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	9	1.000
Total	**9**	**1.000**

Table A.425: Cluster 45 of 183 for `MinPts`=2

Host species	Count	Proportion
Ground Squirrel	9	1.000
Total	**9**	**1.000**

Table A.426: Cluster 46 of 183 for `MinPts`=2

Host species	Count	Proportion
Pigeon	5	0.625
Cliff Sparrow	3	0.375
Total	**8**	**1.000**

Table A.427: Cluster 47 of 183 for `MinPts`=2

Host species	Count	Proportion
Cat	8	1.000
Total	**8**	**1.000**

Table A.428: Cluster 48 of 183 for `MinPts`=2

Host species	Count	Proportion
Sea Otter	8	1.000
Total	**8**	**1.000**

Table A.429: Cluster 49 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Human	7	0.875
Cow	1	0.125
Total	**8**	**1.000**

Table A.430: Cluster 50 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	8	1.000
Total	**8**	**1.000**

Table A.431: Cluster 51 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	7	1.000
Total	**7**	**1.000**

Table A.432: Cluster 52 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Ground Squirrel	7	1.000
Total	**7**	**1.000**

Table A.433: Cluster 53 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	7	1.000
Total	**7**	**1.000**

Table A.434: Cluster 54 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Pigeon	5	0.714
Pig	2	0.286
Total	**7**	**1.000**

Table A.435: Cluster 55 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	6	0.857
Wild Turkey	1	0.143
Total	**7**	**1.000**

Table A.436: Cluster 56 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	6	0.857
Chicken	1	0.143
Total	**7**	**1.000**

Table A.437: Cluster 57 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Human	5	0.714
Ground Squirrel	1	0.143
Pelican	1	0.143
Total	**7**	**1.000**

Table A.438: Cluster 58 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Ground Squirrel	7	1.000
Total	**7**	**1.000**

Table A.439: Cluster 59 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	7	1.000
Total	**7**	**1.000**

Table A.440: Cluster 60 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Bear	5	0.833
Ground Squirrel	1	0.167
Total	**6**	**1.000**

Table A.441: Cluster 61 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	5	0.833
Wild Turkey	1	0.167
Total	**6**	**1.000**

Table A.442: Cluster 62 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	6	1.000
Total	**6**	**1.000**

Table A.443: Cluster 63 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Ground Squirrel	6	1.000
Total	**6**	**1.000**

Table A.444: Cluster 64 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Pigeon	6	1.000
Total	**6**	**1.000**

Table A.445: Cluster 65 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	6	1.000
Total	**6**	**1.000**

Table A.446: Cluster 66 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Human	6	1.000
Total	**6**	**1.000**

Table A.447: Cluster 67 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Mountain Lion	6	1.000
Total	**6**	**1.000**

Table A.448: Cluster 68 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Human	5	0.833
Chicken	1	0.167
Total	**6**	**1.000**

Table A.449: Cluster 69 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	6	1.000
Total	**6**	**1.000**

Table A.450: Cluster 70 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.451: Cluster 71 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.452: Cluster 72 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.453: Cluster 73 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.454: Cluster 74 of 183 for `MinPts`=2

Host species	Count	Proportion
Pig	2	0.400
Mountain Lion	2	0.400
Human	1	0.200
Total	**5**	**1.000**

Table A.455: Cluster 75 of 183 for `MinPts`=2

Host species	Count	Proportion
Pigeon	5	1.000
Total	**5**	**1.000**

Table A.456: Cluster 76 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.457: Cluster 77 of 183 for `MinPts`=2

Host species	Count	Proportion
Pigeon	5	1.000
Total	**5**	**1.000**

Table A.458: Cluster 78 of 183 for `MinPts`=2

Host species	Count	Proportion
Pig	5	1.000
Total	**5**	**1.000**

Table A.459: Cluster 79 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.460: Cluster 80 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.461: Cluster 81 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.462: Cluster 82 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.463: Cluster 83 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	5	1.000
Total	**5**	**1.000**

Table A.464: Cluster 84 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.465: Cluster 85 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.466: Cluster 86 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.467: Cluster 87 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.468: Cluster 88 of 183 for `MinPts`=2

Host species	Count	Proportion
Opossum	4	1.000
Total	**4**	**1.000**

Table A.469: Cluster 89 of 183 for `MinPts`=2

Host species	Count	Proportion
Pigeon	2	0.500
Tree Swallow	2	0.500
Total	**4**	**1.000**

Table A.470: Cluster 90 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.471: Cluster 91 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.472: Cluster 92 of 183 for `MinPts`=2

Host species	Count	Proportion
Mountain Lion	4	1.000
Total	**4**	**1.000**

Table A.473: Cluster 93 of 183 for `MinPts`=2

Host species	Count	Proportion
Rabbit	4	1.000
Total	**4**	**1.000**

Table A.474: Cluster 94 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.475: Cluster 95 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.476: Cluster 96 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Human	3	0.750
Dog	1	0.250
Total	**4**	**1.000**

Table A.477: Cluster 97 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Human	4	1.000
Total	**4**	**1.000**

Table A.478: Cluster 98 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	3	0.750
Cliff Sparrow	1	0.250
Total	**4**	**1.000**

Table A.479: Cluster 99 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.480: Cluster 100 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.481: Cluster 101 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	4	1.000
Total	**4**	**1.000**

Table A.482: Cluster 102 of 183 for `MinPts`=2

Host species	Count	Proportion
Eared Grebe	2	0.500
Red Tailed Hawk	1	0.250
Mallard Duck	1	0.250
Total	**4**	**1.000**

Table A.483: Cluster 103 of 183 for `MinPts`=2

Host species	Count	Proportion
Cliff Sparrow	2	0.500
Cow	2	0.500
Total	**4**	**1.000**

Table A.484: Cluster 104 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.485: Cluster 105 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.486: Cluster 106 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Human	4	1.000
Total	**4**	**1.000**

Table A.487: Cluster 107 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Pigeon	4	1.000
Total	**4**	**1.000**

Table A.488: Cluster 108 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Seagull	2	0.500
California Sea Lion	2	0.500
Total	**4**	**1.000**

Table A.489: Cluster 109 of 183 for `MinPts`=2

Host species	Count	Proportion
Pigeon	4	1.000
Total	**4**	**1.000**

Table A.490: Cluster 110 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.491: Cluster 111 of 183 for `MinPts`=2

Host species	Count	Proportion
Horse	2	0.500
Pigeon	1	0.250
Cow	1	0.250
Total	**4**	**1.000**

Table A.492: Cluster 112 of 183 for `MinPts`=2

Host species	Count	Proportion
Pig	4	1.000
Total	**4**	**1.000**

Table A.493: Cluster 113 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.494: Cluster 114 of 183 for `MinPts`=2

Host species	Count	Proportion
Wild Turkey	4	1.000
Total	**4**	**1.000**

Table A.495: Cluster 115 of 183 for `MinPts`=2

Host species	Count	Proportion
Pig	4	1.000
Total	**4**	**1.000**

Table A.496: Cluster 116 of 183 for `MinPts`=2

Host species	Count	Proportion
White Crowned Sparrow	4	1.000
Total	**4**	**1.000**

Table A.497: Cluster 117 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.498: Cluster 118 of 183 for `MinPts`=2

Host species	Count	Proportion
Pigeon	4	1.000
Total	**4**	**1.000**

Table A.499: Cluster 119 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.500: Cluster 120 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.501: Cluster 121 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.502: Cluster 122 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	3	0.750
Ground Squirrel	1	0.250
Total	**4**	**1.000**

Table A.503: Cluster 123 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Pigeon	2	0.500
Human	1	0.250
Cow	1	0.250
Total	**4**	**1.000**

Table A.504: Cluster 124 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.505: Cluster 125 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.506: Cluster 126 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Pigeon	2	0.500
Chicken	1	0.250
Cow	1	0.250
Total	**4**	**1.000**

Table A.507: Cluster 127 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Pigeon	4	1.000
Total	**4**	**1.000**

Table A.508: Cluster 128 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cat	3	1.000
Total	**3**	**1.000**

Table A.509: Cluster 129 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Chicken	3	1.000
Total	**3**	**1.000**

Table A.510: Cluster 130 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Sheep	3	1.000
Total	**3**	**1.000**

Table A.511: Cluster 131 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.512: Cluster 132 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Sea Otter	2	0.667
Pigeon	1	0.333
Total	**3**	**1.000**

Table A.513: Cluster 133 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cliff Sparrow	2	0.667
Pigeon	1	0.333
Total	**3**	**1.000**

Table A.514: Cluster 134 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.515: Cluster 135 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Human	3	1.000
Total	**3**	**1.000**

Table A.516: Cluster 136 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.517: Cluster 137 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Common Loon	2	0.667
Red Tailed Hawk	1	0.333
Total	**3**	**1.000**

Table A.518: Cluster 138 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Pig	2	0.667
Cow	1	0.333
Total	**3**	**1.000**

Table A.519: Cluster 139 of 183 for `MinPts`=2

Host species	Count	Proportion
Wild Turkey	3	1.000
Total	**3**	**1.000**

Table A.520: Cluster 140 of 183 for `MinPts`=2

Host species	Count	Proportion
Human	3	1.000
Total	**3**	**1.000**

Table A.521: Cluster 141 of 183 for `MinPts`=2

Host species	Count	Proportion
Seagull	3	1.000
Total	**3**	**1.000**

Table A.522: Cluster 142 of 183 for `MinPts`=2

Host species	Count	Proportion
American Kestrel	2	0.667
Red Tailed Hawk	1	0.333
Total	**3**	**1.000**

Table A.523: Cluster 143 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	3	1.000
Total	**3**	**1.000**

Table A.524: Cluster 144 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	3	1.000
Total	**3**	**1.000**

Table A.525: Cluster 145 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Human	3	1.000
Total	**3**	**1.000**

Table A.526: Cluster 146 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.527: Cluster 147 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.528: Cluster 148 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Human	3	1.000
Total	**3**	**1.000**

Table A.529: Cluster 149 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.530: Cluster 150 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Horse	3	1.000
Total	**3**	**1.000**

Table A.531: Cluster 151 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.532: Cluster 152 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.533: Cluster 153 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.534: Cluster 154 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.535: Cluster 155 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	3	1.000
Total	**3**	**1.000**

Table A.536: Cluster 156 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	3	1.000
Total	**3**	**1.000**

Table A.537: Cluster 157 of 183 for `MinPts`=2

Host species	Count	Proportion
Dog	3	1.000
Total	**3**	**1.000**

Table A.538: Cluster 158 of 183 for `MinPts`=2

Host species	Count	Proportion
Wild Turkey	3	1.000
Total	**3**	**1.000**

Table A.539: Cluster 159 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Dog	3	1.000
Total	**3**	**1.000**

Table A.540: Cluster 160 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Human	3	1.000
Total	**3**	**1.000**

Table A.541: Cluster 161 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Human	3	1.000
Total	**3**	**1.000**

Table A.542: Cluster 162 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.543: Cluster 163 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Wild Turkey	3	1.000
Total	**3**	**1.000**

Table A.544: Cluster 164 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	3	1.000
Total	**3**	**1.000**

Table A.545: Cluster 165 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Pig	2	0.667
Cow	1	0.333
Total	**3**	**1.000**

Table A.546: Cluster 166 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Human	2	0.667
Red-shoulder Hawk	1	0.333
Total	**3**	**1.000**

Table A.547: Cluster 167 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Pig	3	1.000
Total	**3**	**1.000**

Table A.548: Cluster 168 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Ground Squirrel	3	1.000
Total	**3**	**1.000**

Table A.549: Cluster 169 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	3	1.000
Total	**3**	**1.000**

Table A.550: Cluster 170 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	3	1.000
Total	**3**	**1.000**

Table A.551: Cluster 171 of 183 for `MinPts`=2

Host species	Count	Proportion
Horse	2	0.667
Cow	1	0.333
Total	**3**	**1.000**

Table A.552: Cluster 172 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	3	1.000
Total	**3**	**1.000**

Table A.553: Cluster 173 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	2	0.667
Human	1	0.333
Total	**3**	**1.000**

Table A.554: Cluster 174 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	3	1.000
Total	**3**	**1.000**

Table A.555: Cluster 175 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Sheep	3	1.000
Total	**3**	**1.000**

Table A.556: Cluster 176 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Human	3	1.000
Total	**3**	**1.000**

Table A.557: Cluster 177 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Dog	2	0.667
Cat	1	0.333
Total	**3**	**1.000**

Table A.558: Cluster 178 of 183 for `MinPts`=2

Host species	**Count**	**Proportion**
Cow	2	0.667
Sheep	1	0.333
Total	**3**	**1.000**

Table A.559: Cluster 179 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	3	1.000
Total	**3**	**1.000**

Table A.560: Cluster 180 of 183 for `MinPts`=2

Host species	Count	Proportion
Cow	3	1.000
Total	**3**	**1.000**

Table A.561: Cluster 181 of 183 for `MinPts`=2

Host species	Count	Proportion
Dog	3	1.000
Total	**3**	**1.000**

Table A.562: Cluster 182 of 183 for `MinPts`=2

Host species	Count	Proportion
Cat	3	1.000
Total	**3**	**1.000**

Table A.563: Cluster 183 of 183 for `MinPts`=2

A.3 Cluster Counts for `MinPts` = 3

Host species	**Count**	**Proportion**
Human	168	0.469
Cow	165	0.461
Pig	7	0.020
Pigeon	5	0.014
Sheep	3	0.008
Ground Squirrel	3	0.008
Chicken	3	0.008
Dog	2	0.006
Horse	1	0.003
Coyote	1	0.003
Total	**358**	**1.000**

Table A.564: Cluster 1 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Human	248	1.000
Total	**248**	**1.000**

Table A.565: Cluster 2 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	159	0.736
Pigeon	12	0.056
Dog	10	0.046
Human	8	0.037
Wild Turkey	7	0.032
Ground Squirrel	5	0.023
Horse	3	0.014
Chicken	3	0.014
Cliff Sparrow	3	0.014
Sheep	2	0.009
Turkey Vulture	1	0.005
Pelican	1	0.005
Deer Mouse	1	0.005
Pig	1	0.005
Total	**216**	**1.000**

Table A.566: Cluster 3 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	199	1.000
Total	**199**	**1.000**

Table A.567: Cluster 4 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	161	0.843
Cow	15	0.079
Mountain Lion	8	0.042
Deer	6	0.031
Horse	1	0.005
Total	**191**	**1.000**

Table A.568: Cluster 5 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	101	0.971
Mountain Lion	2	0.019
Coyote	1	0.010
Total	**104**	**1.000**

Table A.569: Cluster 6 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	99	1.000
Total	**99**	**1.000**

Table A.570: Cluster 7 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	75	0.765
Dog	8	0.082
Ground Squirrel	5	0.051
Grey Fox	4	0.041
Mountain Lion	2	0.020
Cat	1	0.010
Pigeon	1	0.010
Coyote	1	0.010
Orangutan	1	0.010
Total	**98**	**1.000**

Table A.571: Cluster 8 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	53	0.883
Sheep	3	0.050
Cliff Sparrow	2	0.033
Cat	2	0.033
Total	**60**	**1.000**

Table A.572: Cluster 9 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	50	0.943
Cow	3	0.057
Total	**53**	**1.000**

Table A.573: Cluster 10 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	49	0.980
Cow	1	0.020
Total	**50**	**1.000**

Table A.574: Cluster 11 of 122 for `MinPts`=3

Host species	Count	Proportion
Dog	48	1.000
Total	**48**	**1.000**

Table A.575: Cluster 12 of 122 for `MinPts`=3

Host species	Count	Proportion
Bat	36	0.837
Cow	6	0.140
Human	1	0.023
Total	**43**	**1.000**

Table A.576: Cluster 13 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	41	1.000
Total	**41**	**1.000**

Table A.577: Cluster 14 of 122 for `MinPts`=3

Host species	Count	Proportion
Dog	31	0.939
Wild Turkey	2	0.061
Total	**33**	**1.000**

Table A.578: Cluster 15 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	30	0.968
Wild Turkey	1	0.032
Total	**31**	**1.000**

Table A.579: Cluster 16 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	22	0.733
Sheep	4	0.133
Seagull	1	0.033
Mallard Duck	1	0.033
Pig	1	0.033
Western Kingbird	1	0.033
Total	**30**	**1.000**

Table A.580: Cluster 17 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Ground Squirrel	18	0.621
Human	2	0.069
Red-shoulder Hawk	2	0.069
Great Horned Owl	2	0.069
Horse	2	0.069
Cow	1	0.034
Red Shoulder Hawk	1	0.034
Cat	1	0.034
Total	**29**	**1.000**

Table A.581: Cluster 18 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	26	1.000
Total	**26**	**1.000**

Table A.582: Cluster 19 of 122 for `MinPts`=3

Host species	Count	Proportion
Sheep	21	1.000
Total	**21**	**1.000**

Table A.583: Cluster 20 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	21	1.000
Total	**21**	**1.000**

Table A.584: Cluster 21 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	20	1.000
Total	**20**	**1.000**

Table A.585: Cluster 22 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	18	0.947
Pigeon	1	0.053
Total	**19**	**1.000**

Table A.586: Cluster 23 of 122 for `MinPts`=3

Host species	Count	Proportion
Sheep	17	0.895
Chicken	2	0.105
Total	**19**	**1.000**

Table A.587: Cluster 24 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	18	1.000
Total	**18**	**1.000**

Table A.588: Cluster 25 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	18	1.000
Total	**18**	**1.000**

Table A.589: Cluster 26 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	10	0.588
Ground Squirrel	7	0.412
Total	**17**	**1.000**

Table A.590: Cluster 27 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	16	1.000
Total	**16**	**1.000**

Table A.591: Cluster 28 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	14	0.875
Ground Squirrel	2	0.125
Total	**16**	**1.000**

Table A.592: Cluster 29 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	12	0.750
Cliff Sparrow	3	0.188
Pigeon	1	0.062
Total	**16**	**1.000**

Table A.593: Cluster 30 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Human	15	1.000
Total	**15**	**1.000**

Table A.594: Cluster 31 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	14	1.000
Total	**14**	**1.000**

Table A.595: Cluster 32 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	13	1.000
Total	**13**	**1.000**

Table A.596: Cluster 33 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Human	13	1.000
Total	**13**	**1.000**

Table A.597: Cluster 34 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	12	1.000
Total	**12**	**1.000**

Table A.598: Cluster 35 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.599: Cluster 36 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.600: Cluster 37 of 122 for `MinPts`=3

Host species	Count	Proportion
Ground Squirrel	12	1.000
Total	**12**	**1.000**

Table A.601: Cluster 38 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	4	0.364
Wild Turkey	3	0.273
Red-Winged Blackbird	2	0.182
Pigeon	1	0.091
Pig	1	0.091
Total	**11**	**1.000**

Table A.602: Cluster 39 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Human	3	0.273
Wild Turkey	2	0.182
Elephant Seal	2	0.182
California Sea Lion	2	0.182
Sea Lion	2	0.182
Total	**11**	**1.000**

Table A.603: Cluster 40 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Human	11	1.000
Total	**11**	**1.000**

Table A.604: Cluster 41 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Sheep	8	0.800
Pig	1	0.100
Cow	1	0.100
Total	**10**	**1.000**

Table A.605: Cluster 42 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Deer	10	1.000
Total	**10**	**1.000**

Table A.606: Cluster 43 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Wild Turkey	6	0.600
Cow	4	0.400
Total	**10**	**1.000**

Table A.607: Cluster 44 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	9	1.000
Total	**9**	**1.000**

Table A.608: Cluster 45 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Pelican	4	0.444
Red Throated Loon	2	0.222
Common Loon	2	0.222
Common Murre	1	0.111
Total	**9**	**1.000**

Table A.609: Cluster 46 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	9	1.000
Total	**9**	**1.000**

Table A.610: Cluster 47 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Ground Squirrel	9	1.000
Total	**9**	**1.000**

Table A.611: Cluster 48 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	7	0.875
Ground Squirrel	1	0.125
Total	**8**	**1.000**

Table A.612: Cluster 49 of 122 for `MinPts`=3

Host species	Count	Proportion
Pigeon	5	0.625
Cliff Sparrow	3	0.375
Total	**8**	**1.000**

Table A.613: Cluster 50 of 122 for `MinPts`=3

Host species	Count	Proportion
Cat	8	1.000
Total	**8**	**1.000**

Table A.614: Cluster 51 of 122 for `MinPts`=3

Host species	Count	Proportion
Sea Otter	8	1.000
Total	**8**	**1.000**

Table A.615: Cluster 52 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	7	0.875
Cow	1	0.125
Total	**8**	**1.000**

Table A.616: Cluster 53 of 122 for `MinPts`=3

Host species	Count	Proportion
Ground Squirrel	7	1.000
Total	**7**	**1.000**

Table A.617: Cluster 54 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	7	1.000
Total	**7**	**1.000**

Table A.618: Cluster 55 of 122 for `MinPts`=3

Host species	Count	Proportion
Pigeon	5	0.714
Pig	2	0.286
Total	**7**	**1.000**

Table A.619: Cluster 56 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	6	0.857
Wild Turkey	1	0.143
Total	**7**	**1.000**

Table A.620: Cluster 57 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	6	0.857
Chicken	1	0.143
Total	**7**	**1.000**

Table A.621: Cluster 58 of 122 for `MinPts`=3

Host species	Count	Proportion
Ground Squirrel	7	1.000
Total	**7**	**1.000**

Table A.622: Cluster 59 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	7	1.000
Total	**7**	**1.000**

Table A.623: Cluster 60 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	7	1.000
Total	**7**	**1.000**

Table A.624: Cluster 61 of 122 for `MinPts`=3

Host species	Count	Proportion
Bear	5	0.833
Ground Squirrel	1	0.167
Total	**6**	**1.000**

Table A.625: Cluster 62 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	5	0.833
Wild Turkey	1	0.167
Total	**6**	**1.000**

Table A.626: Cluster 63 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	6	1.000
Total	**6**	**1.000**

Table A.627: Cluster 64 of 122 for `MinPts`=3

Host species	Count	Proportion
Ground Squirrel	6	1.000
Total	**6**	**1.000**

Table A.628: Cluster 65 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	4	0.667
Ground Squirrel	1	0.167
Pelican	1	0.167
Total	**6**	**1.000**

Table A.629: Cluster 66 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	6	1.000
Total	**6**	**1.000**

Table A.630: Cluster 67 of 122 for `MinPts`=3

Host species	Count	Proportion
Mountain Lion	6	1.000
Total	**6**	**1.000**

Table A.631: Cluster 68 of 122 for `MinPts`=3

Host species	Count	Proportion
Pigeon	6	1.000
Total	**6**	**1.000**

Table A.632: Cluster 69 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	5	0.833
Chicken	1	0.167
Total	**6**	**1.000**

Table A.633: Cluster 70 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	6	1.000
Total	**6**	**1.000**

Table A.634: Cluster 71 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.635: Cluster 72 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.636: Cluster 73 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.637: Cluster 74 of 122 for `MinPts`=3

Host species	Count	Proportion
Pigeon	5	1.000
Total	**5**	**1.000**

Table A.638: Cluster 75 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Pigeon	5	1.000
Total	**5**	**1.000**

Table A.639: Cluster 76 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Pig	5	1.000
Total	**5**	**1.000**

Table A.640: Cluster 77 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	5	1.000
Total	**5**	**1.000**

Table A.641: Cluster 78 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	5	1.000
Total	**5**	**1.000**

Table A.642: Cluster 79 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Human	5	1.000
Total	**5**	**1.000**

Table A.643: Cluster 80 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	5	1.000
Total	**5**	**1.000**

Table A.644: Cluster 81 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	5	1.000
Total	**5**	**1.000**

Table A.645: Cluster 82 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	5	1.000
Total	**5**	**1.000**

Table A.646: Cluster 83 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	5	1.000
Total	**5**	**1.000**

Table A.647: Cluster 84 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.648: Cluster 85 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Opossum	4	1.000
Total	**4**	**1.000**

Table A.649: Cluster 86 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.650: Cluster 87 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Mountain Lion	4	1.000
Total	**4**	**1.000**

Table A.651: Cluster 88 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Rabbit	4	1.000
Total	**4**	**1.000**

Table A.652: Cluster 89 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.653: Cluster 90 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	3	0.750
Dog	1	0.250
Total	**4**	**1.000**

Table A.654: Cluster 91 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	4	1.000
Total	**4**	**1.000**

Table A.655: Cluster 92 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.656: Cluster 93 of 122 for `MinPts`=3

Host species	Count	Proportion
Eared Grebe	2	0.500
Red Tailed Hawk	1	0.250
Mallard Duck	1	0.250
Total	**4**	**1.000**

Table A.657: Cluster 94 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.658: Cluster 95 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.659: Cluster 96 of 122 for `MinPts`=3

Host species	Count	Proportion
Human	4	1.000
Total	**4**	**1.000**

Table A.660: Cluster 97 of 122 for `MinPts`=3

Host species	Count	Proportion
Pigeon	4	1.000
Total	**4**	**1.000**

Table A.661: Cluster 98 of 122 for `MinPts`=3

Host species	Count	Proportion
Seagull	2	0.500
California Sea Lion	2	0.500
Total	**4**	**1.000**

Table A.662: Cluster 99 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Pigeon	4	1.000
Total	**4**	**1.000**

Table A.663: Cluster 100 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Pig	4	1.000
Total	**4**	**1.000**

Table A.664: Cluster 101 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.665: Cluster 102 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Wild Turkey	4	1.000
Total	**4**	**1.000**

Table A.666: Cluster 103 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Human	4	1.000
Total	**4**	**1.000**

Table A.667: Cluster 104 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
White Crowned Sparrow	4	1.000
Total	**4**	**1.000**

Table A.668: Cluster 105 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.669: Cluster 106 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Pig	4	1.000
Total	**4**	**1.000**

Table A.670: Cluster 107 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Pigeon	4	1.000
Total	**4**	**1.000**

Table A.671: Cluster 108 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.672: Cluster 109 of 122 for `MinPts`=3

Host species	Count	Proportion
Pigeon	2	0.500
Human	1	0.250
Cow	1	0.250
Total	**4**	**1.000**

Table A.673: Cluster 110 of 122 for `MinPts`=3

Host species	Count	Proportion
Cliff Sparrow	2	0.500
Cow	2	0.500
Total	**4**	**1.000**

Table A.674: Cluster 111 of 122 for `MinPts`=3

Host species	Count	Proportion
Pigeon	2	0.500
Tree Swallow	2	0.500
Total	**4**	**1.000**

Table A.675: Cluster 112 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.676: Cluster 113 of 122 for `MinPts`=3

Host species	Count	Proportion
Pigeon	4	1.000
Total	**4**	**1.000**

Table A.677: Cluster 114 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.678: Cluster 115 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.679: Cluster 116 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.680: Cluster 117 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.681: Cluster 118 of 122 for `MinPts`=3

Host species	**Count**	**Proportion**
Cow	4	1.000
Total	**4**	**1.000**

Table A.682: Cluster 119 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.683: Cluster 120 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.684: Cluster 121 of 122 for `MinPts`=3

Host species	Count	Proportion
Cow	4	1.000
Total	**4**	**1.000**

Table A.685: Cluster 122 of 122 for `MinPts`=3

A.4 Cluster Counts for `MinPts` = 4

Host species	Count	Proportion
Human	168	0.473
Cow	162	0.456
Pig	7	0.020
Pigeon	5	0.014
Sheep	3	0.008
Ground Squirrel	3	0.008
Chicken	3	0.008
Dog	2	0.006
Horse	1	0.003
Coyote	1	0.003
Total	**355**	**1.000**

Table A.686: Cluster 1 of 83 for `MinPts`=4

Host species	Count	Proportion
Human	246	1.000
Total	**246**	**1.000**

Table A.687: Cluster 2 of 83 for `MinPts`=4

Host species	Count	Proportion
Human	199	1.000
Total	**199**	**1.000**

Table A.688: Cluster 3 of 83 for `MinPts`=4

Host species	Count	Proportion
Human	161	0.852
Cow	13	0.069
Mountain Lion	8	0.042
Deer	6	0.032
Horse	1	0.005
Total	**189**	**1.000**

Table A.689: Cluster 4 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Human	101	0.971
Mountain Lion	2	0.019
Coyote	1	0.010
Total	**104**	**1.000**

Table A.690: Cluster 5 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Human	99	1.000
Total	**99**	**1.000**

Table A.691: Cluster 6 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Human	75	0.765
Dog	8	0.082
Ground Squirrel	5	0.051
Grey Fox	4	0.041
Mountain Lion	2	0.020
Cat	1	0.010
Pigeon	1	0.010
Coyote	1	0.010
Orangutan	1	0.010
Total	**98**	**1.000**

Table A.692: Cluster 7 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Cow	77	0.786
Pigeon	9	0.092
Human	3	0.031
Horse	2	0.020
Dog	2	0.020
Wild Turkey	2	0.020
Deer Mouse	1	0.010
Pig	1	0.010
Cliff Sparrow	1	0.010
Total	**98**	**1.000**

Table A.693: Cluster 8 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Cow	75	0.882
Wild Turkey	2	0.024
Horse	2	0.024
Ground Squirrel	2	0.024
Sheep	2	0.024
Dog	1	0.012
Human	1	0.012
Total	**85**	**1.000**

Table A.694: Cluster 9 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Cow	53	0.883
Sheep	3	0.050
Cliff Sparrow	2	0.033
Cat	2	0.033
Total	**60**	**1.000**

Table A.695: Cluster 10 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Human	50	0.943
Cow	3	0.057
Total	**53**	**1.000**

Table A.696: Cluster 11 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Human	49	0.980
Cow	1	0.020
Total	**50**	**1.000**

Table A.697: Cluster 12 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Dog	48	1.000
Total	**48**	**1.000**

Table A.698: Cluster 13 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Cow	39	1.000
Total	**39**	**1.000**

Table A.699: Cluster 14 of 83 for `MinPts`=4

Host species	Count	Proportion
Bat	36	0.973
Human	1	0.027
Total	**37**	**1.000**

Table A.700: Cluster 15 of 83 for `MinPts`=4

Host species	Count	Proportion
Dog	31	0.939
Wild Turkey	2	0.061
Total	**33**	**1.000**

Table A.701: Cluster 16 of 83 for `MinPts`=4

Host species	Count	Proportion
Dog	7	0.219
Cow	6	0.188
Human	5	0.156
Pigeon	3	0.094
Wild Turkey	3	0.094
Chicken	3	0.094
Cliff Sparrow	2	0.062
Turkey Vulture	1	0.031
Pelican	1	0.031
Sheep	1	0.031
Total	**32**	**1.000**

Table A.702: Cluster 17 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	30	0.968
Wild Turkey	1	0.032
Total	**31**	**1.000**

Table A.703: Cluster 18 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Ground Squirrel	18	0.621
Human	2	0.069
Red-shoulder Hawk	2	0.069
Great Horned Owl	2	0.069
Horse	2	0.069
Cow	1	0.034
Red Shoulder Hawk	1	0.034
Cat	1	0.034
Total	**29**	**1.000**

Table A.704: Cluster 19 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	26	1.000
Total	**26**	**1.000**

Table A.705: Cluster 20 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	22	0.880
Seagull	1	0.040
Mallard Duck	1	0.040
Western Kingbird	1	0.040
Total	**25**	**1.000**

Table A.706: Cluster 21 of 83 for `MinPts`=4

Host species	Count	Proportion
Sheep	21	1.000
Total	**21**	**1.000**

Table A.707: Cluster 22 of 83 for `MinPts`=4

Host species	Count	Proportion
Human	21	1.000
Total	**21**	**1.000**

Table A.708: Cluster 23 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	18	0.947
Pigeon	1	0.053
Total	**19**	**1.000**

Table A.709: Cluster 24 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	18	1.000
Total	**18**	**1.000**

Table A.710: Cluster 25 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Sheep	16	0.889
Chicken	2	0.111
Total	**18**	**1.000**

Table A.711: Cluster 26 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Human	18	1.000
Total	**18**	**1.000**

Table A.712: Cluster 27 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Cow	10	0.588
Ground Squirrel	7	0.412
Total	**17**	**1.000**

Table A.713: Cluster 28 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Human	17	1.000
Total	**17**	**1.000**

Table A.714: Cluster 29 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	14	0.875
Ground Squirrel	2	0.125
Total	**16**	**1.000**

Table A.715: Cluster 30 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	16	1.000
Total	**16**	**1.000**

Table A.716: Cluster 31 of 83 for `MinPts`=4

Host species	Count	Proportion
Human	15	1.000
Total	**15**	**1.000**

Table A.717: Cluster 32 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	14	1.000
Total	**14**	**1.000**

Table A.718: Cluster 33 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	13	1.000
Total	**13**	**1.000**

Table A.719: Cluster 34 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	10	0.769
Cliff Sparrow	3	0.231
Total	**13**	**1.000**

Table A.720: Cluster 35 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.721: Cluster 36 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.722: Cluster 37 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.723: Cluster 38 of 83 for `MinPts`=4

Host species	Count	Proportion
Ground Squirrel	12	1.000
Total	**12**	**1.000**

Table A.724: Cluster 39 of 83 for `MinPts`=4

Host species	Count	Proportion
Human	12	1.000
Total	**12**	**1.000**

Table A.725: Cluster 40 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	4	0.364
Wild Turkey	3	0.273
Red-Winged Blackbird	2	0.182
Pigeon	1	0.091
Pig	1	0.091
Total	**11**	**1.000**

Table A.726: Cluster 41 of 83 for `MinPts`=4

Host species	Count	Proportion
Human	3	0.273
Wild Turkey	2	0.182
Elephant Seal	2	0.182
California Sea Lion	2	0.182
Sea Lion	2	0.182
Total	**11**	**1.000**

Table A.727: Cluster 42 of 83 for `MinPts`=4

Host species	Count	Proportion
Human	11	1.000
Total	**11**	**1.000**

Table A.728: Cluster 43 of 83 for `MinPts`=4

Host species	Count	Proportion
Sheep	8	0.800
Pig	1	0.100
Cow	1	0.100
Total	**10**	**1.000**

Table A.729: Cluster 44 of 83 for `MinPts`=4

Host species	Count	Proportion
Deer	10	1.000
Total	**10**	**1.000**

Table A.730: Cluster 45 of 83 for `MinPts`=4

Host species	Count	Proportion
Wild Turkey	6	0.600
Cow	4	0.400
Total	**10**	**1.000**

Table A.731: Cluster 46 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	9	1.000
Total	**9**	**1.000**

Table A.732: Cluster 47 of 83 for `MinPts`=4

Host species	Count	Proportion
Pelican	4	0.444
Red Throated Loon	2	0.222
Common Loon	2	0.222
Common Murre	1	0.111
Total	**9**	**1.000**

Table A.733: Cluster 48 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	9	1.000
Total	**9**	**1.000**

Table A.734: Cluster 49 of 83 for `MinPts`=4

Host species	Count	Proportion
Ground Squirrel	9	1.000
Total	**9**	**1.000**

Table A.735: Cluster 50 of 83 for `MinPts`=4

Host species	Count	Proportion
Pigeon	5	0.625
Cliff Sparrow	3	0.375
Total	**8**	**1.000**

Table A.736: Cluster 51 of 83 for `MinPts`=4

Host species	Count	Proportion
Cat	8	1.000
Total	**8**	**1.000**

Table A.737: Cluster 52 of 83 for `MinPts`=4

Host species	Count	Proportion
Sea Otter	8	1.000
Total	**8**	**1.000**

Table A.738: Cluster 53 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	7	1.000
Total	**7**	**1.000**

Table A.739: Cluster 54 of 83 for `MinPts`=4

Host species	Count	Proportion
Pigeon	5	0.714
Pig	2	0.286
Total	**7**	**1.000**

Table A.740: Cluster 55 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	6	0.857
Wild Turkey	1	0.143
Total	**7**	**1.000**

Table A.741: Cluster 56 of 83 for `MinPts`=4

Host species	Count	Proportion
Ground Squirrel	7	1.000
Total	**7**	**1.000**

Table A.742: Cluster 57 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	6	0.857
Chicken	1	0.143
Total	**7**	**1.000**

Table A.743: Cluster 58 of 83 for `MinPts`=4

Host species	Count	Proportion
Human	7	1.000
Total	**7**	**1.000**

Table A.744: Cluster 59 of 83 for `MinPts`=4

Host species	Count	Proportion
Ground Squirrel	7	1.000
Total	**7**	**1.000**

Table A.745: Cluster 60 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	7	1.000
Total	**7**	**1.000**

Table A.746: Cluster 61 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	7	1.000
Total	**7**	**1.000**

Table A.747: Cluster 62 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	6	0.857
Human	1	0.143
Total	**7**	**1.000**

Table A.748: Cluster 63 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	5	0.833
Ground Squirrel	1	0.167
Total	**6**	**1.000**

Table A.749: Cluster 64 of 83 for `MinPts`=4

Host species	Count	Proportion
Bear	5	0.833
Ground Squirrel	1	0.167
Total	**6**	**1.000**

Table A.750: Cluster 65 of 83 for `MinPts`=4

Host species	Count	Proportion
Mountain Lion	6	1.000
Total	**6**	**1.000**

Table A.751: Cluster 66 of 83 for `MinPts`=4

Host species	Count	Proportion
Human	4	0.667
Ground Squirrel	1	0.167
Pelican	1	0.167
Total	**6**	**1.000**

Table A.752: Cluster 67 of 83 for `MinPts`=4

Host species	Count	Proportion
Pigeon	6	1.000
Total	**6**	**1.000**

Table A.753: Cluster 68 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.754: Cluster 69 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Cow	5	1.000
Total	**5**	**1.000**

Table A.755: Cluster 70 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Cow	5	1.000
Total	**5**	**1.000**

Table A.756: Cluster 71 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Pigeon	5	1.000
Total	**5**	**1.000**

Table A.757: Cluster 72 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Cow	5	1.000
Total	**5**	**1.000**

Table A.758: Cluster 73 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Ground Squirrel	5	1.000
Total	**5**	**1.000**

Table A.759: Cluster 74 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Pig	5	1.000
Total	**5**	**1.000**

Table A.760: Cluster 75 of 83 for `MinPts`=4

Host species	Count	Proportion
Human	5	1.000
Total	**5**	**1.000**

Table A.761: Cluster 76 of 83 for `MinPts`=4

Host species	Count	Proportion
Pigeon	5	1.000
Total	**5**	**1.000**

Table A.762: Cluster 77 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.763: Cluster 78 of 83 for `MinPts`=4

Host species	Count	Proportion
Cow	5	1.000
Total	**5**	**1.000**

Table A.764: Cluster 79 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Cow	5	1.000
Total	**5**	**1.000**

Table A.765: Cluster 80 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Cow	5	1.000
Total	**5**	**1.000**

Table A.766: Cluster 81 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Cow	5	1.000
Total	**5**	**1.000**

Table A.767: Cluster 82 of 83 for `MinPts`=4

Host species	**Count**	**Proportion**
Human	5	1.000
Total	**5**	**1.000**

Table A.768: Cluster 83 of 83 for `MinPts`=4

A.5 Cluster Counts for `MinPts` = 5

Host species	Count	Proportion
Human	168	0.476
Cow	160	0.453
Pig	7	0.020
Pigeon	5	0.014
Sheep	3	0.008
Ground Squirrel	3	0.008
Chicken	3	0.008
Dog	2	0.006
Horse	1	0.003
Coyote	1	0.003
Total	**353**	**1.000**

Table A.769: Cluster 1 of 67 for `MinPts`=5

Host species	Count	Proportion
Human	225	1.000
Total	**225**	**1.000**

Table A.770: Cluster 2 of 67 for `MinPts`=5

Host species	Count	Proportion
Human	199	1.000
Total	**199**	**1.000**

Table A.771: Cluster 3 of 67 for `MinPts`=5

Host species	Count	Proportion
Human	161	0.866
Cow	10	0.054
Mountain Lion	8	0.043
Deer	6	0.032
Horse	1	0.005
Total	**186**	**1.000**

Table A.772: Cluster 4 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Human	101	0.971
Mountain Lion	2	0.019
Coyote	1	0.010
Total	**104**	**1.000**

Table A.773: Cluster 5 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Human	99	1.000
Total	**99**	**1.000**

Table A.774: Cluster 6 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Human	75	0.781
Dog	8	0.083
Ground Squirrel	4	0.042
Grey Fox	4	0.042
Mountain Lion	2	0.021
Cat	1	0.010
Coyote	1	0.010
Orangutan	1	0.010
Total	**96**	**1.000**

Table A.775: Cluster 7 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Cow	73	0.777
Pigeon	9	0.096
Human	3	0.032
Horse	2	0.021
Dog	2	0.021
Wild Turkey	2	0.021
Deer Mouse	1	0.011
Pig	1	0.011
Cliff Sparrow	1	0.011
Total	**94**	**1.000**

Table A.776: Cluster 8 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Cow	74	0.881
Ground Squirrel	2	0.024
Wild Turkey	2	0.024
Horse	2	0.024
Sheep	2	0.024
Dog	1	0.012
Human	1	0.012
Total	**84**	**1.000**

Table A.777: Cluster 9 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Cow	53	0.898
Sheep	2	0.034
Cliff Sparrow	2	0.034
Cat	2	0.034
Total	**59**	**1.000**

Table A.778: Cluster 10 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Human	50	0.943
Cow	3	0.057
Total	**53**	**1.000**

Table A.779: Cluster 11 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Human	49	0.980
Cow	1	0.020
Total	**50**	**1.000**

Table A.780: Cluster 12 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Dog	48	1.000
Total	**48**	**1.000**

Table A.781: Cluster 13 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Cow	39	1.000
Total	**39**	**1.000**

Table A.782: Cluster 14 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Bat	36	0.973
Human	1	0.027
Total	**37**	**1.000**

Table A.783: Cluster 15 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Dog	31	0.939
Wild Turkey	2	0.061
Total	**33**	**1.000**

Table A.784: Cluster 16 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Cow	30	0.968
Wild Turkey	1	0.032
Total	**31**	**1.000**

Table A.785: Cluster 17 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Dog	7	0.233
Human	5	0.167
Cow	5	0.167
Wild Turkey	3	0.100
Chicken	3	0.100
Cliff Sparrow	2	0.067
Pigeon	2	0.067
Turkey Vulture	1	0.033
Pelican	1	0.033
Sheep	1	0.033
Total	**30**	**1.000**

Table A.786: Cluster 18 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Ground Squirrel	18	0.621
Human	2	0.069
Red-shoulder Hawk	2	0.069
Great Horned Owl	2	0.069
Horse	2	0.069
Cow	1	0.034
Red Shoulder Hawk	1	0.034
Cat	1	0.034
Total	**29**	**1.000**

Table A.787: Cluster 19 of 67 for `MinPts`=5

Host species	Count	Proportion
Human	23	1.000
Total	**23**	**1.000**

Table A.788: Cluster 20 of 67 for `MinPts`=5

Host species	Count	Proportion
Cow	22	1.000
Total	**22**	**1.000**

Table A.789: Cluster 21 of 67 for `MinPts`=5

Host species	Count	Proportion
Cow	20	0.909
Mallard Duck	1	0.045
Western Kingbird	1	0.045
Total	**22**	**1.000**

Table A.790: Cluster 22 of 67 for `MinPts`=5

Host species	Count	Proportion
Sheep	21	1.000
Total	**21**	**1.000**

Table A.791: Cluster 23 of 67 for `MinPts`=5

Host species	Count	Proportion
Human	21	1.000
Total	**21**	**1.000**

Table A.792: Cluster 24 of 67 for `MinPts`=5

Host species	Count	Proportion
Cow	18	0.947
Pigeon	1	0.053
Total	**19**	**1.000**

Table A.793: Cluster 25 of 67 for `MinPts`=5

Host species	Count	Proportion
Sheep	16	0.889
Chicken	2	0.111
Total	**18**	**1.000**

Table A.794: Cluster 26 of 67 for `MinPts`=5

Host species	Count	Proportion
Cow	17	1.000
Total	**17**	**1.000**

Table A.795: Cluster 27 of 67 for `MinPts`=5

Host species	Count	Proportion
Cow	10	0.588
Ground Squirrel	7	0.412
Total	**17**	**1.000**

Table A.796: Cluster 28 of 67 for `MinPts`=5

Host species	Count	Proportion
Human	17	1.000
Total	**17**	**1.000**

Table A.797: Cluster 29 of 67 for `MinPts`=5

Host species	Count	Proportion
Cow	16	1.000
Total	**16**	**1.000**

Table A.798: Cluster 30 of 67 for `MinPts`=5

Host species	Count	Proportion
Human	15	1.000
Total	**15**	**1.000**

Table A.799: Cluster 31 of 67 for `MinPts`=5

Host species	Count	Proportion
Cow	14	1.000
Total	**14**	**1.000**

Table A.800: Cluster 32 of 67 for `MinPts`=5

Host species	Count	Proportion
Cow	13	1.000
Total	**13**	**1.000**

Table A.801: Cluster 33 of 67 for `MinPts`=5

Host species	Count	Proportion
Cow	13	1.000
Total	**13**	**1.000**

Table A.802: Cluster 34 of 67 for `MinPts`=5

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.803: Cluster 35 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Cow	12	1.000
Total	**12**	**1.000**

Table A.804: Cluster 36 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Cow	12	1.000
Total	**12**	**1.000**

Table A.805: Cluster 37 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Ground Squirrel	12	1.000
Total	**12**	**1.000**

Table A.806: Cluster 38 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Human	12	1.000
Total	**12**	**1.000**

Table A.807: Cluster 39 of 67 for `MinPts`=5

Host species	Count	Proportion
Human	3	0.273
Wild Turkey	2	0.182
Elephant Seal	2	0.182
California Sea Lion	2	0.182
Sea Lion	2	0.182
Total	**11**	**1.000**

Table A.808: Cluster 40 of 67 for `MinPts`=5

Host species	Count	Proportion
Human	11	1.000
Total	**11**	**1.000**

Table A.809: Cluster 41 of 67 for `MinPts`=5

Host species	Count	Proportion
Cow	4	0.364
Wild Turkey	3	0.273
Red-Winged Blackbird	2	0.182
Pigeon	1	0.091
Pig	1	0.091
Total	**11**	**1.000**

Table A.810: Cluster 42 of 67 for `MinPts`=5

Host species	Count	Proportion
Cow	8	0.727
Cliff Sparrow	3	0.273
Total	**11**	**1.000**

Table A.811: Cluster 43 of 67 for `MinPts`=5

Host species	Count	Proportion
Human	11	1.000
Total	**11**	**1.000**

Table A.812: Cluster 44 of 67 for `MinPts`=5

Host species	Count	Proportion
Sheep	8	0.800
Pig	1	0.100
Cow	1	0.100
Total	**10**	**1.000**

Table A.813: Cluster 45 of 67 for `MinPts`=5

Host species	Count	Proportion
Deer	10	1.000
Total	**10**	**1.000**

Table A.814: Cluster 46 of 67 for `MinPts`=5

Host species	Count	Proportion
Cow	9	1.000
Total	**9**	**1.000**

Table A.815: Cluster 47 of 67 for `MinPts`=5

Host species	Count	Proportion
Pelican	4	0.444
Red Throated Loon	2	0.222
Common Loon	2	0.222
Common Murre	1	0.111
Total	**9**	**1.000**

Table A.816: Cluster 48 of 67 for `MinPts`=5

Host species	Count	Proportion
Cow	9	1.000
Total	**9**	**1.000**

Table A.817: Cluster 49 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Ground Squirrel	9	1.000
Total	**9**	**1.000**

Table A.818: Cluster 50 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Pigeon	5	0.625
Cliff Sparrow	3	0.375
Total	**8**	**1.000**

Table A.819: Cluster 51 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Cat	8	1.000
Total	**8**	**1.000**

Table A.820: Cluster 52 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Sea Otter	8	1.000
Total	**8**	**1.000**

Table A.821: Cluster 53 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Cow	7	1.000
Total	**7**	**1.000**

Table A.822: Cluster 54 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Pigeon	5	0.714
Pig	2	0.286
Total	**7**	**1.000**

Table A.823: Cluster 55 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Cow	6	0.857
Wild Turkey	1	0.143
Total	**7**	**1.000**

Table A.824: Cluster 56 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Ground Squirrel	7	1.000
Total	**7**	**1.000**

Table A.825: Cluster 57 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Cow	7	1.000
Total	**7**	**1.000**

Table A.826: Cluster 58 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Cow	7	1.000
Total	**7**	**1.000**

Table A.827: Cluster 59 of 67 for `MinPts`=5

Host species	Count	Proportion
Human	7	1.000
Total	**7**	**1.000**

Table A.828: Cluster 60 of 67 for `MinPts`=5

Host species	Count	Proportion
Wild Turkey	6	0.857
Cow	1	0.143
Total	**7**	**1.000**

Table A.829: Cluster 61 of 67 for `MinPts`=5

Host species	Count	Proportion
Mountain Lion	6	1.000
Total	**6**	**1.000**

Table A.830: Cluster 62 of 67 for `MinPts`=5

Host species	Count	Proportion
Bear	5	0.833
Ground Squirrel	1	0.167
Total	**6**	**1.000**

Table A.831: Cluster 63 of 67 for `MinPts`=5

Host species	Count	Proportion
Human	4	0.667
Ground Squirrel	1	0.167
Pelican	1	0.167
Total	**6**	**1.000**

Table A.832: Cluster 64 of 67 for `MinPts`=5

Host species	Count	Proportion
Cow	6	1.000
Total	**6**	**1.000**

Table A.833: Cluster 65 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Ground Squirrel	6	1.000
Total	**6**	**1.000**

Table A.834: Cluster 66 of 67 for `MinPts`=5

Host species	**Count**	**Proportion**
Cow	5	0.833
Ground Squirrel	1	0.167
Total	**6**	**1.000**

Table A.835: Cluster 67 of 67 for `MinPts`=5

A.6 Cluster Counts for MinPts = 6

Host species	Count	Proportion
Human	168	0.476
Cow	160	0.453
Pig	7	0.020
Pigeon	5	0.014
Sheep	3	0.008
Ground Squirrel	3	0.008
Chicken	3	0.008
Dog	2	0.006
Horse	1	0.003
Coyote	1	0.003
Total	**353**	**1.000**

Table A.836: Cluster 1 of 61 for MinPts=6

Host species	Count	Proportion
Human	225	1.000
Total	**225**	**1.000**

Table A.837: Cluster 2 of 61 for MinPts=6

Host species	Count	Proportion
Human	199	1.000
Total	**199**	**1.000**

Table A.838: Cluster 3 of 61 for MinPts=6

Host species	Count	Proportion
Human	161	0.866
Cow	10	0.054
Mountain Lion	8	0.043
Deer	6	0.032
Horse	1	0.005
Total	**186**	**1.000**

Table A.839: Cluster 4 of 61 for MinPts=6

Host species	**Count**	**Proportion**
Human	101	0.971
Mountain Lion	2	0.019
Coyote	1	0.010
Total	**104**	**1.000**

Table A.840: Cluster 5 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Human	98	1.000
Total	**98**	**1.000**

Table A.841: Cluster 6 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Human	75	0.815
Dog	8	0.087
Ground Squirrel	4	0.043
Mountain Lion	2	0.022
Cat	1	0.011
Grey Fox	1	0.011
Orangutan	1	0.011
Total	**92**	**1.000**

Table A.842: Cluster 7 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Cow	70	0.769
Pigeon	9	0.099
Human	3	0.033
Horse	2	0.022
Dog	2	0.022
Wild Turkey	2	0.022
Deer Mouse	1	0.011
Pig	1	0.011
Cliff Sparrow	1	0.011
Total	**91**	**1.000**

Table A.843: Cluster 8 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Cow	73	0.924
Ground Squirrel	2	0.025
Wild Turkey	2	0.025
Dog	1	0.013
Horse	1	0.013
Total	**79**	**1.000**

Table A.844: Cluster 9 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Cow	52	0.897
Sheep	2	0.034
Cliff Sparrow	2	0.034
Cat	2	0.034
Total	**58**	**1.000**

Table A.845: Cluster 10 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Human	50	0.943
Cow	3	0.057
Total	**53**	**1.000**

Table A.846: Cluster 11 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Human	49	0.980
Cow	1	0.020
Total	**50**	**1.000**

Table A.847: Cluster 12 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Dog	48	1.000
Total	**48**	**1.000**

Table A.848: Cluster 13 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Cow	37	1.000
Total	**37**	**1.000**

Table A.849: Cluster 14 of 61 for `MinPts`=6

Host species	Count	Proportion
Bat	36	0.973
Human	1	0.027
Total	**37**	**1.000**

Table A.850: Cluster 15 of 61 for `MinPts`=6

Host species	Count	Proportion
Dog	30	0.938
Wild Turkey	2	0.062
Total	**32**	**1.000**

Table A.851: Cluster 16 of 61 for `MinPts`=6

Host species	Count	Proportion
Cow	29	0.967
Wild Turkey	1	0.033
Total	**30**	**1.000**

Table A.852: Cluster 17 of 61 for `MinPts`=6

Host species	Count	Proportion
Ground Squirrel	17	0.607
Human	2	0.071
Red-shoulder Hawk	2	0.071
Great Horned Owl	2	0.071
Horse	2	0.071
Cow	1	0.036
Red Shoulder Hawk	1	0.036
Cat	1	0.036
Total	**28**	**1.000**

Table A.853: Cluster 18 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Dog	7	0.259
Cow	5	0.185
Human	4	0.148
Chicken	3	0.111
Cliff Sparrow	2	0.074
Pigeon	2	0.074
Turkey Vulture	1	0.037
Pelican	1	0.037
Wild Turkey	1	0.037
Sheep	1	0.037
Total	**27**	**1.000**

Table A.854: Cluster 19 of 61 for `MinPts`=6

Host species	Count	Proportion
Cow	20	0.909
Mallard Duck	1	0.045
Western Kingbird	1	0.045
Total	**22**	**1.000**

Table A.855: Cluster 20 of 61 for `MinPts`=6

Host species	Count	Proportion
Sheep	21	1.000
Total	**21**	**1.000**

Table A.856: Cluster 21 of 61 for `MinPts`=6

Host species	Count	Proportion
Human	21	1.000
Total	**21**	**1.000**

Table A.857: Cluster 22 of 61 for `MinPts`=6

Host species	Count	Proportion
Human	21	1.000
Total	**21**	**1.000**

Table A.858: Cluster 23 of 61 for `MinPts`=6

Host species	Count	Proportion
Cow	18	0.947
Pigeon	1	0.053
Total	**19**	**1.000**

Table A.859: Cluster 24 of 61 for `MinPts`=6

Host species	Count	Proportion
Human	17	1.000
Total	**17**	**1.000**

Table A.860: Cluster 25 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Cow	10	0.588
Ground Squirrel	7	0.412
Total	**17**	**1.000**

Table A.861: Cluster 26 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Cow	16	1.000
Total	**16**	**1.000**

Table A.862: Cluster 27 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Cow	16	1.000
Total	**16**	**1.000**

Table A.863: Cluster 28 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Human	15	1.000
Total	**15**	**1.000**

Table A.864: Cluster 29 of 61 for `MinPts`=6

Host species	Count	Proportion
Cow	14	1.000
Total	**14**	**1.000**

Table A.865: Cluster 30 of 61 for `MinPts`=6

Host species	Count	Proportion
Cow	13	1.000
Total	**13**	**1.000**

Table A.866: Cluster 31 of 61 for `MinPts`=6

Host species	Count	Proportion
Cow	13	1.000
Total	**13**	**1.000**

Table A.867: Cluster 32 of 61 for `MinPts`=6

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.868: Cluster 33 of 61 for `MinPts`=6

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.869: Cluster 34 of 61 for `MinPts`=6

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.870: Cluster 35 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Ground Squirrel	12	1.000
Total	**12**	**1.000**

Table A.871: Cluster 36 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Human	12	1.000
Total	**12**	**1.000**

Table A.872: Cluster 37 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Cow	12	1.000
Total	**12**	**1.000**

Table A.873: Cluster 38 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Human	3	0.273
Wild Turkey	2	0.182
Elephant Seal	2	0.182
California Sea Lion	2	0.182
Sea Lion	2	0.182
Total	**11**	**1.000**

Table A.874: Cluster 39 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Sheep	11	1.000
Total	**11**	**1.000**

Table A.875: Cluster 40 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Cow	4	0.364
Wild Turkey	3	0.273
Red-Winged Blackbird	2	0.182
Pigeon	1	0.091
Pig	1	0.091
Total	**11**	**1.000**

Table A.876: Cluster 41 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Human	11	1.000
Total	**11**	**1.000**

Table A.877: Cluster 42 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Deer	10	1.000
Total	**10**	**1.000**

Table A.878: Cluster 43 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Cow	9	1.000
Total	**9**	**1.000**

Table A.879: Cluster 44 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Sheep	7	0.778
Pig	1	0.111
Cow	1	0.111
Total	**9**	**1.000**

Table A.880: Cluster 45 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Cow	9	1.000
Total	**9**	**1.000**

Table A.881: Cluster 46 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Human	9	1.000
Total	**9**	**1.000**

Table A.882: Cluster 47 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Ground Squirrel	9	1.000
Total	**9**	**1.000**

Table A.883: Cluster 48 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Pelican	4	0.444
Red Throated Loon	2	0.222
Common Loon	2	0.222
Common Murre	1	0.111
Total	**9**	**1.000**

Table A.884: Cluster 49 of 61 for `MinPts`=6

Host species	Count	Proportion
Sheep	7	0.778
Chicken	2	0.222
Total	**9**	**1.000**

Table A.885: Cluster 50 of 61 for `MinPts`=6

Host species	Count	Proportion
Cow	7	0.778
Cliff Sparrow	2	0.222
Total	**9**	**1.000**

Table A.886: Cluster 51 of 61 for `MinPts`=6

Host species	Count	Proportion
Pigeon	5	0.625
Cliff Sparrow	3	0.375
Total	**8**	**1.000**

Table A.887: Cluster 52 of 61 for `MinPts`=6

Host species	Count	Proportion
Cat	8	1.000
Total	**8**	**1.000**

Table A.888: Cluster 53 of 61 for `MinPts`=6

Host species	Count	Proportion
Sea Otter	8	1.000
Total	**8**	**1.000**

Table A.889: Cluster 54 of 61 for `MinPts`=6

Host species	Count	Proportion
Cow	8	1.000
Total	**8**	**1.000**

Table A.890: Cluster 55 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Cow	7	1.000
Total	**7**	**1.000**

Table A.891: Cluster 56 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Ground Squirrel	7	1.000
Total	**7**	**1.000**

Table A.892: Cluster 57 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Cow	7	1.000
Total	**7**	**1.000**

Table A.893: Cluster 58 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Cow	7	1.000
Total	**7**	**1.000**

Table A.894: Cluster 59 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Wild Turkey	6	0.857
Cow	1	0.143
Total	**7**	**1.000**

Table A.895: Cluster 60 of 61 for `MinPts`=6

Host species	**Count**	**Proportion**
Cow	6	0.857
Wild Turkey	1	0.143
Total	**7**	**1.000**

Table A.896: Cluster 61 of 61 for `MinPts`=6

A.7 Cluster Counts for MinPts = 7

Host species	Count	Proportion
Human	168	0.477
Cow	159	0.452
Pig	7	0.020
Pigeon	5	0.014
Sheep	3	0.009
Ground Squirrel	3	0.009
Chicken	3	0.009
Dog	2	0.006
Horse	1	0.003
Coyote	1	0.003
Total	**352**	**1.000**

Table A.897: Cluster 1 of 56 for MinPts=7

Host species	Count	Proportion
Human	225	1.000
Total	**225**	**1.000**

Table A.898: Cluster 2 of 56 for MinPts=7

Host species	Count	Proportion
Human	161	0.870
Cow	9	0.049
Mountain Lion	8	0.043
Deer	6	0.032
Horse	1	0.005
Total	**185**	**1.000**

Table A.899: Cluster 3 of 56 for MinPts=7

Host species	Count	Proportion
Human	168	1.000
Total	**168**	**1.000**

Table A.900: Cluster 4 of 56 for MinPts=7

Host species	Count	Proportion
Human	101	0.981
Mountain Lion	2	0.019
Total	**103**	**1.000**

Table A.901: Cluster 5 of 56 for `MinPts`=7

Host species	Count	Proportion
Human	98	1.000
Total	**98**	**1.000**

Table A.902: Cluster 6 of 56 for `MinPts`=7

Host species	Count	Proportion
Human	74	0.813
Dog	8	0.088
Ground Squirrel	4	0.044
Mountain Lion	2	0.022
Cat	1	0.011
Grey Fox	1	0.011
Orangutan	1	0.011
Total	**91**	**1.000**

Table A.903: Cluster 7 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	64	0.762
Pigeon	8	0.095
Human	3	0.036
Wild Turkey	2	0.024
Horse	2	0.024
Dog	2	0.024
Pig	1	0.012
Deer Mouse	1	0.012
Cliff Sparrow	1	0.012
Total	**84**	**1.000**

Table A.904: Cluster 8 of 56 for `MinPts`=7

Host species	**Count**	**Proportion**
Cow	72	0.923
Ground Squirrel	2	0.026
Wild Turkey	2	0.026
Dog	1	0.013
Horse	1	0.013
Total	**78**	**1.000**

Table A.905: Cluster 9 of 56 for `MinPts`=7

Host species	Count	Proportion
Human	50	0.962
Cow	2	0.038
Total	**52**	**1.000**

Table A.906: Cluster 10 of 56 for `MinPts`=7

Host species	Count	Proportion
Human	49	0.980
Cow	1	0.020
Total	**50**	**1.000**

Table A.907: Cluster 11 of 56 for `MinPts`=7

Host species	Count	Proportion
Dog	48	1.000
Total	**48**	**1.000**

Table A.908: Cluster 12 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	37	1.000
Total	**37**	**1.000**

Table A.909: Cluster 13 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	37	1.000
Total	**37**	**1.000**

Table A.910: Cluster 14 of 56 for `MinPts`=7

Host species	Count	Proportion
Bat	36	0.973
Human	1	0.027
Total	**37**	**1.000**

Table A.911: Cluster 15 of 56 for `MinPts`=7

Host species	Count	Proportion
Human	33	1.000
Total	**33**	**1.000**

Table A.912: Cluster 16 of 56 for `MinPts`=7

Host species	Count	Proportion
Dog	30	0.938
Wild Turkey	2	0.062
Total	**32**	**1.000**

Table A.913: Cluster 17 of 56 for `MinPts`=7

Host species	Count	Proportion
Ground Squirrel	17	0.607
Human	2	0.071
Red-shoulder Hawk	2	0.071
Great Horned Owl	2	0.071
Horse	2	0.071
Cow	1	0.036
Red Shoulder Hawk	1	0.036
Cat	1	0.036
Total	**28**	**1.000**

Table A.914: Cluster 18 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	27	0.964
Wild Turkey	1	0.036
Total	**28**	**1.000**

Table A.915: Cluster 19 of 56 for `MinPts`=7

Host species	**Count**	**Proportion**
Cow	20	0.909
Mallard Duck	1	0.045
Western Kingbird	1	0.045
Total	**22**	**1.000**

Table A.916: Cluster 20 of 56 for `MinPts`=7

Host species	**Count**	**Proportion**
Sheep	21	1.000
Total	**21**	**1.000**

Table A.917: Cluster 21 of 56 for `MinPts`=7

Host species	**Count**	**Proportion**
Human	21	1.000
Total	**21**	**1.000**

Table A.918: Cluster 22 of 56 for `MinPts`=7

Host species	**Count**	**Proportion**
Cow	14	0.700
Sheep	2	0.100
Cliff Sparrow	2	0.100
Cat	2	0.100
Total	**20**	**1.000**

Table A.919: Cluster 23 of 56 for `MinPts`=7

Host species	**Count**	**Proportion**
Cow	18	0.947
Pigeon	1	0.053
Total	**19**	**1.000**

Table A.920: Cluster 24 of 56 for `MinPts`=7

Host species	Count	Proportion
Human	18	1.000
Total	**18**	**1.000**

Table A.921: Cluster 25 of 56 for `MinPts`=7

Host species	Count	Proportion
Human	16	1.000
Total	**16**	**1.000**

Table A.922: Cluster 26 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	16	1.000
Total	**16**	**1.000**

Table A.923: Cluster 27 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	16	1.000
Total	**16**	**1.000**

Table A.924: Cluster 28 of 56 for `MinPts`=7

Host species	Count	Proportion
Human	15	1.000
Total	**15**	**1.000**

Table A.925: Cluster 29 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	10	0.667
Ground Squirrel	5	0.333
Total	**15**	**1.000**

Table A.926: Cluster 30 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	14	1.000
Total	**14**	**1.000**

Table A.927: Cluster 31 of 56 for `MinPts`=7

Host species	Count	Proportion
Dog	7	0.500
Chicken	3	0.214
Sheep	1	0.071
Cow	1	0.071
Cliff Sparrow	1	0.071
Pigeon	1	0.071
Total	**14**	**1.000**

Table A.928: Cluster 32 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	13	1.000
Total	**13**	**1.000**

Table A.929: Cluster 33 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	13	1.000
Total	**13**	**1.000**

Table A.930: Cluster 34 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.931: Cluster 35 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.932: Cluster 36 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.933: Cluster 37 of 56 for `MinPts`=7

Host species	Count	Proportion
Ground Squirrel	12	1.000
Total	**12**	**1.000**

Table A.934: Cluster 38 of 56 for `MinPts`=7

Host species	Count	Proportion
Human	12	1.000
Total	**12**	**1.000**

Table A.935: Cluster 39 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.936: Cluster 40 of 56 for `MinPts`=7

Host species	Count	Proportion
Human	3	0.273
Wild Turkey	2	0.182
Elephant Seal	2	0.182
California Sea Lion	2	0.182
Sea Lion	2	0.182
Total	**11**	**1.000**

Table A.937: Cluster 41 of 56 for `MinPts`=7

Host species	Count	Proportion
Sheep	11	1.000
Total	**11**	**1.000**

Table A.938: Cluster 42 of 56 for `MinPts`=7

Host species	Count	Proportion
Deer	10	1.000
Total	**10**	**1.000**

Table A.939: Cluster 43 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	4	0.400
Red-Winged Blackbird	2	0.200
Wild Turkey	2	0.200
Pigeon	1	0.100
Pig	1	0.100
Total	**10**	**1.000**

Table A.940: Cluster 44 of 56 for `MinPts`=7

Host species	**Count**	**Proportion**
Cow	9	1.000
Total	**9**	**1.000**

Table A.941: Cluster 45 of 56 for `MinPts`=7

Host species	**Count**	**Proportion**
Human	9	1.000
Total	**9**	**1.000**

Table A.942: Cluster 46 of 56 for `MinPts`=7

Host species	**Count**	**Proportion**
Ground Squirrel	9	1.000
Total	**9**	**1.000**

Table A.943: Cluster 47 of 56 for `MinPts`=7

Host species	**Count**	**Proportion**
Pelican	4	0.444
Red Throated Loon	2	0.222
Common Loon	2	0.222
Common Murre	1	0.111
Total	**9**	**1.000**

Table A.944: Cluster 48 of 56 for `MinPts`=7

Host species	**Count**	**Proportion**
Cow	9	1.000
Total	**9**	**1.000**

Table A.945: Cluster 49 of 56 for `MinPts`=7

Host species	Count	Proportion
Cow	7	0.778
Cliff Sparrow	2	0.222
Total	**9**	**1.000**

Table A.946: Cluster 50 of 56 for `MinPts`=7

Host species	Count	Proportion
Human	4	0.444
Cow	3	0.333
Wild Turkey	1	0.111
Pelican	1	0.111
Total	**9**	**1.000**

Table A.947: Cluster 51 of 56 for `MinPts`=7

Host species	Count	Proportion
Sheep	7	0.875
Cow	1	0.125
Total	**8**	**1.000**

Table A.948: Cluster 52 of 56 for `MinPts`=7

Host species	Count	Proportion
Pigeon	5	0.625
Cliff Sparrow	3	0.375
Total	**8**	**1.000**

Table A.949: Cluster 53 of 56 for `MinPts`=7

Host species	Count	Proportion
Cat	8	1.000
Total	**8**	**1.000**

Table A.950: Cluster 54 of 56 for `MinPts`=7

Host species	**Count**	**Proportion**
Sea Otter	8	1.000
Total	**8**	**1.000**

Table A.951: Cluster 55 of 56 for `MinPts`=7

Host species	**Count**	**Proportion**
Cow	8	1.000
Total	**8**	**1.000**

Table A.952: Cluster 56 of 56 for `MinPts`=7

A.8 Cluster Counts for MinPts = 8

Host species	**Count**	**Proportion**
Human	168	0.477
Cow	159	0.452
Pig	7	0.020
Pigeon	5	0.014
Sheep	3	0.009
Ground Squirrel	3	0.009
Chicken	3	0.009
Dog	2	0.006
Horse	1	0.003
Coyote	1	0.003
Total	**352**	**1.000**

Table A.953: Cluster 1 of 50 for MinPts=8

Host species	**Count**	**Proportion**
Human	225	1.000
Total	**225**	**1.000**

Table A.954: Cluster 2 of 50 for MinPts=8

Host species	**Count**	**Proportion**
Human	161	0.885
Cow	9	0.049
Deer	6	0.033
Mountain Lion	5	0.027
Horse	1	0.005
Total	**182**	**1.000**

Table A.955: Cluster 3 of 50 for MinPts=8

Host species	**Count**	**Proportion**
Human	168	1.000
Total	**168**	**1.000**

Table A.956: Cluster 4 of 50 for MinPts=8

Host species	**Count**	**Proportion**
Human	101	0.981
Mountain Lion	2	0.019
Total	**103**	**1.000**

Table A.957: Cluster 5 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Human	98	1.000
Total	**98**	**1.000**

Table A.958: Cluster 6 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Human	67	0.798
Dog	8	0.095
Ground Squirrel	4	0.048
Mountain Lion	2	0.024
Grey Fox	1	0.012
Cat	1	0.012
Orangutan	1	0.012
Total	**84**	**1.000**

Table A.959: Cluster 7 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	72	0.923
Ground Squirrel	2	0.026
Wild Turkey	2	0.026
Dog	1	0.013
Horse	1	0.013
Total	**78**	**1.000**

Table A.960: Cluster 8 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	57	0.781
Pigeon	5	0.068
Wild Turkey	2	0.027
Human	2	0.027
Horse	2	0.027
Dog	2	0.027
Pig	1	0.014
Deer Mouse	1	0.014
Cliff Sparrow	1	0.014
Total	**73**	**1.000**

Table A.961: Cluster 9 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Human	50	0.962
Cow	2	0.038
Total	**52**	**1.000**

Table A.962: Cluster 10 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Human	49	0.980
Cow	1	0.020
Total	**50**	**1.000**

Table A.963: Cluster 11 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Dog	48	1.000
Total	**48**	**1.000**

Table A.964: Cluster 12 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	37	1.000
Total	**37**	**1.000**

Table A.965: Cluster 13 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	37	1.000
Total	**37**	**1.000**

Table A.966: Cluster 14 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Bat	36	0.973
Human	1	0.027
Total	**37**	**1.000**

Table A.967: Cluster 15 of 50 for `MinPts`=8

Host species	Count	Proportion
Dog	30	0.968
Wild Turkey	1	0.032
Total	**31**	**1.000**

Table A.968: Cluster 16 of 50 for `MinPts`=8

Host species	Count	Proportion
Human	29	1.000
Total	**29**	**1.000**

Table A.969: Cluster 17 of 50 for `MinPts`=8

Host species	Count	Proportion
Ground Squirrel	17	0.607
Human	2	0.071
Red-shoulder Hawk	2	0.071
Great Horned Owl	2	0.071
Horse	2	0.071
Cow	1	0.036
Red Shoulder Hawk	1	0.036
Cat	1	0.036
Total	**28**	**1.000**

Table A.970: Cluster 18 of 50 for `MinPts`=8

Host species	Count	Proportion
Cow	26	0.963
Wild Turkey	1	0.037
Total	**27**	**1.000**

Table A.971: Cluster 19 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	20	0.909
Mallard Duck	1	0.045
Western Kingbird	1	0.045
Total	**22**	**1.000**

Table A.972: Cluster 20 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Sheep	21	1.000
Total	**21**	**1.000**

Table A.973: Cluster 21 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Human	21	1.000
Total	**21**	**1.000**

Table A.974: Cluster 22 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	18	0.947
Pigeon	1	0.053
Total	**19**	**1.000**

Table A.975: Cluster 23 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	12	0.667
Sheep	2	0.111
Cliff Sparrow	2	0.111
Cat	2	0.111
Total	**18**	**1.000**

Table A.976: Cluster 24 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Human	17	1.000
Total	**17**	**1.000**

Table A.977: Cluster 25 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Human	16	1.000
Total	**16**	**1.000**

Table A.978: Cluster 26 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	16	1.000
Total	**16**	**1.000**

Table A.979: Cluster 27 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Human	15	1.000
Total	**15**	**1.000**

Table A.980: Cluster 28 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	15	1.000
Total	**15**	**1.000**

Table A.981: Cluster 29 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	14	1.000
Total	**14**	**1.000**

Table A.982: Cluster 30 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Dog	7	0.500
Chicken	3	0.214
Sheep	1	0.071
Cow	1	0.071
Cliff Sparrow	1	0.071
Pigeon	1	0.071
Total	**14**	**1.000**

Table A.983: Cluster 31 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	13	1.000
Total	**13**	**1.000**

Table A.984: Cluster 32 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	13	1.000
Total	**13**	**1.000**

Table A.985: Cluster 33 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	12	1.000
Total	**12**	**1.000**

Table A.986: Cluster 34 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Ground Squirrel	12	1.000
Total	**12**	**1.000**

Table A.987: Cluster 35 of 50 for `MinPts`=8

Host species	Count	Proportion
Cow	9	0.750
Ground Squirrel	3	0.250
Total	**12**	**1.000**

Table A.988: Cluster 36 of 50 for `MinPts`=8

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.989: Cluster 37 of 50 for `MinPts`=8

Host species	Count	Proportion
Cow	12	1.000
Total	**12**	**1.000**

Table A.990: Cluster 38 of 50 for `MinPts`=8

Host species	Count	Proportion
Cow	10	1.000
Total	**10**	**1.000**

Table A.991: Cluster 39 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Deer	10	1.000
Total	**10**	**1.000**

Table A.992: Cluster 40 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	4	0.400
Red-Winged Blackbird	2	0.200
Wild Turkey	2	0.200
Pigeon	1	0.100
Pig	1	0.100
Total	**10**	**1.000**

Table A.993: Cluster 41 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Human	10	1.000
Total	**10**	**1.000**

Table A.994: Cluster 42 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	9	1.000
Total	**9**	**1.000**

Table A.995: Cluster 43 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Ground Squirrel	9	1.000
Total	**9**	**1.000**

Table A.996: Cluster 44 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Pelican	4	0.444
Red Throated Loon	2	0.222
Common Loon	2	0.222
Common Murre	1	0.111
Total	**9**	**1.000**

Table A.997: Cluster 45 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	7	0.778
Cliff Sparrow	2	0.222
Total	**9**	**1.000**

Table A.998: Cluster 46 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Cow	9	1.000
Total	**9**	**1.000**

Table A.999: Cluster 47 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Human	4	0.444
Cow	3	0.333
Wild Turkey	1	0.111
Pelican	1	0.111
Total	**9**	**1.000**

Table A.1000: Cluster 48 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Human	9	1.000
Total	**9**	**1.000**

Table A.1001: Cluster 49 of 50 for `MinPts`=8

Host species	**Count**	**Proportion**
Sheep	9	1.000
Total	**9**	**1.000**

Table A.1002: Cluster 50 of 50 for `MinPts`=8