COMPUTATIONAL METHODS FOR ESTIMATING GENETIC RELATIONSHIPS

by

Yunyun Zhou

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

WASHINGTON STATE UNIVERSITY
School of Electrical Engineering and Computer Science

August 2012

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of Yunyun Zhou find it satisfactory and recommend that it be accepted.

Shira L. Broschat, Chair

Douglas R. Call

John H. Miller

ACKNOWLEDGMENTS

My first and most earnest acknowledgment must go to my outstanding advisor and chair of my thesis committee Dr. Shira L. Broschat who has always been willing to support me, who is extremely understanding and helpful, and who has encouraged me and shown patience with so many hours of mentoring me. Without her, it's impossible for me to complete the Ph.D. degree. She is also a great friend, full of love, and I could not have a better mentor than she. Many thanks also go to my committee members Drs. Douglas R. Call and John H. Miller for their valuable time and suggestions, which significantly improved the quality of this dissertation.

I also want to thank my parents Yiming Zhou and Keping Li. Their love and commitment will remain for an eternity with me. I hope I have fulfilled their dreams for me. Thanks to my brother Xuechi Zhou whom I cherish. We both had a very good time during childhood, and I cannot wait to see what the future holds for him. Thanks to my grandparents Yuxi Zhou and Lianli Yang for their love and who made my childhood very special and enjoyable.

Besides my family, there have been a number of people whose support was essential for the completion of my degree. One of my best friends Svetlana Lockwood has given me unconditional and constant support. She has been a role model in both the academic and personal realms. I would like to express my sincere gratitude to her for all she has taught me and for all the effort she has spent helping me. In my daily work I have been blessed with a friendly and cheerful group of fellow students. Yunbing Tan has been invaluable in his willingness to help. Thanks also go to my

friend Gokcen Cilingir.

Special thanks go to Mrs. Lihua Wang for taking good care of me and my life over the past 5 years. She is the most powerful and wonderful woman with whom I have been closely associated. Other special thanks go to Mrs. Trude Marston for her spiritual consulting with me. She shows God's love to me and has helped me to obtain a more peaceful and quiet existence.

I would like to thank the School of Electrical Engineering and Computer Science and the Department of Veterinary Microbiology and Pathology for providing support for my research. Also, I want to acknowledge the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, which gave me great financial support for my studies.

Last but not least, I would like to thank my husband for his companionship over the past 10 years. I want to express my most heartfelt gratitude to my little son Aaron Ruibo Zhang for his love and joy. He is a little angel and gift from God. It's so precious to watch him growing up and learning about the world every day.

COMPUTATIONAL METHODS FOR ESTIMATING GENETIC RELATIONSHIPS

Abstract

by Yunyun Zhou, Ph.D.
Washington State University
August 2012

Chair: Shira Broschat

The advent of molecular methods has altered our approach to the study of the genetic relationships of microbes. In particular, we now have the unprecedented ability to estimate genetic relationships from whole genome sequences whose numbers are increasing exponentially. In this dissertation we examine several computational methods for using genome sequences to infer the genetic relationships of both plasmids and bacteria.

First we describe a method for relating 527 Gram-negative bacterial plasmids based on their genetic sequences. Initial classification of their genetic relationships was accomplished using a computational approach analogous to hybridization of ”mixed-genome microarrays.” Relationships were refined for several clusters by identifying conserved proteins within a cluster. The replication of consistent results produced in a separate study for a small group of IncA/C plasmids and clusters of *Borrelia* plasmids provides evidence that the approach used can correctly predict genetic relationships.

Second, we use the *pClust* program to estimate the genetic relationships of the same 527 plasmid genomes. Protein clusters generated by *pClust* are used to create profiles for each plasmid in the tree, which are then used as correlation filters for classification of a new bacterial plasmid. The major contribution of this work is the development of a method that can be used to construct a tree and, more importantly, to insert a new taxon a posteriori. While this method was developed specifically for plasmids, it can be used with genomes of any kind.

The third project is a study of the genetic relationships of bacteria, more specifically species of the alphaproteobacteria class. Typically phylogeny studies of bacteria are based on the 16S rRNA gene. In this work, however, we again use the software program *pClust* with twelve genomes to generate homologous protein clusters which are then used to construct a tree. The results are compared with a tree constructed using 16S rRNA; while certain features in both trees are similar, the differences indicate that the use of whole-genome sequences may provide a better estimate of genetic relationships.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Understanding the distribution of microbes and their relationships to other organisms is very important for understanding their evolution (Eisen, 2000). In this chapter, we discuss the motivation for our research and explain some popular algorithms and procedures for identifying and classifiying relationships based on genetic traits.

1.1 Biological background and motivation

Plasmids are typically circular DNA molecules that are hosted by organisms in the three major domains: Archaea, Bacteria, and Eukarya. Their genomes vary in size from a few to several hundred kilobase pairs. Plasmids are mosaic in composition with a maintenance “backbone” as well as “accessory” genes obtained via horizontal gene transfer (HGT) (Christopher, 2000). A single plasmid or many plasmids may be carried in the same host cell. Plasmids hosted by bacteria often serve as vectors for the dissemination of drug resistance or virulence in the environment. Thus, it is important to understand their genetic relationships as well as that of their bacterial hosts.

Bacteria form a large domain of prokaryotic organisms. Bacterial genomes consist of conservative regions and dynamic regions. The conservative region usually contains housekeeping genes and has relatively low mutation rates and more stable G-C nucleotide content; the dynamic region accepts DNA from foreign sources or from their environment (Wolska, 2003). Bacteria evolve because of random gene recombination and mutation together with adaptation to environmental changes by means of the mechanism of horizontal gene transfer.

The study of bacterial evolution is challenging because of variance within and between species. This is in large part due to HGT that can change both the morphologies and metabolisms of closely related bacteria (Boucher et al., 2003). The study of plasmid evolution is equally—or even more—challenging because of HGT. However, the recent availability of whole-genome sequences for both many species of bacteria and for many different plasmids has opened up entirely new possibilities for studying the genetic relationships of these microbes in a manner that has been impossible until now. Developing methods to study these relationships poses new challenges requiring collaboration between microbiologists and computational specialists.

The research described in this dissertation is a result of collaboration between microbiologists and computational specialists to utilize whole-genome sequences of both plasmids and bacteria. Our goal was to develop computational methods for estimating the genetic relationships of both plasmids and bacteria.

1.2 Molecular phylogenetic studies

Studying genetic relationships is very important in understanding microbial evolution. Over the past several decades, as technological innovations have been introduced, molecular phylogenetic methods have become more popular than traditional methods based on phenotypic properties or morphological or physiological information (Wiens, 2001). In particular, the advent of polymerase chain reaction (PCR), DNA microarrays, gel electrophoresis, and most recently, relatively inexpensive DNA sequencing techniques has caused successive changes in our molecular phylogenetic approaches. PCR has allowed us to amplify small amounts of DNA. These were used with DNA microarrays to provide quantitative measures of similarities between genomes of two different species. More recently amplified DNA has been used for sequencing of entire genomes, and methods of homologous sequence detection have been employed for many uses, including phylogeny studies (Broschat et al., 2010; Davis et al., 2003).

1.2.1 Homologous sequence detection

The Basic Local Alignment Search Tool (BLAST) is probably the most popular program for homologous sequence detection. It is used for large databases and is very fast at performing local pairwise searches (Altschul et al., 1997). It can search both for DNA sequences or amino acid sequences or a combination of both. ClustalW is a general purpose, multiple alignment algorithm for sequences. Its main advantage is its comparatively low memory usage so it is an optimal choice only in limited cases when memory size is an issue (Chenna et al., 2003). *pClust* is a parallel sequence comparison algorithm that uses the Smith-Waterman algorithm to perform pairwise sequence alignment and the shingling algorithm to form clusters of homologous proteins. It has been shown to be more efficient and accurate than BLAST (Wu et al., 2012).

For homolog detection, there are trade-offs between local and global alignment, speed and accuracy, and pairwise and multiple alignment. Thus, there is no absolutely perfect algorithm for all problems, and which algorithm to use depends on the specific requirements. For our work, we actually use a combination of these algorithms. For example, in chapter 2, we construct the initial tree using BLAST for pairwise sequence comparison but then refine clusters using ClustalW for multiple alignment of conserved sequences. The reason for this is because the initial tree was constructed using the neighbor-joining distance algorithm that requires a pairwise distance matrix as input. For the subtree refinement, we construct the tree based on multiple alignment of conserved genes which are more stable, closely related and, most importantly, much smaller in number. In chapter 3, we use *pClust* to output homologous protein clusters for construction of the 527-plasmid tree and classify new plasmids using a BLAST comparison with the clusters. Although for the new plasmid classification, we could run *pClust* again, computational speed is the key point of consideration.

1.2.2 Computational algorithms for tree reconstruction

The common clustering algorithms for tree constructions are distance-based, maximum parsimony, and maximum likelihood.

Distance-based analysis

Distance-based methods compute the pairwise distance between microorganisms and estimate their genetic relationships based on cluster techniques. The major advantage is that they are computationally fast and therefore good for large data sets. The disadvantage is that the phylogenetic signal within the sequence is lost when all the sequence variation is reduced to a single value and, thus, it performs poorly for very divergent sequences. Distance methods are also sensitive to gaps in the sequence alignment. UPGMA (Unweighted Pair Group Method with Arithmetic mean) and the neighbor-joining method are the most popular ways of reconstructing phylogenetic trees.

UPGMA assumes a constant rate of evolution and uses unweighted distance to cluster the two nearest sequences together. A new node is added at the midpoint of the two, and the distance from the new node to other nodes is the average distance of the two original taxa. The process is repeated and leads to a rooted tree (Sokal and Michener, 1958). The neighbor-joining method starts with an unsolved star tree structure and successively chooses the nearest pair of taxa to join together based on their distances (Saitou and Nei, 1987).

Maximum parsimony analysis

Maximum parsimony assumes minimum evolution and chooses the tree that requires the smallest number of changes in the sequences from a common ancestral sequence. A maximum parsimony tree is obtained from the output of a multiple sequence alignment. Analysis is performed at every position in the sequence alignment (Sober, 1983), and the trees that produce the smallest number of changes for all sequence positions are identified. Maximum parsimony is well-suited for sequences that are very similar and is limited to small numbers of sequences (Galtier and Guoy, 1998).

Maximum likelihood analysis

Maximum likelihood analysis is a model-based algorithm for fitting the best parameters for modeling sequence evolution. The goal is to find the tree that has the highest probability under this model. Maximum likelihood can also be used to successively refine a model (Sullivan and Joyce, 2005) and is good for its accuracy in fitting a realistic model for sequence evolution, but poor in computational efficiency.

1.2.3 Tree reliability assessment

Bootstrapping is used to assess the reliability of the topology of a constructed phylogenetic tree. In some cases, a single tree may not represent reliable relationships, and it is necessary to use a number of trees to obtain more reliable relationships. The use of consensus trees and supertrees is one approach for doing so.

Bootstrapping

Tree building algorithms produce phylogenetic trees but with no measure of how reliable they are. Bootstrapping is a method for assessing the significance of a given topology for a tree. Suppose we have an alignment of M sequences and we randomly pick N sequences from the original M with replacement to construct a tree. We do this repeatedly, each time constructing a tree, and we track the number of times the same topology occurs at each node. The frequency at which the same topology occurs gives us a confidence measure at each node, and this measure is known as the bootstrapping value (Efron et al., 1996).

Consensus tree methods

As consensus tree methods result in one 'representative' tree from many phylogenetic trees, we can describe them as determining pairwise compatible subsets from the union of many hierarchies. There are three types of consensus methods: strict consensus, majority-rule consensus, and greedy consensus. Strict consensus collects the subsets from the input set that are common to all input phylogenetic trees. Majority-rule consensus picks the subsets with more than half that occur in the input phylogenetic trees. Greedy consensus is more computationally expensive; it sorts subsets in descending order according to their frequencies and iteratively adds the subset with the highest frequency to the consensus set if it matches all previously added subsets (Margush and Mcmorris, 1981). In chapter 2, we use majority-rule consensus to determine our consensus tree.

Supertree methods

Supertree methods combine multiple phylogenetic trees to produce the overall best 'supertree' (Bansal et al., 2010). They can be used to combine phylogenetic information from datasets only partially overlapping and from different types of sources, e.g., both molecular and morphological data, and they can be used to reduce problems so that they are more computationally tractable. If given a set of completely overlapping source trees, a supertree method should work exactly like a consensus method in combining phylogenetic information (Gordon, 1986).

1.3 Overview of computational tools

1.3.1 Software packages

There are several software packages, including PAUP, PAML, PHYLIP, and MEGA, that provide popular methods for phylogenetic analysis. PAUP (Phylogenetic Analysis Using Parsimony) was originally used for maximum parsimony analysis, but the latest version 4.0 also includes some simple distance-based and maximum likelihood methods (Swofford, 2002). PAUP supports different working environments including Windows and Unix.

PAML (Phylogenetic Analysis by Maximum Likelihood) is a powerful software package for maximum likelihood analysis of sequences using different models and estimating sophisticated model parameters (Yang, 2007). When we understand the possible evolutionary models for our sequences, this software is the best choice. However, this package cannot do multiple alignment. Therefore, it must be combined with a multiple alignment software package such as ClustalW for complete analysis.

PHYLIP (PHYLogeny Inference Package) supports distance-based, maximum parsimony, and maximum likelihood for bootstrapping or consensus tree analysis (Felsenstein, 1993). PHYLIP is the most complete package for phylogeny studies, covering a broad selection of programs for phylogenetic analysis. However, each program requires users to set options based on complex menus which makes it somewhat difficult to use.

MEGA (Molecular Evolutionary Genetics Analysis) is a visual interface software program for sequence alignment and phylogenetic tree construction, fitting evolutional models and rates, mining online databases, inferring ancestral sequences, and testing evolutionary hypotheses (Margush and Mcmorris, 2005). When no prior information is available for sequence properties, MEGA 5 is helpful for getting a general idea of genetic relationships (Tamura et al., 2011). The models and parameters provided are not very complicated which makes them easy and practical to use. The ability to color highlight interesting taxa is one of the most attractive features of MEGA for phylogenetic tree analysis. In chapter 2, we made use of this feature to assist in the analysis of the distribution of four antibiotic resistance genes.

1.3.2 Programming script languages

The libraries or toolboxes available for phylogenetic analysis are well established and use popular scripting languages such as biopython http://www.biopython.org and bioperl http://www.bioperl.org. These and the bioinformatics toolbox in MATLAB greatly save on coding time. Python and perl are not as powerful as Matlab, but they are open-source and, thus, free. Python is more readable and manageable than perl and thus easier for beginners. However, the bioperl library is more complete than that of biopython because of its long history in bioinformatics. This is especially the case for the application of parsing BLAST/FASTA or for online sequence database searches for which bioperl is much better than biopython. The Matlab Bioinformatics Toolbox http://www.mathworks.com/products/bioinfo/ provides broad usage in algorithm optimization and visualization analysis for Next Generation Sequencing, microarray analysis, mass spectrometry, and gene ontology detection. For our research we actually used a combination of scripts as well as the bioinformatics toolbox. Bioperl and biopython scripts were used to download genomes from the NCBI database in batches and to parse the genomes into protein coding sequences. The Matlab Bioinformatics Toolbox was used to develop optimized algorithms for reconstructing phylogenetic trees. MEGA5 was used for post-stage processing of the phylogenetic trees.

1.4 Overview of this dissertation

As mentioned previously, the availability of whole-genome sequences has given us new ways to study genetic relationships. The task requires cooperation between microbiologists and computational scientists, and it is very challenging work to obtain biologically interesting information. To do so, requires the development of suitable classification methods which is the focus of this dissertation.

In chapter 2, we describe a method for relating a large number of Gram-negative (GN) bacterial plasmids based on their genetic sequences. Complete coding gene sequences of 527 GN bacterial plasmids were obtained from NCBI. Initial classification of their genetic relationships was accomplished using a computational approach analogous to hybridization of “mixed-genome microarrays” (Wan et al., 2007). Because of this similarity, the phrase “virtual hybridization” is used to describe this approach. Protein sequences generated from the gene sequences were randomly chosen to serve as “probes” for the virtual arrays, and virtual hybridization for each GN plasmid was achieved using BLASTp. Each resulting intensity matrix was used to generate a distance matrix from which an initial tree was constructed. Relationships were refined for several clusters by identifying conserved proteins within a cluster. Multiple-sequence alignment was applied to the concatenated conserved proteins, and maximum likelihood was used to generate relationships from the results of the alignment. While it is not possible to prove that the genetic relationships among the 527 GN bacterial plasmids obtained in this study are correct, replication of identical results produced in a separate study for a small group of IncA/C plasmids provides evidence that the approach used can correctly predict genetic relationships. In addition, results obtained for clusters of *Borrelia* plasmids are consistent with the expected exclusivity for plasmids from this genus. Finally, the 527-plasmid tree was used to study the distribution of four common antibiotic resistance genes.

In chapter 3, a software program called *pClust* was used with plasmid genomes to establish the genetic relationships among the 527 Gram-negative bacterial plasmids. In order to classify a new plasmid within the tree structure, an algorithm is developed that is based on a technique used to identify aircraft from radar range profiles is developed. The protein clusters generated by *pClust* are used to create profiles for each plasmid in the tree that are then used as correlation filters for classification of a new Gram-negative bacterial plasmid. A profile for the new plasmid is created using BLASTp. The method is tested using several of the 527 plasmids, and new plasmids are added to the original tree. While the method described in this chapter was developed specifically for plasmids, it can be used with genomes of any kind.

In chapter 4, we studied the genetic relationships of twelve alphaproteobacteria species. Alphaproteobacteria are Gram-negative bacteria that belong to the proteobacteria class. Studying the phylogenetic relationships of species of alphaproteobacteria is very important because they include some major human and animal pathogens (Gupta, 2005). Usually, phylogeny studies for bacteria are based on the 16S rRNA gene. In this work, we use whole-genome sequences to obtain phylogenetic relationships among the twelve alphaproteobacteria. The software program *pClust* was used to generate homologous protein clusters which were used to create profiles for each bacterial genome. The resulting trees (both Jaccard and Euclidean distance metrics were used) were compared with trees obtained using the 16S rRNA gene. While the trees share some similarity, one interesting difference indicates that the analysis using whole-genome sequences may provide some insight into the interrelationships of different families within the alphaprotobacteria class.

A summary of the research reported in this dissertaton and possible directions for future work are presented in chapter 5.

Chapter 2

Genetic study for 527 GN bacterial plasmids

2.1 Introduction

Plasmids are extrachromosomal DNA molecules that are found in many species of bacteria and within taxa from archaea, eukaryota, and bacteria (Bapteste et al., 2007). Sequenced plasmids vary in size from less than 1 kbp to more than 2500 kbp, and plasmids vary in their compatibility with different hosts and with other plasmids within the same host cell (Couturier et al., 1988). Plasmids are considered "mosaic" in composition containing both backbone genes for maintenance and mobile and transmissible genes that encode "accessory" traits (Christopher, 2000). Plasmid genes can be obtained from multiple sources (Boyd et al., 1996) and disseminated by horizontal gene transfer (HGT). HGT is responsible for the dissemination of many of the undesirable traits associated with bacteria, including antibiotic resistance and virulence. In addition, broad-host-range plasmids play an important role in bacterial adaptation to new environments. This provides much of the motivation for understanding the relationships among plasmids. Knowledge of these relationships will help us to better understand how genes are shared horizontally across species boundaries as well as to understand microbial evolution. Unfortunately, HGT itself can obstruct phylogenetic signal, making the study of plasmid evolution challenging.

There are several ways to identify genes that have arisen from divergent sources, including comparison of GC frequency, codon usage, and genomic signatures (Campbell et al., 1999; Karlin, 2001; Karlin and Burge, 1995; Mark et al., 2006; Suzuki et al., 2008). However, there is some debate over whether plasmid mosaicism can be understood from such features (Campbell et al., 1999; Mark et al., 2006). In addition, while molecular methods are frequently used to characterize plasmids (Smalla et al., 2000), there is no sequence analogous to the 16S rRNA sequence in bacteria with which to examine their phylogenetic relationships. Several network-based representations have been used to explore genetic relationships among plasmids (Halary and Leigh, 2009; Popa et al., 2011; Brilli et al., 2008). In particular, Brilli et al. (2008) studied the evolutionary relationships of several Gram-negative bacterial plasmids, including those hosted by *Escherichia*, *Salmonella*, and *Shigella*, using the Blast2Network method. Our work is the first to simultaneously classify a broad and diverse group of Gram-negative bacterial plasmids.

In this chapter we introduce a method for investigating the genetic relationships of 527 Gram negative (GN) bacterial plasmids using their complete gene sequences. We start with a modified virtual mixed-genome microarray method to create an initial tree that describes overall genetic similarity for these plasmids (Wan et al., 2007) using proteins rather than DNA for both “probes” and “targets.” Because virtual hybridization of mixed-genome microarrays is an entirely computational method, protein sequences can be used as readily as DNA sequences. We choose to use protein “probes” and “targets” because doing so is more efficient computationally (amino acid sequences are one-third as long as their nucleotide counterparts) and because differences in silent nucleotide mutations are absent in amino acid sequences. To overcome representational bias due to gene repetition, we use BLASTp on the concatenated amino acid sequences of a plasmid with itself and remove duplicate proteins for each plasmid. After removal of the duplicate proteins, protein sequences are randomly chosen to serve as "probes" for the virtual arrays, and virtual hybridization for each GN plasmid is achieved using BLASTp. Each resulting intensity matrix is used to generate a distance matrix from which the initial tree is constructed. After completion of the initial tree, conserved proteins within a cluster are identified and used to refine the relation ships within the cluster by means of multiple sequence alignment of the conserved proteins.

2.2 Materials and Methods

2.2.1 Data preparation

. In July 2010 the complete gene sequences for 2,171 bacterial plasmids were available in the NCBI genome database (http://www.ncbi.nlm.nih.gov/). Of these, 527 sequences were for Gram-negative (GN) bacterial plasmids with more than 50 putative coding genes (CDS) (supplemental file 1). These were downloaded in FASTA format and translated into amino acid sequences based on putative open reading frames. BLASTp with default parameters was used to remove duplicate proteins within plasmid sequences by blasting the sequence with itself. Duplicate proteins were not removed across plasmids because of the need to reflect a representative distribution within the entire protein population. A protein was considered to be a duplicate for the similarity value as (length of matching sequence)*(BLAST similarity score)/(length of reference protein) ≥ 0.45 (Call et al., 2010). The resulting set of proteins for all 527 GN plasmids after removal of duplications-more than 97,000 in total-was used to obtain "probes" that were randomly selected to create the virtual arrays. Each array consisted of 20,000 proteins, roughly 20% of the total protein population. The probe selection procedure utilized independent sampling with-out replacement.

2.2.2 Selection of number of arrays

Using 20% of the protein pool to construct an array corresponds, on average, to 20% representation of each plasmid on an array; this degree of representation is sufficient for discrimination (Wan et al., 2007). Nevertheless, because probe selection is random, there is no guarantee that plasmids will have equal representation, and therefore sampling bias might be a concern. To overcome potential bias, we can construct a number of virtual arrays and generate the initial tree using a consensus method based on all the array results. The problem then is to determine the number of arrays needed for the analysis. In terms of accuracy of the relationship results, we assume the more arrays that are used, the better. However, the computational expense involved in using BLASTp or "virtual hybridization" for each array makes it necessary to determine an optimum number of arrays—i.e., a number that minimizes the computational cost while minimizing variance.

To determine the optimum number of arrays, we used the average absolute difference (ADD) (Fig. 2.1) as a function of the number of arrays. After virtual hybridization of an array for N different plasmids, an $N \times N$ distance matrix is obtained. Pairwise comparison of two distance matrices results in an $M = N(N-1)/2$ distance vector. For the ADD metric, we sum the absolute difference of the mean distances for $n+1$ arrays and n arrays over all M values. The formula is as follows:

$$ADD = \sum_{i=1}^{M} | < D(n+1, i) > - < D(n, i) > | \tag{2.1}$$

where i takes on values between 1 and M, n is the number of virtual arrays, and $D(n, i)$ is the mean distance value for the n virtual arrays. The ADD is a robust estimator of the absolute difference between two populations of distance matrices. When two populations are exact, the ADD will be zero. As we increase the number of arrays used, we expect the differences between the mean values to decrease (i.e., the mean values will not change as much, for example, between 50 and 51 arrays compared to between 2 and 3 arrays), and thus we expect the value of the ADD to decrease as the number of arrays used increases. For $n = 1$ the ADD is approximately 60,000 (Fig. 2.1). It then declines quickly, and for $n = 20$, the value is less than 2,000. For 50 and 99 arrays, the values are 679 and 379, respectively. Note that to calculate the value of the ADD for 99 arrays, 100 arrays are used. If we assume the distance between each location in the two matrices is the maximum distance we obtained minus the minimum possible distance of zero, the maximum value for the ADD is 301,707. The ADDs for 50 and 99 arrays are then approximately 0.002 and 0.001 of this maximum value, respectively. While the small fractional difference of 0.002 indicates that the use of 50 arrays is probably sufficient, we conservatively chose to use 100 arrays.

2.3 Results

2.3.1 Initial tree construction

As indicated previously, each array was populated by 20,000 randomly chosen protein "probes" (entire proteins were used) and a total of 100 arrays were "constructed."

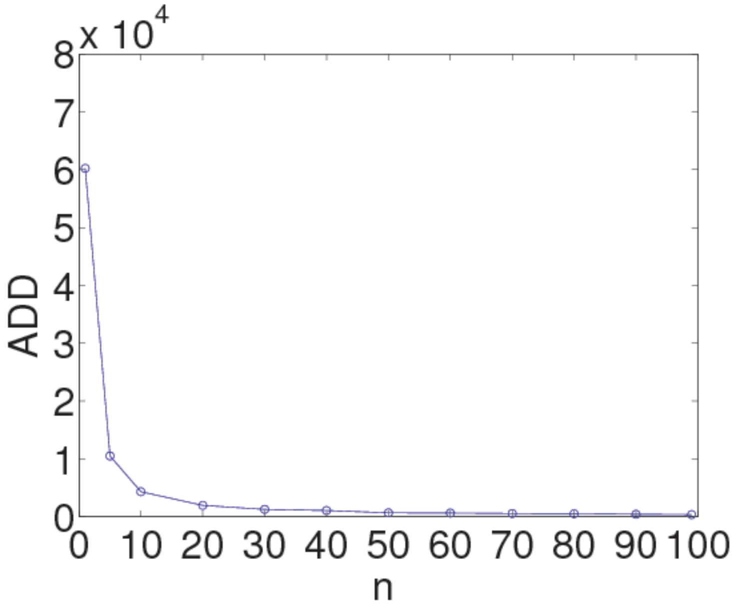


Figure 2.1 Absolute distance difference (ADD) as a function of the number of arrays (n) used for averaging. The ADD value decreases as the number of arrays increases—i.e., the ADD curve is more steep for $n = 20$ than $n = 50$ or 99.

For each array virtual hybridization was performed for the 527 GN plasmids. Hybridization was simulated using stand-alone BLASTp with default parameters and normalized intensities between 0 and 1 were obtained using the formula 2*(length of matching sequence)*(BLASTp similarity score)/(length of reference protein + length of matching sequence protein) resulting in 100 ($527 \times 20,000$) normalized intensity matrices. From each intensity matrix, a distance matrix was calculated using pairwise Euclidean distances. Two methods were used to obtain consensus trees from the 100 distance matrices: majority voting (Margush and Mcmorris, 1981) and averaging. Because majority voting is order dependent for an unrooted tree, human and cattle proteins were used as an out-group to force clustering of the GN plasmids (these proteins were added to the arrays and during virtual hybridization all intensity values for these proteins were zero for the plasmids). Consensus from Phylip was then used to obtain the initial tree (Felsenstein, 1989). For the second method the 100 distance matrices were averaged to obtain a single distance matrix and neighbor joining was used to construct the initial tree.

The shear size of the consensus trees for the 527 GN bacterial plasmids prevents inclusion of them as figures. Instead, files for these trees in Newick standard format are included as supple-mental files 2 and 3 that may be viewed in circular format using freely available, tree-generating software packages such as MEGA5 (Tamura et al., 2011). To compare results for the two types of consensus trees, we considered the intensity results for a subset of 50 of the 527 GN plasmids. Each virtual array was populated by randomly selecting 1,400 probes from a protein pool of 6,795. As with the 527 plasmids, 100 arrays were used to obtain 100 distance matrices. Figs. 2.2 and 2.3 show the majority voting and averaged distance results, respectively. There are some limitations to comparing the two trees because; for example, one is unrooted (majority voting) and the other is rooted. Nevertheless, recall that both trees are generated using all proteins and reflect only the initial relationships. Hence, distances are unimportant, and we focus on comparison of the clusters. While the trees have different topologies, the clusters are virtually identical.

2.3.2 Cluster refinement

The initial tree is constructed using the entire pool of proteins including those obtained via HGT. Relationships between clusters can be further refined using conserved genes (Call et al., 2010). Conserved genes are the genes shared among clusters and are considered to be orthologous genes and, thus, are less likely to represent cases of HGT-that is, genetic content from a recent common ancestor can be used to refine relationships within clusters, while genes from a common ancestor should contain more information about the evolutionary relationship among plasmids. It would take considerable computational time to refine all the clusters comprising the 527-plasmid tree, and there would be no meaningful way to assemble the resulting clusters because distances would be inconsistent across clusters. In addition, there is no benefit to refining a cluster unless it is of interest to a particular investigator. To illustrate the results of refining a cluster, we consider the two clusters shown in Figs. 2.4 and 2.6, one from the *Borrelia* group and the other a group of plasmids from several species. Both clusters form part of the full 527-plasmid tree obtained using majority voting. Conserved proteins were identified via BLASTp searches using the same approach described in (Call et al., 2010) except using proteins rather than genes; the similarity cut-off value was 0.3. In addition, when either the length of an aligned sequence was short relative to the reference sequence or the BLASTp identity score was low, the protein was excluded from the analysis. The number of conserved proteins identified at each cluster node is shown. Figs. 2.5 and 2.7 shows the results after multiple-sequence alignment has been performed.

2.4 Discussion

The method described in this chapter provides a new in silico approach to study genetic relation-ships among plasmids. One useful outcome of this type of analysis is that we can consider questions about the dissemination of antibiotic resistance genes among bacteria as discussed below. While there is no “gold standard” to assess the “accuracy” of the proposed relationships Fig. 2.2-2.5, there are several independent observations that support the validity of our results. One consistent finding was the clear demarcation of *Borrelia* plasmids as separate from plasmids from Enterobacteriaceae (Baker et al., 2007) both in the 50-plasmid analysis Fig. 2.2 and in the full plasmid analysis (supplemental files 2 and 3). *Borrelia* is a distinct genus of organisms from the spirochete phylum and is mostly known for pathogenic species that are responsible for vector-borne zoonotic infections (e.g., Lyme disease). Plasmids available for *Borrelia* in Genbank include both circular and linear plasmids. The latter are rare for members of Enterobacteriaceae so we would not expect them to be genetically related. Furthermore, the only plasmid sequences from relapsing fever cluster closely together (supplemental files 4 and 5) as should be expected given that *B. recurrentis* (NC_011246) is considered to be very closely related to *B. duttonii* (NC_ 011247) (Lescot, 2008). The *Borrelia* plasmids that are classified as "cp32" are also grouped together (additional files 4 and 5; NC_011720, NC_011842, NC_012106, NC_012253) as expected given that these plasmids are considered to arise from prophage (Eggers and Samuels, 1999). *Borrelia* plasmids that are required for infectivity, including lp25 (NC_012166, NC_011856), lp36 (NC_011857, NC_011867, NC_001855, NC_012184, NC_012202), and lp38 (NC_012167, NC_012182) are also clustered together (Purser and Norris, 2000)(supplemental files 4 and 5). Another notable and consistent grouping in our analysis includes incompatibility A/C plasmids (IncAC; Figs. 2.5 and 2.7 that have been well described in the literature (Call et al., 2006; Fricke et al., 2009; Welch et al., 2009) and should be expected to cluster closely together.

2.4.1 Distribution of Four Antibiotic-Resistance Genes

Importantly, the 527-plasmid tree can be used to study the distribution of genes of particular interest. To illustrate this we considered four common antibiotic resistance genes, *tetA*, *tetB*, *cat*, and *floR*. We first performed a sensitivity analysis to identify orthologous resistance genes using the coefficient of variation to identify a similarity threshold (45% similarity in this case; data not shown). Next a BLASTp search of all 527 plasmid sequences resulted in identification of 40 plasmids with one or two of the four AR genes (Fig. 2.8). In addition to the relatively few plasmids with these antibiotic resistance genes, it is striking that 37 of the 40 plasmids group within a single 115-plasmid cluster (Fig.2.8). We examined the 115-plasmid cluster further (supplemental file 6) and found several features of interest. Plasmids from Escherichia coli compose 51.3% of the 115 plasmids. This contrasts with their representation in the remaining 412 plasmids, which is only 1.5%. Of the 37 plasmids with antibiotic resistance genes within the cluster, 15 (41.7%) are from E. coli. While most of the 115 plasmids are from bacteria belonging to the Enterobacteriaceae family, a number of them are not, although all belong to the class Gammoproteobacteria. Importantly, not all plasmids associated with the Enterobacteriaceae family or the Gammoproteobacteria class are in the cluster. Removal of the antibiotic resistance genes had an insignificant impact on the clustering of the 115 plasmids confirming that this grouping was not a consequence of the antibiotic resistance genes themselves. Finally, the size distribution (number of proteins) of the 115 plasmids is significantly smaller than the size distribution of the entire set of 527 plasmids. Thus, the antibiotic resistance genes we examined are most closely associated with smaller plasmids of the Enterobacteriaceae family, which indicates that a subpopulation of plasmids is probably responsible for dissemination of these resistance traits in nature.

2.5 Conclusions

We present an in silico approach for simultaneously establishing the genetic relationships among 527 Gram-negative bacterial plasmids. The method uses complete gene sequences for plasmids with at least 50 coding genes to create 100 virtual arrays. These arrays are used to construct an initial tree by consensus that can be refined cluster by cluster using multiple-sequence alignment of conserved proteins within each cluster. While it is not possible to confirm the accuracy of the consensus trees, known relationships from both *Borrelia* and IncA/C plasmids were reflected accurately in our analysis. Based on our results, one can construct additional hypotheses about both inter- and intra-genus transmission of plasmids.

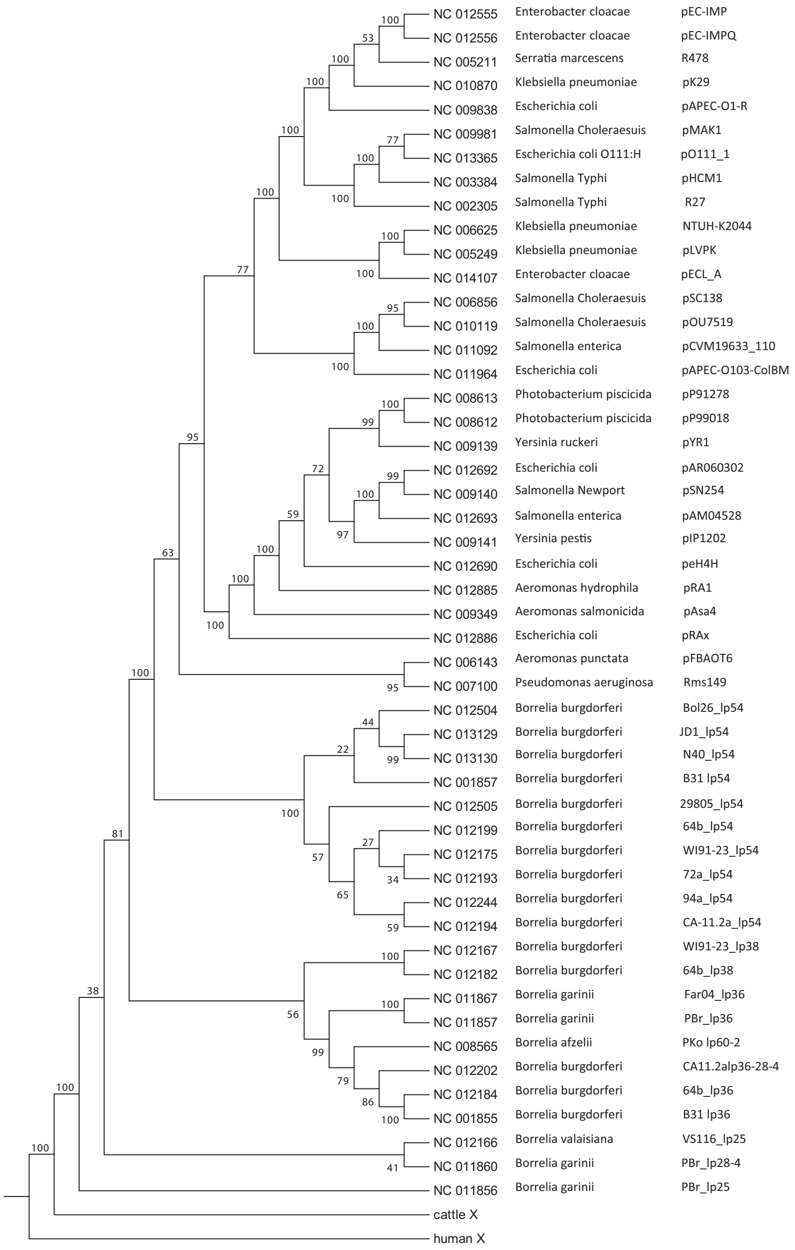


Figure 2.2 Majority voting consensus tree. The majority value of clusters has been listed on each node. A cluster is confirmed if it is in a majority (50%) of a tree.

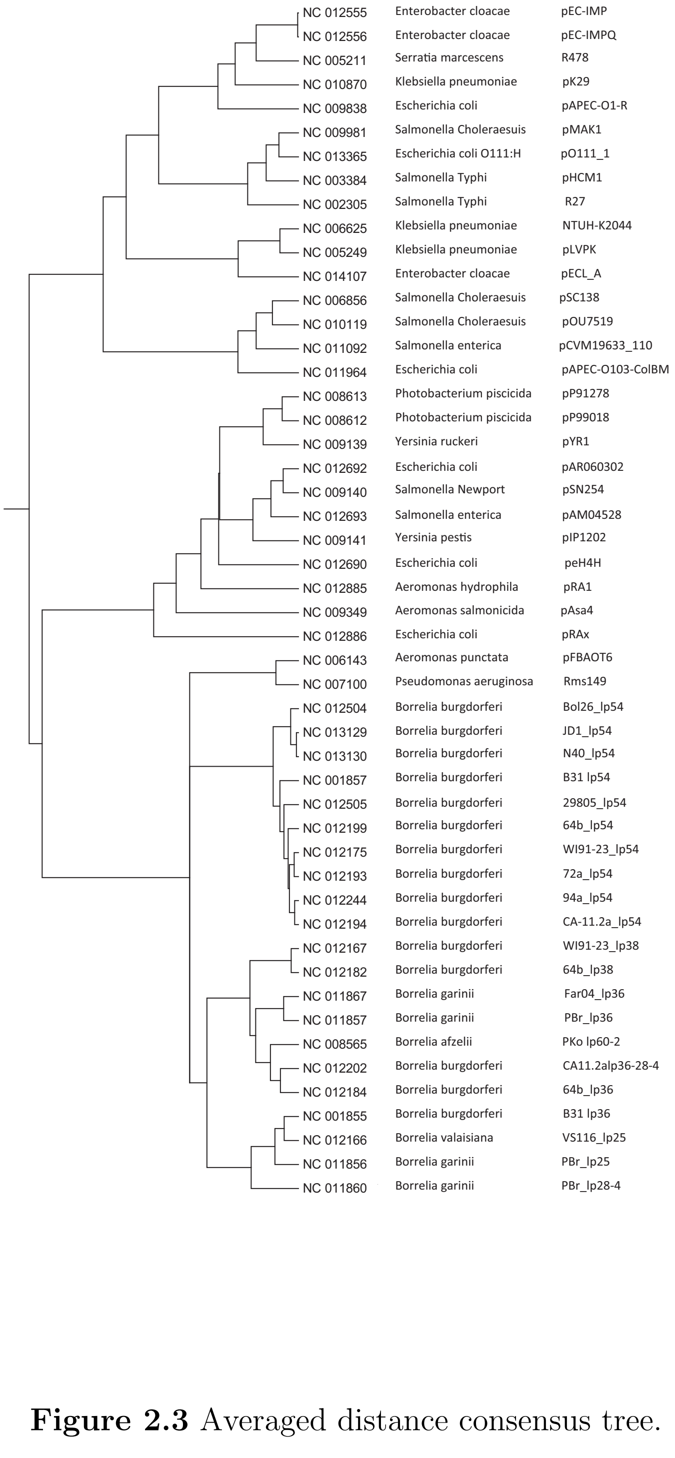


Figure 2.3 Averaged distance consensus tree.

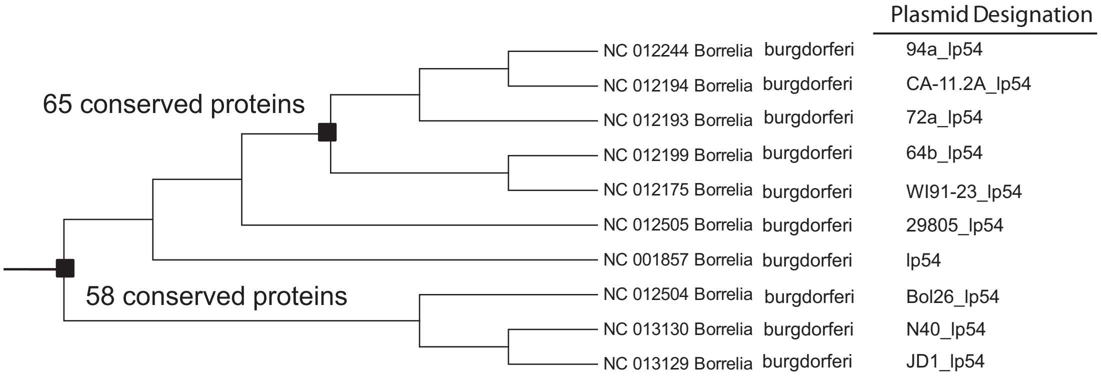


Figure 2.4 *Borrelia* cluster from initial tree; NC_012175 is the reference plasmid for determining the conserved proteins.

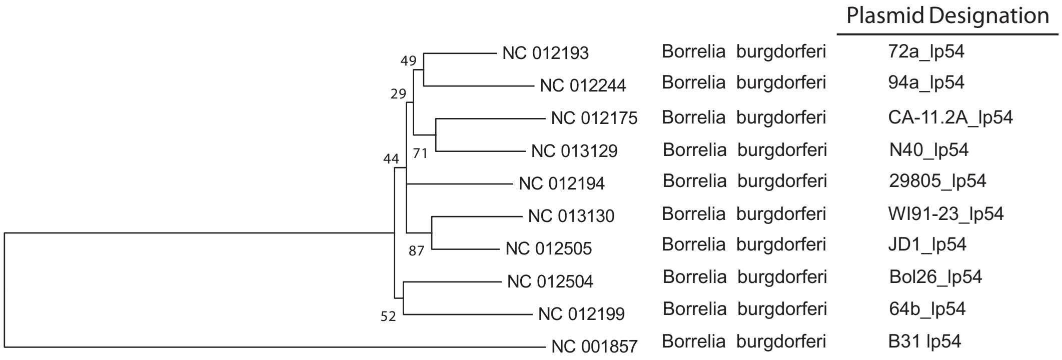


Figure 2.5 *Borrelia* cluster after refinement.

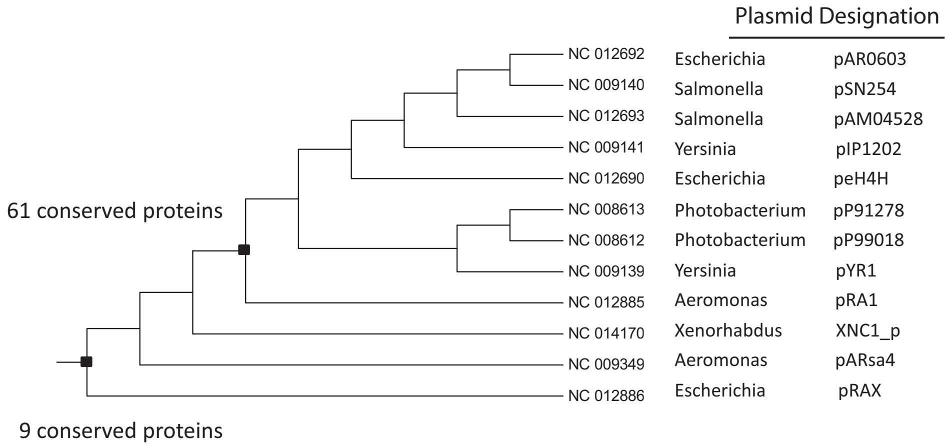


Figure 2.6 Mixed bacterial plasmids cluster from initial tree including 8 IncAC plasmids; NC_012693 is the reference plasmid.

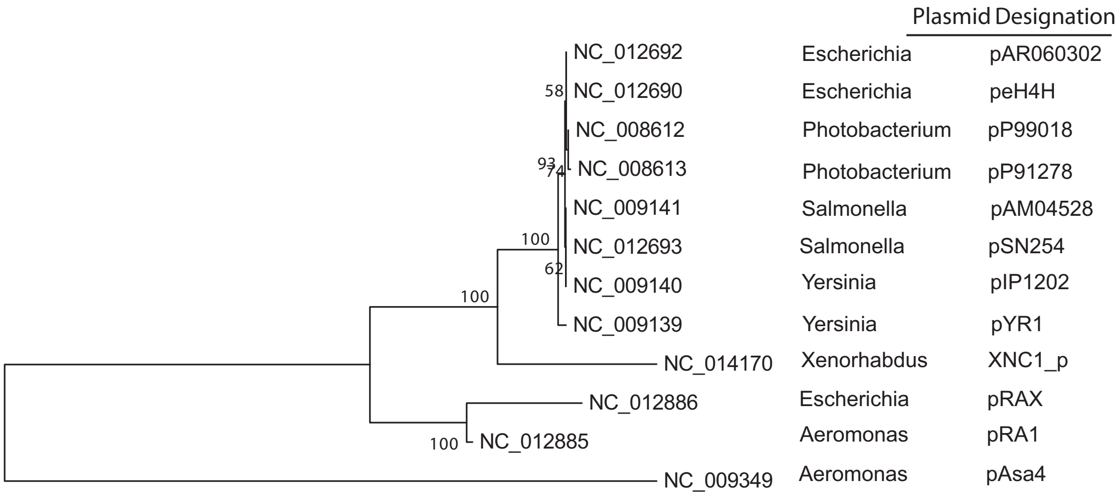


Figure 2.7 Mixed bacterial plasmids cluster after refinement.

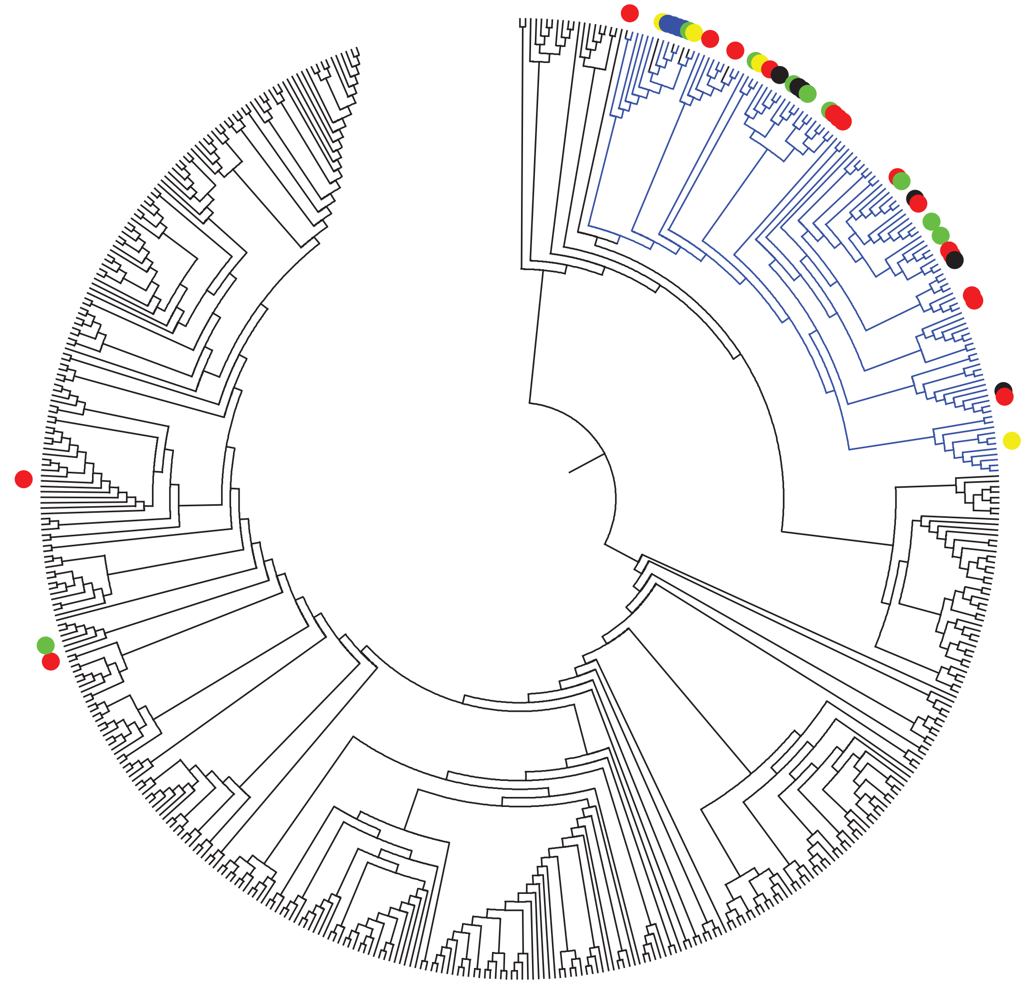


Figure 2.8 Distribution of antibiotic resistance genes *tetA* (red), *cat* (yellow), *tetB* (green), *tetA* and *floR* (blue), and *tetB* and *cat* (black). Thirty-seven of the 40 plasmids carrying antibiotic resistance genes are within a 115-plasmid cluster. Labeling of taxa in the figure is not possible due to size limitations.

Chapter 3

Using correlation filters from protein clusters for classification

3.1 Introduction

Plasmids are typically circular DNA sequences that can transfer and replicate in bacteria and can be classified as broad or narrow host-range (Bapteste et al., 2004; Boyd et al., 1996). Understanding the genetic relationships among plasmids is important in the study of microbial evolution, in medical epidemiology, and in assessing the dissemination of antibiotic resistance genes (Dennis, 2005; Couturier et al., 1988). The gene composition of microbes such as bacteria is more conservative than that of plasmids; this is true regardless of evolutionary distance and is mainly due to the mechanism of horizontal gene transfer (HGT) (Ochman et al., 2000). Genes of plasmids coding for environmentally adapted traits such as antibiotic resistance are transferred between organisms via HGT (Thomas, 2000; Thomas and Neilsen, 2005). Our objective is to develop an accurate and efficient method for establishing genetic relationships among a diverse group of plasmids that also allows the addition of a new plasmid to the resulting network or tree. This method is not restricted to plasmids, but in principle can be used for genomes of any kind.

There are a number of approaches to understanding plasmid relationships. Some researchers focus on the identification of HGT cores, which can reflect how gene transfer or replication occurs among plasmids or hosts (Karlin and Burge, 1995; Huang and Gogarten, 2006). Some work is related to comparison of compositional features such as genomic signatures and codon usage (Karlin, 2001; Passel et al., 2006). Some researchers use network-based representations to explore genetic relationships among plasmids (Halary et al., 2009; Popa et al., 2011; Brilli, 2008). The most popular approach for sequence comparison is computing all-against-all pairwise sequence similarity and then using heuristic techniques to cluster sequences (Bateman et al., 2004; Sasson et al., 2003). However, the problem with all-against-all comparison it that it is computationally expensive. To increase the speed of comparison, BLAST is widely used, but BLAST has relatively low sensitivity (Altschul et al., 1997; Shpaer et al., 1996).

Because BLAST is not guaranteed for optimality of an alignment, in this work we use an alternative sequence comparison algorithm *pClust*. The output of *pClust* is clusters of homologous proteins (Wu et al., 2010). We use these clusters to construct a tree for 527 Gram-Negative (GN) bacterial plasmids and then predict the relationship of new plasmids within the structure of this tree—that is, we insert a new plasmid into the existing tree. This is much more efficient than constructing an entirely new tree which is computationally costly. The approach used to accomplish this is based on a technique introduced for discriminating between military and commercial aircraft using a library of radar range profiles as correlation filters. We create a library of 527 plasmid protein cluster profiles that are used as correlation filters to determine whether the new plasmid fits within the tree structure.

To utilize the classification algorithm for new plasmids, we focus on 12 pAKD plasmids isolated from Norwegian soil that encode mercury resistance (Sen et al., 2011). These 12 pAKD plasmids belong to incompatibility groups IncP-1(β) and IncP-1(ε). A phylogenetic tree constructed using multiple alignment of the relaxase gene *traI* is presented in Sen et al. (2011), and we compare the results for these 12 plasmids with our results.

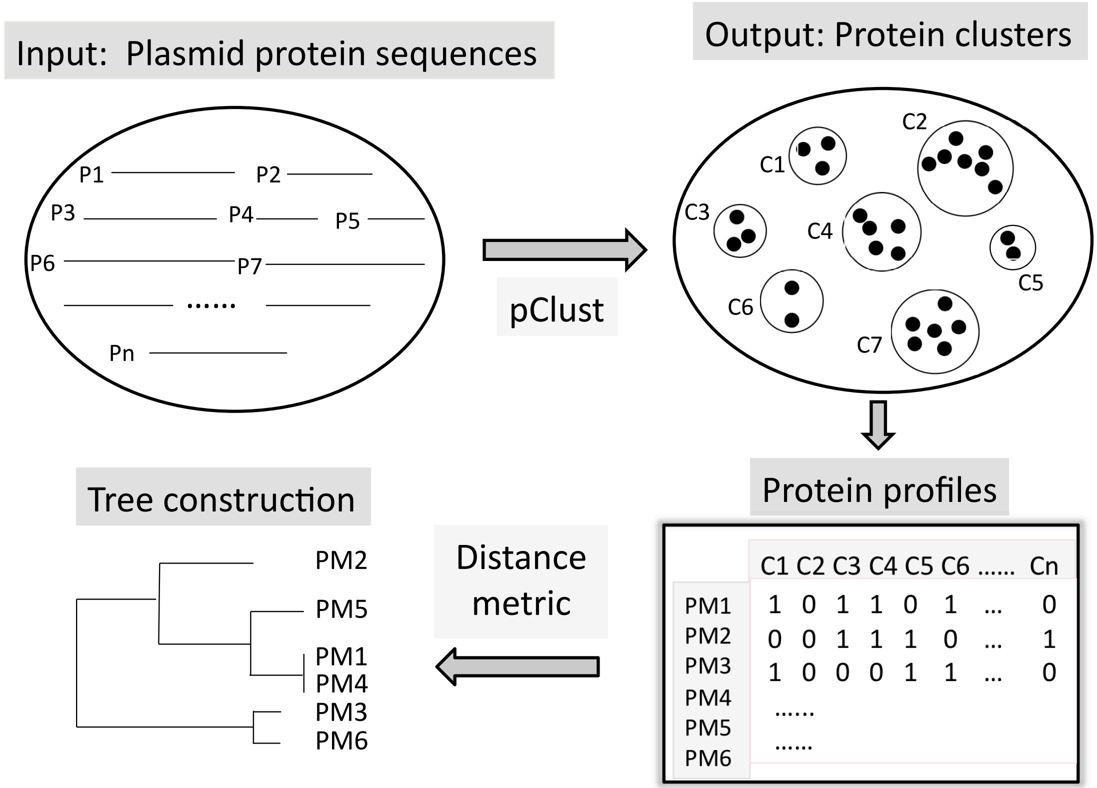
3.2 Methods

3.2.1 Data preparation

In Zhou et al., (2012), a virtual hybridization method was used to construct a tree for 527 GN bacterial plasmids. To allow comparison with this earlier work, we use the same group of plasmids in the current work. Complete gene sequences of GN bacterial plasmids with more than 50 putative coding genes were converted to protein sequences. BLASTp with default parameters was used to remove duplicate proteins within plasmid sequences using the formula (length of matching sequence)*(BLAST similarity score)/(length of reference protein) ≥ 0.45—that is, proteins with scores ≥ 0.45 were considered to be duplicates (Doug et al., 2010). After removal of duplicate proteins, more than 97,000 protein sequences remained.

3.2.2 Clustering algorithm

Our clustering algorithm consists of two parts, plasmid tree construction and insertion of a new plasmid into this tree, as depicted in Figs. 3.1 and 3.2.



	C1	C2	C3	C4	C5	C6	......	Cn
PM1	1	0	1	1	0	1	...	0
PM2	0	0	1	1	1	0	...	1
PM3	1	0	0	0	1	1	...	0
PM4	......							
PM5	......							
PM6								

Figure 3.1

Tree construction

The flowchart in Fig. 3.1 shows the approach used for constructing a tree from the 97k+ plasmid protein sequences. The protein sequences *P1, P2,...,Pn* are used as input into the *pClust* program (Wu et al., 2010), which employs the Smith-Waterman algorithm to perform pairwise comparison of a subset of the sequences. The output from *pClust* is clusters *C1, C2,...,Cm* of similar proteins. Protein profiles *PM1, PM2,...,PMl* are then created for all the plasmids from the *pClust* output files. As

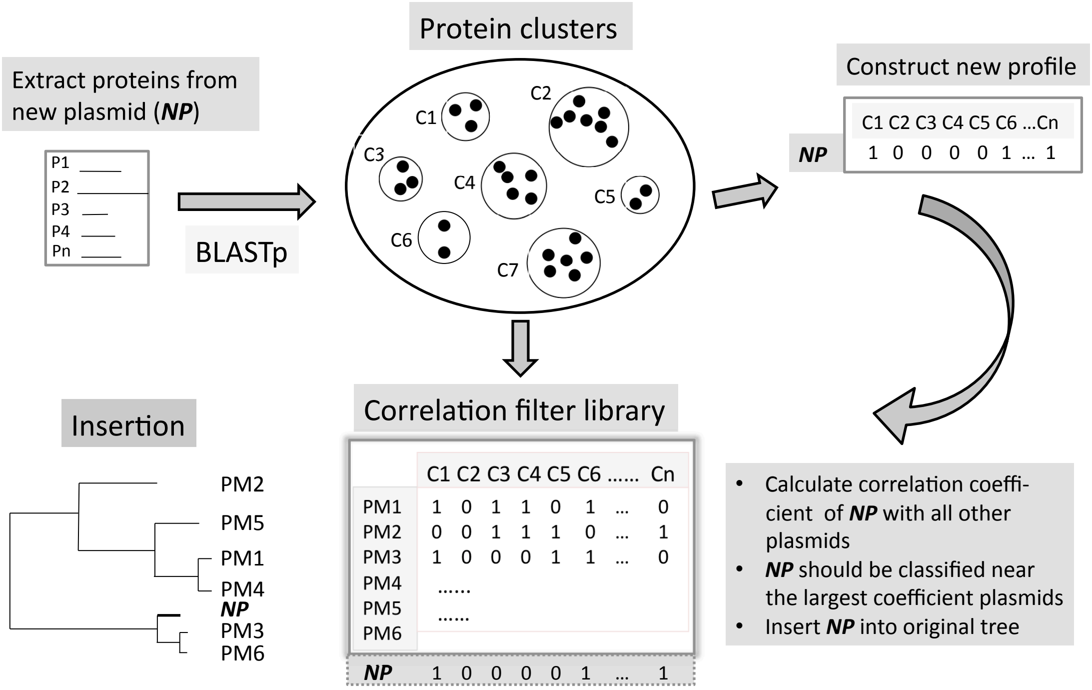


Figure 3.2

shown in Fig. 3.1, each profile consists of a binary sequence with 1 indicating the presence of a protein and 0 indicating absence. The *pClust* software was used with default settings except for ExactMatchLen for which a value of 4 was used. A total of 6,618 clusters (defined as having at least two proteins) were found by *pClust*.

The resulting 527×6618 binary matrix was used to construct the tree. Two different distance measures were used, the Jaccard distance metric, which is one of the measures used for binary matrices, and the Euclidean distance metric. Neighbor joining was used to obtain the final tree.

Insertion of new plasmids

As additional plasmid gene sequences becomes available, we can repeat the procedure described in the previous section to obtain a new tree. However, the amount of computation and time required to accomplish this task is excessive considering the incremental gain achieved. For example, execution time for the 527-plasmid tree was 72 hours on an Intel Xeon CPU E5420 machine with 32GB of memory. Instead it is preferable to have a means of inserting new plasmids into the existing tree structure as described in this section, where execution of the described insertional algorithm takes a few minutes on a laptop computer.

The procedure for inserting a new plasmid into an existing tree constructed by means of *pClust* is shown in Fig. 3.1. Proteins *P1,P2,...Pn* from a new plasmid are extracted from the plasmid genome following its conversion into a protein sequence. BLASTp is performed with these proteins against all the proteins in the 6,618 clusters to determine the protein profile for the new plasmid. A protein is considered to be a member of a cluster when its similarity score is above 0.2. The similarity score is given by (length of matching sequence)*(BLAST similarity score)/(length of reference protein). The cutoff value of 0.2 is consistent with the 40% sequence similarity used as a parameter setting in *pClust*. Correlation filtering is then performed with the correlation filter library that consists of the protein profiles of the original 527 GN bacterial plasmids. The Pearson's product-moment correlation coefficient, whose absolute value is less than or equal to 1, is used to measure the correlation between two profiles (Rodgers and Nicewander, 1988; Stigler, 1989). The larger the value, the more similar two profiles are. This value is used to determine whether the plasmid fits into the tree and, if so, where it should be located; this is explained in the discussion section. When appropriate, the new phylogenetic profile is added to the binary matrix, and a tree is constructed from the entire matrix as described in the previous section.

To test the performance of our insertion algorithm, we used a leave-one-out, cross-validation strategy. We randomly selected a plasmid and constructed a tree from $n-1$ ($n = 527$) plasmid genomes. We then used our insertion algorithm to reinsert the missing plasmid. Tests were performed for five different plasmids, and each time the plasmid was inserted into its original location in the tree.

3.3 Results

3.3.1 527-Plasmid Tree

Following the procedure described above, a tree was constructed for 527 GN bacterial plasmids. Because of its size, it is not shown, but it is available as supplementary information in Newick standard format (.nwk) for both Jaccard and Euclidean distance metrics and can be viewed using MEGA5 (Tamura et al., 2011). A tree constructed using the Jaccard distance metric for the same subset of 50 plasmids used in Zhou et al. (2012) is shown in Fig. 3.3, and the Euclidean distance version is shown in Fig. 3.4. Comparison of the tree in Fig. 3.3 with its counterpart in Zhou et al. (2012) shows a few small differences. The tree constructed using the Euclidean distance metric is actually closer to the one shown in (Zhou et al., 2012), but the Jaccard tree does a better job of clustering the *Borrelia* plasmids (Lescot et al., 2008; Purser and Norris, 2000).

3.3.2 Algorithm Testing and Insertion of New Plasmids

To validate our insertion algorithm, we randomly selected two plasmids, NC_008613 (*Photobacterium*) and NC_013365 (*Escherichia*), from different bacterial species and, excluding each one in turn from the original matrix, verified that they were correctly inserted into their original locations in the 527-plasmid tree. Protein profiles for each of these were generated and filtered using the remaining 526 correlation filters in the library. The largest correlation coefficient for NC_008613 was found to be 0.8957 with NC_008612. Thus, after generation of the tree with inclusion of the profile for NC_008613, we expect it to be closest to NC_008612 as was the case. The same was true for NC_013365 which correlated best with NC_003384 having a correlation coefficient of 0.8912.

Next we applied our correlation filter classification algorithm to 12 new plasmids from the pAKD family (Sen et al., 2011). The 12 plasmids cluster together and are most closely grouped with genera typical of other soil bacteria. The correlation coefficient values among the pAKD plasmids were found to be ¿0.7 and to decrease relative to the other plasmids with distance to ¿0.5 (Fig. 3.5). pAKD plasmids 16, 25, and 34 belong to the IncP-1(ε) compatibility group and form a cluster; pAKD plasmids 1, 14, 15, 17, 18, 29, 31, and 33 cluster as the IncP-1(β) comparability group. Although pAKD26 falls into the IncP-1(ε) clade, it should be in the IncP-1(β) group. However, the placement is distal from the eight other plasmids in the β group, and pAKD26 was actually designated as IncP-1β-2 to differentiate it from the other eight plasmids as recently described in Norberg et al. (2011). Our results are consistent with Sen et al. (2011).

3.4 Discussion

In this work, we presented a new method for constructing a tree from plasmid genome sequences based on the use of *pClust*, a software program developed for homology detection for large-scale protein sequence analyses, and for efficiently inserting a new plasmid into the tree based on the use of a correlation filter library. Two aspects of this method deserve further discussion. The first is a question of which distance metric, Euclidean or Jaccard, gives a more accurate tree or whether a different distance metric would give superior results. We will not address this latter point, but regarding the former, we suggest that users should use both metrics, and the decision as to which one is more accurate should be determined on the basis of the biology of the system. In theory, the Jaccard distance metric gives better results for a binary matrix. However, the results based on Euclidean distance compare favorably with those obtained for a non-binary intensity matrix using a different approach (Zhou et al., 2012).

The second aspect for further discussion is interpretation of the correlation coefficient. The correlation coefficient is used to check the final tree—i.e., a new plasmid should be located near the plasmid with which it is most highly correlated. However, the correlation coefficient is used additionally to determine whether a plasmid should even be inserted into a tree. In other words, how large should the correlation coefficient be before it can be considered to be biologically meaningful? Several works offer guidelines for the interpretation of a correlation coefficient (Buda, 2010; Cohen, 1988), but all criteria are in some way arbitrary and ultimately interpretation of a correlation coefficient depends on the purpose. For our case, we chose a value of 0.5, but we also require biological evidence—for example, that a plasmid is, in fact, from a Gram-negative bacterium.

To further examine the correlation coefficient, we randomly selected 10 Grampositive bacterial plasmid genomes from 10 different genera. The largest correlation coefficient was 0.234 and the smallest 0.112. GP bacterial plasmids clearly do not belong in our GN bacterial plasmid tree and our minimum value of 0.5 suffices to differentiate between these groups; nonetheless, the correlation coefficient can still be used to determine their nearest GN neighbors. Note that the 527 GN bacterial plasmids we considered do not represent the full spectrum of GN plasmids. Thus, it is possible to obtain a small correlation coefficient value for a new, uncharacterized GN plasmid. This plasmid can be classified within the tree structure, and by incorporating the new plasmid sequence information into our library, we can predict future plasmids that may be closely related to it.

While the method of inserting new plasmids into an existing tree is fast and efficient, at some point generation of a new tree using all proteins from all the taxa will probably be required. We do not know at what point this will occur, but we assume it will be necessary eventually to insure that all possible protein clusters are included. Recall that a cluster must contain at least two proteins to be considered a cluster. Thus, any new plasmid containing a protein that would have formed a cluster with a discarded protein represents incomplete information in the library—i.e., perhaps the total number of clusters will ultimately be much greater than 6,818.

3.5 Conclusions

In this work, we provide a new classification method to first construct a tree from genomes converted to protein sequences and to then insert a new taxon into this tree. The method uses *pClust)* to cluster similar proteins and to create a correlation filter library. BLASTp is used to create a profile for a new taxon, and correlation filters are used to determine whether it belongs in the tree and, if so, where is should be located. The method was developed to examine the genetic relationships among 527 Gram-negative bacterial plasmids and to allow insertion of a new plasmid into a tree, but it can be used for any species.

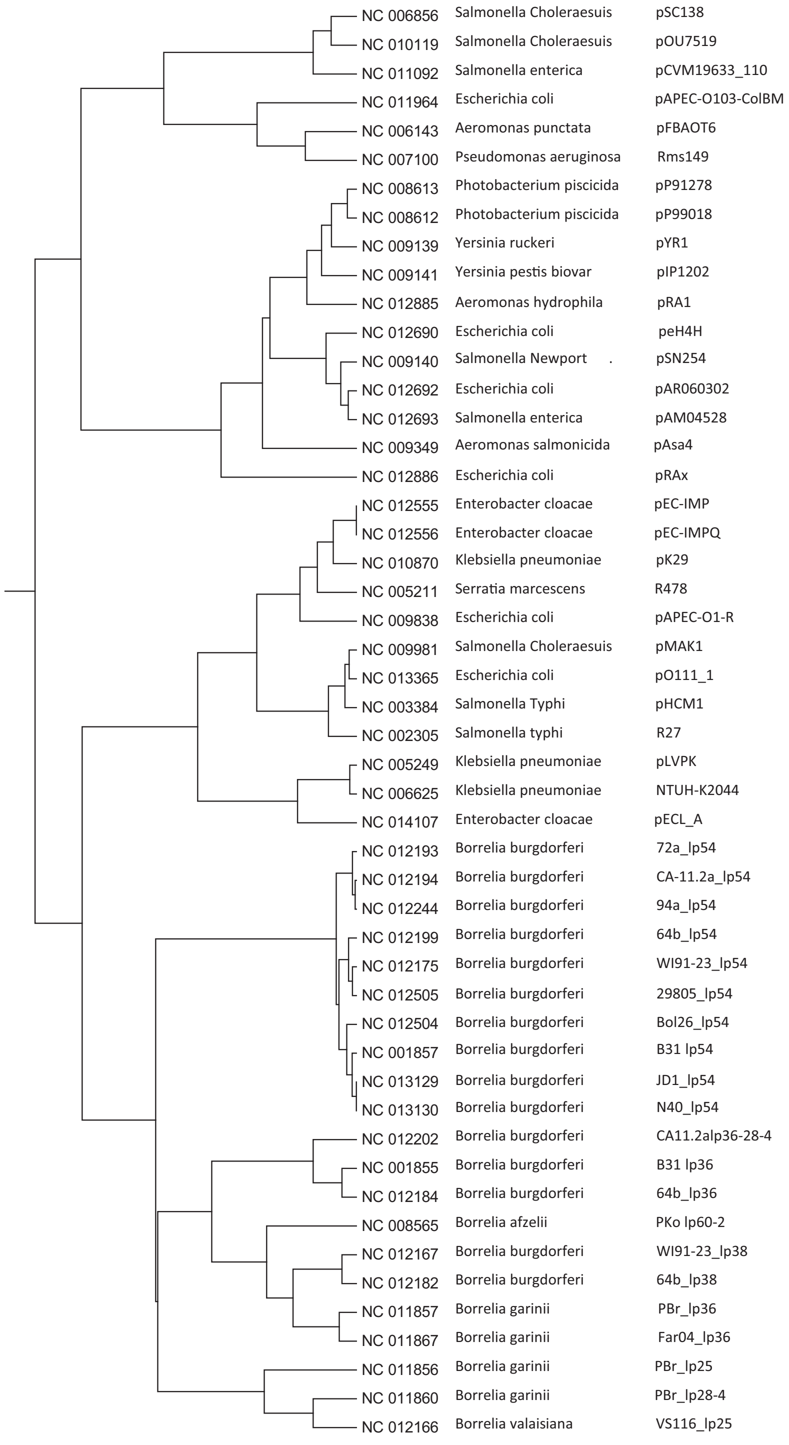


Figure 3.3 Neighbor joining phylogenetic tree of 50 plasmids based on Jaccard distance matrix

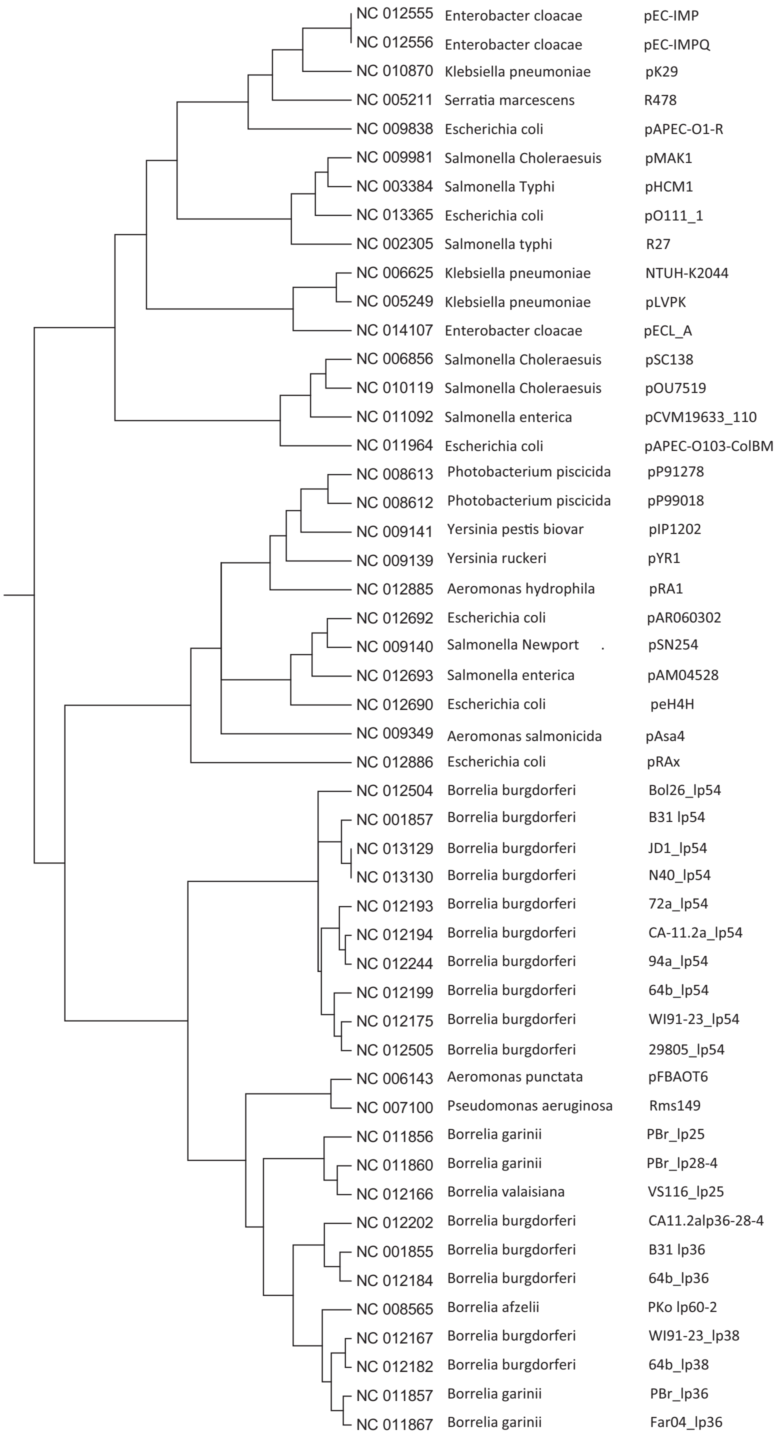


Figure 3.4 Neighbor joining phylogenetic tree of 50 plasmids based on Euclidean distance matrix

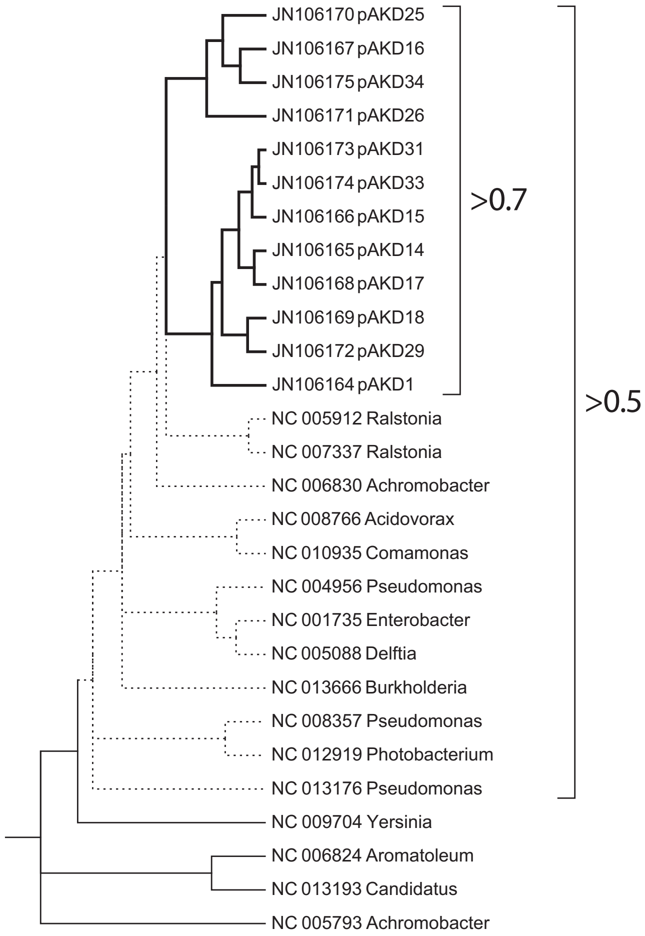


Figure 3.5 pAKD family.

Chapter 4

Phylogenetic classification of bacteria using whole-genome sequences

4.1 Introduction

Because ribosomal RNA is highly conserved and the rate of nucleotide changes is slow and predictable, it has become the standard tool for inferring bacterial phylogeny (Woese et al., 1990). However, a number of reports indicate that it is impossible to explain all bacterial evolution using a single gene (Wheelis et al., 1992; Eisen, 1995; Garrity and Holt, 2001). Because of complicating factors such as horizontal gene transfer, misalignment, and differing evolutionary rates, using 16S rRNA alone to determine genetic relationships can be inaccurate and result in incorrect topologies (Marshall, 1997; Fitz and House, 1999). Also, in spite of the success of rRNA microbial taxonomy, many evolutionary relationships are still unclear. Thus, it is desirable to develop a method for constructing bacterial phylogenies from a large number of genes.

Current methods for studying phylogenetic relationships at the genome level are mainly based on sequence alignment and analysis of a large number of conserved genes (Feng et al., 1997; Brown, 2001; Roux and Rydkina, 1997; Daubin et al., 2001; Eisen, 1995), comparison of the presence or absence of homologous genes (Yeh et al., 2001; Lin et al., 2011), or comparisons of whole genomes (Snel et al., 1999; Tekaia et al., 1999; House and Fitz-Gibbon, 2002; Wolf et al., 2010; Bansal, 2002). In this work we consider an approach that uses all the coding genes in whole-genome sequences to construct a phylogenetic tree.

Alphaproteobacteria are relatively well-characterized taxonomically using traditional methods, and a number of complete genome sequences are available (Sasson et al., 2003). Thus, they serve as a good test case for our approach. Moreover, many genera (e.g., *Rickettsiales*, *Brucella*, and *Bartonella)*) are major human and animal pathogens.

In this study we compare phylogenetic trees for 12 species belonging to 11 genera of the alphaproteobacteria class. The trees are constructed using the 16S rRNA sequence and a new method that uses whole-genome sequences. For the 16S rRNA trees we use both unweighted and Weighbor weighted bootstrapping with the neighbor joining method, and for the whole-genome method we use two different distance metrics, Euclidean and Jaccard, with neighbor joining. The whole-genome approach uses the open-source software program *pClust* to cluster all homologous proteins into groups. The goal of the present study was to extract information from whole-genome sequences that can be used to gain further insight into the taxonomy and evolution of bacterial species.

ORGANISM	ACCESSION NO.	GENOME SIZE (bp)	NO. CDS
Mesorhizobium loti	NC_002678	7036071	6743
Sinorhizobium meliloti	NC_003047	3654135	3359
Bradyrhizobium japonicum	NC_004463	9105828	8317
Rhodopseudomonas palustris	NC_005296	5459213	4813
Bartonella quintana	NC_005955	1581384	1142
Bartonella henselae	NC_005956	1931047	1488
Rickettsia typhi	NC_006142	1111496	837
Beijerinckia indica	NC_010581	4170153	3569
Brucella melitensis	NC_012441	2125701	2063
Rhizobium leguminosarum	NC_012850	4767043	4565
Methylobacterium extorquens	NC_012988	5943768	5594
Rhodomicrobium vannielii	NC_014664	4014469	a3565

Table 4.1 Table 1. Whole-genome sequences used in this study

4.2 Methods

4.2.1 Data preparation

Twelve species were selected from the phylogenetic tree for alphaproteobacteria given in (Gupta, 2005), eleven randomly from one order and the twelfth arbitrarily from another order, and their 16S rRNA genes were downloaded from http://rdp.cme.msu.edu/. The complete genome sequences for these twelve species were downloaded from http://www.ncbi.nlm.nih.gov/ (Table 4.1). As there are many strains for each species—e.g., there were five different strains of *Brucella melitensis*—we randomly selected one to serve as the species representative.

4.2.2 Phylogenetic tree construction using the 16S rRNA gene

Two different methods were applied to build the 16S rRNA tree for the twelve species. One was the unweighted, neighbor-joining bootstrapped consensus dendrogram tree (including bootstrap values) and the other was the Weighbor-weighted neighbor-joining tree constructed using the tree builder tool of the Ribosomal Database Project (RDP) (Cole et al., 2009). For the unweighted method, the neighbor joining tree was obtained using MEGA5 with 500 bootstrapping iterations based on the results of multiple alignment from ClustalW with default settings, which allows gaps. The Weighbor-weighted consensus tree was implemented in the manner described in (Gupta, 2005). Weighbor is a weighted version of neighbor joining that assigns much less weight to longer distances in the distance matrix. The weights are based on variances and covariances expected in a simple Jukes-Cantor model (Bruno et al., 2000).

4.2.3 Phylogenetic tree construction using the whole genome

More than 46k proteins were extracted from the twelve genomes, and these proteins were clustered into homologous groups using *pClust*. Details for using the *pClust* program to generate a tree are given in Zhou et al. (2012). Briefly, *pClust* uses the Smith-Waterman algorithm to perform pairwise comparison on a subset of the total number of protein sequences used as input—in our case the 46k+ genome proteins—obtained after filtering has occurred. The filtering step removes sequences that are shorter than the window size (one of the configuration parameters) and sequence pairs that do not share at least one exact match of length greater than or equal to the cut-off (another of the configuration parameters that contributes most of the filtering), and the strength of filtering is determined by the two parameter settings in the configuration file. The default settings were used except for ExactMatchLen which was set to 4 rather than the default value of 7. The smaller value provides less stringent filtering so that more proteins are compared. A total of 6,325 homologous protein groups (defined as having at least two proteins) were found by pClust. A binary matrix 12×6325 in size was formed with a 1 or 0 indicating presence or absence, respectively, of a given genome protein in each group. This binary matrix was used to construct the tree using two different distance metrics, the Jaccard distance metric, which is used for binary matrices, and the Euclidean distance metric which is a standard distance metric, and neighbor joining was used to obtain the final trees.

4.3 Results and discussion

Following the procedures described above, phylogenetic trees for twelve species of alphaproteobacteria were constructed from 16S rRNA and complete genome sequences. The unweighted and Weighbor-weighted neighbor-joining bootstrapped trees are shown in Figs. 4.1 and 4.2, respectively. The lower parts of the two trees are very similar, but there is a slight difference in the upper part for *Sinorhizobium meliloti*. The whole-genome results for Euclidean distance and Jaccard distance are shown in Figs. 4.3 and 4.4, respectively. These latter two trees have very similar topologies to each other. A comparison of the 16S rRNA trees with the large tree provided in Gupta (2005) shows that the Weighbor-weighted tree corresponds more closely with it. The differences with the 16S tree shown in (Gupta, 2005) are likely due to a difference in the genes that were used, but this is difficult to evaluate because Gupta (2005) does not identify the genes that were analyzed. Eleven of the twelve species studied are in the order Rhisobiales; the twelfth is in the order Rickettsiales. The whole-genome trees recapitulate most of the Rhisobiales topology, clustering the soil-borne species

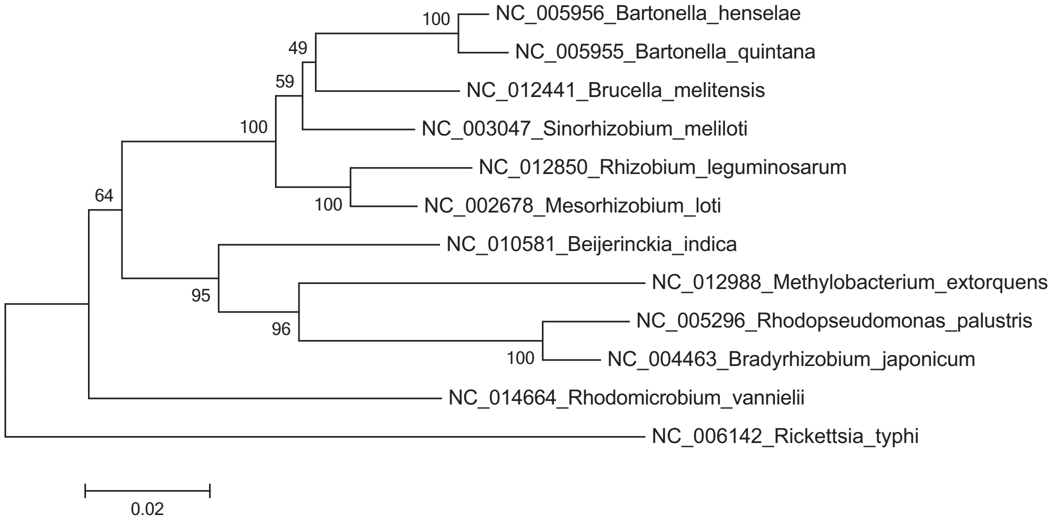


Figure 4.1 Unweighted, neighbor-joining bootstrapped tree using 16S rRNA gene.

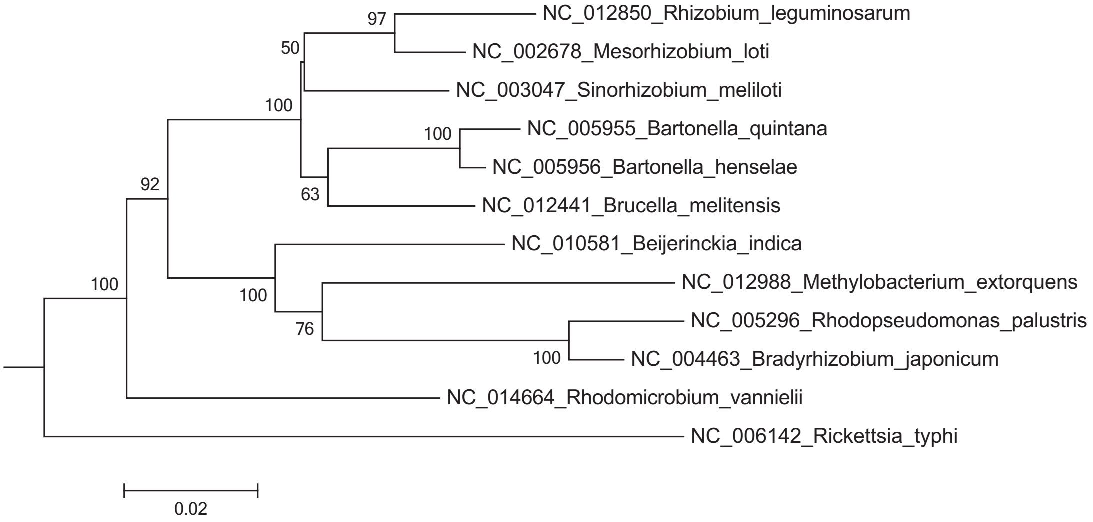


Figure 4.2 Weighbor-weighted neighbor-joining bootstrapped tree using 16S rRNA gene.

of the families Brucellaceae, Rhizobiaceae, and Phyllobacteriaceae separately from the other soil-borne Rhisobiales. However, the one interesting difference is that the Rickettsiales species *Rickettsia typhi* is clustered together with *Bartonella quintana*, *Bartonella henselae* and also with *Brucella melitensis* rather than forming an outlying

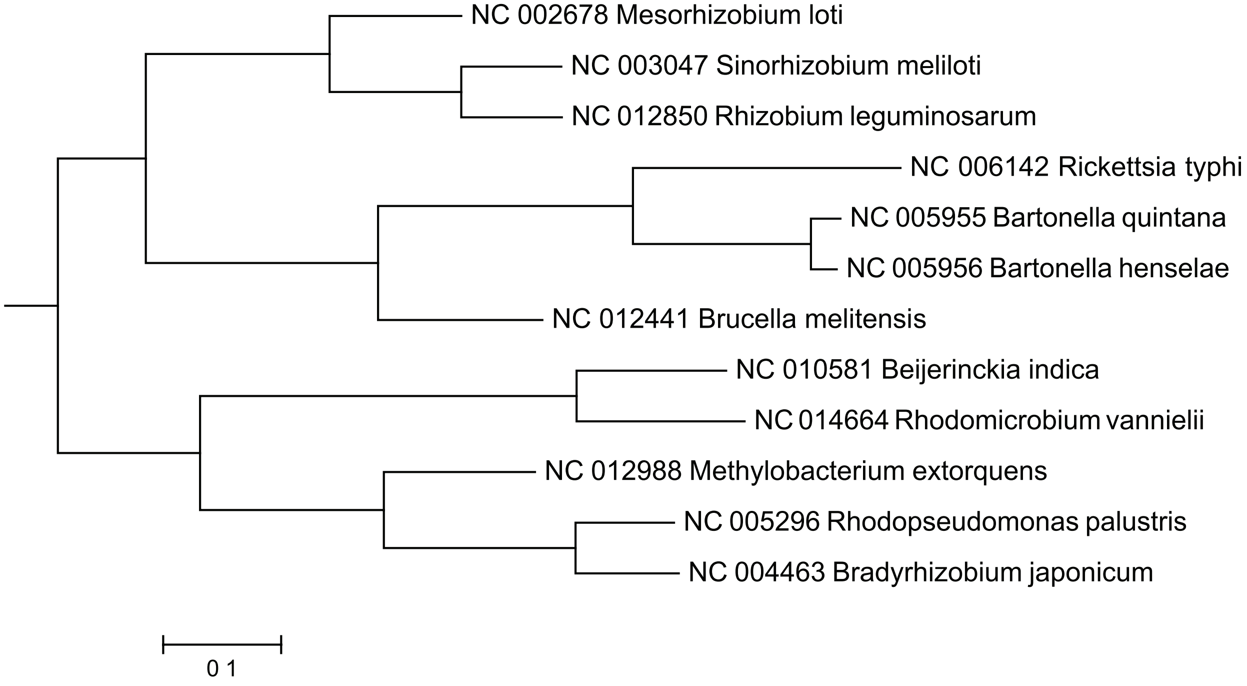


Figure 4.3 Euclidean distance treesusing whole genomes.

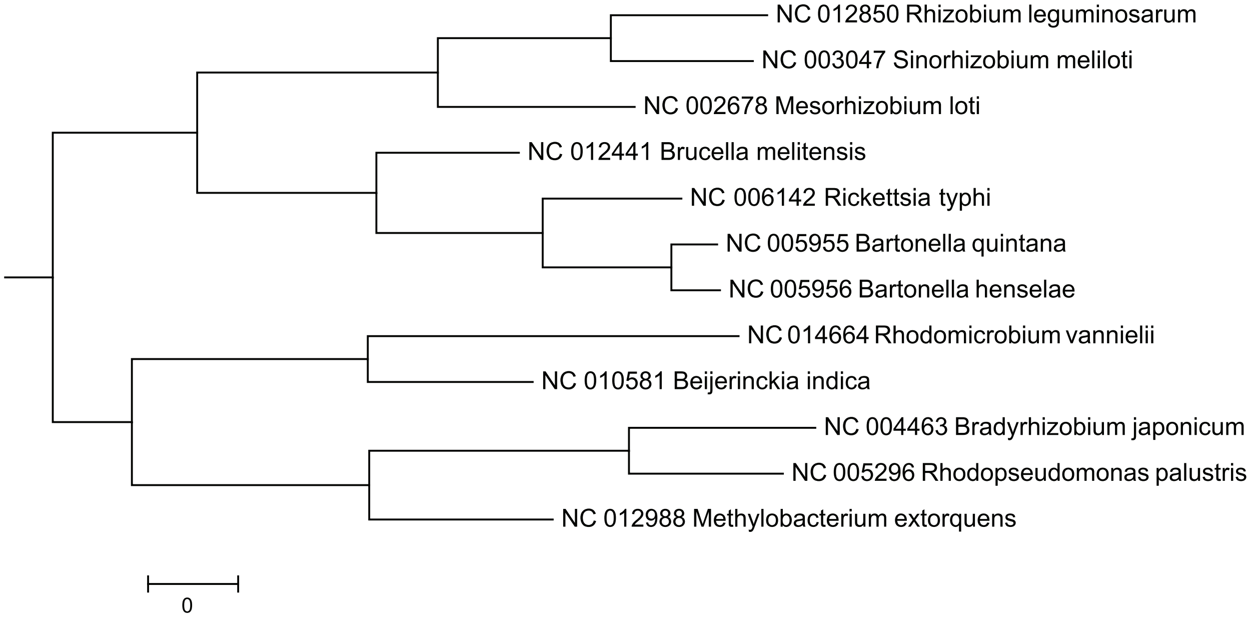


Figure 4.4 Jaccard distance trees using whole genomes.

singlet cluster as it does in the 16S rRNA trees. What makes this interesting is that the four aforementioned species have something in common, which is that they are all pathogens causing, respectively, murine typhus, trench fever, cat scratch disease, and Brucellosis.

It has been challenging to determine the interrelationships among the different alphaproteobacteria families on the basis of the 16S rRNA gene (Ludwig and Klenk, 2001; Kersters et al., 2003). The results discussed above indicate that the use of whole-genome sequences has the potential to illuminate these interrelationships. Clearly, species from two different families share in common the phenotypic trait of animal pathogenicity, and thus the grouping shown in Figs. 4.3 and 4.4 represents a more robust functional grouping of these species.

4.4 Conclusions

It is believed that complete genome sequences will help to clarify phylogenetic relationships among organisms, but until recently no satisfying approach has been proposed to efficiently use these data. While it has not been clear how to benefit from the availability of whole-genome sequences for taxonomic purposes (Daubin et al., 2001), the results shown in this paper give some insight into how this might be accomplished.

The results of the present study show that trees based on whole-genome sequences do not entirely recapitulate the results obtained using 16S rRNA. Instead they reflect a phenotypic relationship between species from two different families and may be the key to giving us a better picture of bacterial phylogeny and taxonomy.

Chapter 5

Conclusions

5.1 Summary

In chapter 2, we described a computational, virtual hybridization method to classify 527 Gram-negative (GN) bacterial plasmids by means of their genomes. This computational approach is analogous to hybridization of mixed-genome microarrays (MGM) (Wan et al., 2007) except that it uses proteins rather than DNA for both 'probes' and 'targets.' While it is not possible to prove that the genetic relationships among the 527 GN bacterial plasmids obtained in this study are correct, replication of identical results produced in a separate study for a small group of IncA/C plasmids provides evidence that the approach can correctly predict genetic relationships. In addition, results obtained for clusters of *Borrelia* plasmids are consistent with the expected exclusivity for plasmids from this genus. Based on our results, one can construct additional hypotheses about both inter- and intra-genus transmission of plasmids.

In chapter 3, we developed a computational method to construct a tree for the 527 GN bacterial plasmids using the software program *pClust* to detect homologous protein clusters. Most importantly the approach permits the insertion of a new plasmid into the existing tree. This is much more efficient than constructing an entirely new tree which is computationally costly (Wu et al., 2012). BLASTp is used to create a profile for a new taxon, and correlation filters are used to determine whether it belongs in the tree and, if so, where is should be located. To utilize the classification algorithm for new plasmids, we focused on 12 pAKD plasmids isolated from Norwegian soil that encode mercury resistance (Sen et al., 2011). Our results agreed with a phylogenetic tree constructed using multiple alignment of the relaxase gene *traI* presented in (Sen et al., 2011). While the method was developed to examine the genetic relationships among 527 Gram-negative bacterial plasmids and to allow insertion of a new plasmid into a tree, it can be used for any genome.

In chapter 4, classification of a group of alphaproteobacteria was discussed. Rather than using the standard approach for creating bacterial phylogenies using 16S rRNA, a method was presented that uses bacterial genomes. We examined 12 species from 11 different genera from the alphaproteobacteria class and constructed trees using both 16S rRNA and our whole-genome approach as described in chapter 3. Comparison of the trees shows that the whole-genome approach better reflects phenotypic traits with all pathogens clustered in one group as opposed to the 16S-rRNA tree where three of the pathogens are grouped, but one is classified as an outlier.

5.2 Future Work

To handle mixed and complicated data or knowledge together, more accurate computational schemes need to be developed. Novel methods for genetic relationship studies should not just focus on phylogenetic tree reconstruction but also should consider other possible systematic and comprehensive approaches such as phylogenetic networks or machine learning classification systems for homolog protein family detection.

In chapter 2, we constructed a tree for 527 GN plasmids using both a consensus algorithm and an average distance algorithm. However, another possibility is the use of a supertree algorithm. Supertree methods combine multiple phylogenetic trees to produce the overall best "supertree" (Bansal et al., 2010). They can be used to combine phylogenetic information from datasets only partially overlapping and from disparate sources (like molecular and morphological data). Another line of future work might focus on some specific coding protein sequences extracted from the completed genomes to identify unique groups. Meanwhile, different distance metrics also have different influences on tree construction. For example, for binary data the Jaccard distance metric is considered to be better than Euclidean distance, but for most of cases we are not sure how to weight the data. Future work might include a comparative study of distance metrics.

In chapter 3, we used a correlation filter method to classify a new plasmid within an existing tree. We also can try machine learning algorithms to classify new plasmids using the existing protein clusters. The advantage of using a machine learning method is that the accuracy is measurable using the receiver operating characteristic. This would give us greater confidence in our results. BlASTp is a local alignment method that is fast but not very accurate. More parallel alignment algorithms can be investigated. Also, the ability to compare trees for large-scale data sets is nonexistent. Currently, we use a visible check to determine whether trees are similar to each other. However, in order to do this, we first must determine what we mean by similarity. If we could develop a method to show a specific similarity value for large-scale trees, it would be a milestone.

The increasing availability of complete genome sequences offers a new pathway for understanding the genetic relationships of microbes, a pathway that could lead to great insight into both their previous and future evolution. However, this can only occur with collaborations between experts in both microbiology and computational methods.

Bibliography

[1] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., et al., (1997). Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs, Nucleic Acids Research, 25:3389-3402.

[2] Baker, S., Hardy, J., Sanderson, K.E., Quail, M., Goodhead, I., et al., (2007). A novel linear plasmid mediates flagellar variation in Salmonella Typhi, PLoS Pathogens, 3:e59.

[3] Bansal, A.K., and Meyer, T.E., (2002). Evolutionary analysis by whole-genome comparisons, J Bacteriol, 184(8):2260-2272.

[4] Bansal, M., Burleigh, J., Eulenstein, O., and Fernandez-Baca, D., (2010). Robinson-Foulds supertrees, Algorithms for molecular biology, 5:18.

[5] Bapteste, E., Boucher, Y., Leigh, J., and Doolittle, W.F., (2004). Phylogenetic reconstruction and lateral gene transfer, Trends in Microbiology, 12:406-411.

[6] Bateman, A., Coin, L., Durbin, R., Finn, R.D., et al., (2004). The Pfam Protein Families Database, Nucleic Acids Research, 32:138-141.

[7] Boucher Y., Douady, C.J., Papke, R.T., Walsh, D.A., Boudreau, M.E., Nesbo, C.L., Case, R.J., Doolittle, W.F., (2003). Lateral gene transfer and the origins of prokaryotic groups. Annu Rev Genet, 37:283-328.

[8] Boyd, E.F., Hill, C.W., Rich, S.M., and Hartl D.L., (1996). Mosaic structure of plasmids from natural populations of Escherichia coli., Genetics, 142:1091-1100.

[9] Brilli, M., Mengoni, A., Mengoni, M., et al., (2008). Analysis of Plasmid Genes by Phylogenetic Profiling and Visualization of Homology Relationships Using Blast2Network, BMC Bioinformatics, 9:551.

[10] Broschat S.L., Call, D.R., Davis, M.A., Meng, D., et al., (2010). Improved Identification of Epidemiologically Related Strains of Salmonella enterica by use of a Fusion Algorithm Based on Pulsed-Field Gel Electrophoresis and Multiple-Locus Variable-Number Tandem-Repeat Analysis, Journal of Clinical Microbiology, 48(11):4072-4082.

[11] Brown, J.R., Douady, C.J., Italia, M.J., Marshall, W.E., and Stanhope, M.J., (2001). Universal trees based on large combined protein sequence data sets, Nat Genet, 28:281-85.

[12] Bruno, W.J., Socci, N.D., and Halpern A.L., (2000). Weighted Neighbor Joining: A Likelihood-Based Approach to Distance-Based Phylogeny Reconstruction, Mol Biol Evol, 17(1):189-197.

[13] Buda, A., and Jarynowski, A., (2010). Life-Time of Correlations and Its Applications, Wydawnictwo Niezalezne, 1:5-21

[14] Call, D.R., Kang, M.S., Daniels, J., Besser, T.E., (2006). Assessing genetic diversity in plasmids from Escherichia coli and Salmonella enterica using a mixed-plasmid microarray, Journal of Applied Microbiology, 100:15-28.

[15] Call, D.R., Singer, R.S., Meng, D., Broschat, S.L., and Orfe, L.H., et al., (2010). blaCMY-2-positive Inc A/C plasmids from Escherichia coli and Salmonella enter-

ica are a distinct component of a larger lineage of plasmids, Antimicrobial Agents and Chemotherapy, 54(2):590-596.

[16] Campbell, A., Mrazek, J., Karlin, S., (1999). Genome signature comparisons among prokaryote, plasmid and mitochondrial DNA, Proc Natl Acad Sci, 96:9184-9189.

[17] Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., Thompson, J.D., (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res, 31(13):3497-3500.

[18] Christopher, M.T., (2000). Paradigms of plasmid organization, Molecular Microbiology, 37(3):485-491.

[19] Coenye, T., and Vandamme, P., (2003). Extracting phylogenetic information from whole-genome sequencing projects: the lactic acid bacteria as a test case, Microbiology, 149(3):507-517.

[20] Cohen, J., (1988), *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.

[21] Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., et al., (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis, Nucleic Acids Res, 37 (Database issue).

[22] Couturier, M., Bex, F., Bergquist, P.L. and Maas, W.K., (1988). Identification and classification of bacterial plasmids, Microbiology Reviews, 52:375-395.

[23] Daubin, V., Gouy, M., and Perriere, G., (2001). Bacterial molecular phylogeny using supertree approach, Genome Inform, 12:155-164.

[24] Daubin, V., Gouy, M., and Perriere, G., (2002). A Phylogenomic Approach to Bacterial Phylogeny: Evidence of a Core of Genes Sharing a Common History, Genome Res, 12:1080-1090.

[25] Davis, M.A., Hancock, D.D., Besser, T.E., and Call, D. R., (2003). Evaluation of pulsed-field gel electrophoresis as a tool for determining the degree of genetic relatedness between strains of Escherichia coli O157:H7, J Clin Microbiol, 41(5):1843-1849.

[26] Dennis, J.J., (2005). The Evolution of IncP Catabolic Plasmids, Current Opinions in Biotechnology, 16(3):291-298.

[27] Efron, B., Halloran, E., and Holmes, S., (1996). Bootstrap confidence levels for phylogenetic trees, Proc Natl Acad Sci, 93(23).

[28] Eggers, C.H., Samuels, D.S., (1999). Molecular evidence for a new bacteriophage of Borrelia burgdorferi, Journal of Bacteriology, 181(23):7308-7313.

[29] Eisen, J.A., (1995). The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species, J Mol Evol, 41(6):1105-123.

[30] Eisen, J.A., (2000). Assessing evolutionary relationships among microbes from whole-genome analysis, Curr Opin Microbiol, 3(5):475-480.

[31] Felsenstein, J., (1989). PHYLIP-Phylogeny Inference Package (Version 3.2), Cladistics, 5:154-166.

[32] Feng, D.F., Cho, G., Doolittle, R.F., (1997). Determining divergence times with a protein clock: update and reevaluation, Proc Natl Acad Sci, 94(1):3028-3033.

[33] Fitz-Gibbon, S., and House, C.H., (1999). Whole genome-based phylogenetic analysis of free-living microorganisms, Nucleic Acids Res, 27(4): 218-222.

[34] Fricke, W.F., Welch, T.J., McDermott, P.F., Mammel, M.K., LeClerc, J.E., et al., (2009). Comparative genomics of the Inc AC multidrug resistance plasmid family, Journal of Bacteriology, 191:4750-4757.

[35] Galtier, N., and Guoy, M., (1998.) Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis, Mol Biol Evol, 15:871-879.

[36] Garrity, G.M., Holt, J.G., (2001). Bergey's Manual of Systematic Bacteriology, New York: Springer. 2nd ed, 1:119-141.

[37] Gordon, A., (1986). Consensus supertrees: the synthesis of rooted trees containing overlapping sets of labeled leaves, Journal of Classification, 3:335-348.

[38] Gupta, R.S., (2005). Protein signatures distinctive of Alpha Proteobacteria and its Subgroups and a model for Alphaproteobacterial evolution, Crit Rev Microbiol, 31:101-135.

[39] Halary, S., Leigh, J.W., Cheaib, B., Lopez P., et al., (2009). Network analyses structure genetic diversity in independent genetic worlds, Proc Natl Acad Sci, 107(1):127-132.

[40] House, C.H. and Fitz-Gibbon, S.T., (2002). Using homolog groups to create a whole-genomic tree of free-living organisms: an update, J Mol Evol, 54:539-547.

[41] Huang, J., and Gogarten, J.P., (2006). Ancient Horizontal Gene Transfer Can Benefit Phylogenetic Reconstruction, Trends in Genetics, 22(7):361-366.

[42] Karlin, S., Burge, C., (1995). Dinucleotide relative abundance extremes: a genomic signature, Trends Genet, 11:283-290.

[43] Karlin, S., (2001). Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes, Trends in Microbiology, 9(7):335-343.

[44] Kersters K., Devos, P., Gillis, M., Vandamme, P., and Stackebrandt, E., (2003). The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community, 3rd ed. Dworkin M., Springer-Verlag, New York.

[45] Lescot, M., Audic, S., Robert, C., Nguyen, T.T., Blanc, G., et al., (2008). The genome of Borrelia recurrentis, the agent of deadly louse-borne relapsing fever, is a degraded subset of thick-borne Borrelia duttonii, PLos Genetics, 4:e1000185.

[46] Lin F.P.-Y., Lan R., Sintchenko V., Gilbert G.L., Kong F., et al., (2011). Computational Bacterial Genome-Wide Analysis of Phylogenetic Profiles Reveals Potential Virulence Genes of Streptococcus agalactiae, PLoS ONE 6(4):e17964.

[47] Ludwig, W., and Klenk, H.P., (2001). Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematics. In Bergey's Manual of Systematic Bacteriology, 2nd ed, pp. 49-50. Edited by D.R. Boone, R.W. Castenholz, and G.M. Garrity. New York: Springer.

[48] Margush, T., and Mcmorris, F.R., (1981). Consensus n-trees, Bulletin of Mathematical Biology, 43:239-244.

[49] Marshall, C.R., (1997). Comput Sci Statist, 29:218-26.

[50] Norberg, P., Bergstrom, M., Jethava, V., Dubhashi, D., and Hermansson, M., (2011). The IncP-1 Plasmid Backbone Adapts to Different Host Bacterial Species and Evolves through Homologous Recombination, Nat Commun, 2.

[51] Ochman, H., Lawrence, J.G., and Groisman, E.A., (2000). Lateral Gene Transfer and the Nature of Bacterial Innovation, Nature, 405:299-304.

[52] Passel, M., Bart, A., Luyf, A., et al., (2006). Compositional discordance between prokaryotic plasmids and host chromosomes, BMC Genomics, 7:26.

[53] Popa, O., Hazkani-Covo, E., Landan, G., Martin, W., Dagan, T., (2011). Directed networks re-veal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes, Genome Res, 21:599-609.

[54] Purser, J.E., Norris, S.J., (2000). Correlation between plasmid content and infectivity in Borrelia burgdorferi, Proc Natl Acad Sci, 97:13865-13870.

[55] Rodgers, J.L., and Nicewander, W.A., (1988). Thirteen Ways to Look at the Correlation Coefficient, The American Statistician, 42(1):59-66.

[56] Roux, V., Rydkina, E., Eremeeva, M., and Raoult, D., (1997). Citrate synthase gene comparison, a new tool for phylogenetic analysis, and its application for the Rickettsiae, Int J Syst Bacteriol 47:252-261.

[57] Saitou N., and Nei M., (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees, Molecular Biology and Evolution, 4(4):406-425.

[58] Sasson, O., Vaaknin, A., Fleischer, H., Portugaly, E., et al., (2003). ProtoNet: Hierarchical Classification of the Protein Space, Nucleic Acids Research, 31(1):348-352.

[59] Sen, D., Van der Auwera, G.A., Rogers, L.M., et.al., (2011). Broad-Host-Range Plasmids from Agricultural Soils Have IncP-1 Backbones with Diverse Accessory Genes, Appl Environ Microbiol, 77(22):7975-7983.

[60] Shpaer, E.G., Robinson, M., Yee, D., Candlin, J.D., et al., (1996). Sensitivity and Selectivity in Protein Similarity Searches: A Comparison of Smith-Waterman in Hardware to BLAST and FASTA, Genomics, 38:179-191.

[61] Smalla, K., Osborn, A., and Wellington, E., (2000). Isolation and characterization of plasmids from bacteria, Harwood Academic Publishers. 424-428.

[62] Snel, B., Bork, P., and Huynen, M.A., (1999). Genome phylogeny based on gene content. Nat Genet, 21(1):108-110.

[63] Sober, E., (1983). Parsimony in Systematics: Philosophical Issues, Annual Review of Ecology and Systematics 14:335-357.

[64] Sokal, R., and Michener, C., (1958). A statistical method for evaluating systematic relationships, University of Kansas Science Bulletin, 38:1409-1438.

[65] Stigler, M.S., (1989). Francis Galton's Account of the Invention of Correlation, Statistical Science, 4(2):73-79.

[66] Sullivan, J., and Joyce, P., (2005). Model Selection in Phylogenetics, Annual Review of Ecology Evolution and Systematics, 36(1):445.

[67] Suzuki, H., Sota, M., Brown, C.J., Top, E.M., (2008). Using Mahalanobis distance to compare genomic signatures between bacterial plasmids and chromosomes, Nucleic Acids Research, 36(22).

[68] Swofford, D.L., (2002). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4. Sinauer Associates, Sunderland, Massachusetts.

[69] Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood,

Evolutionary Distance, and Maximum Parsimony Methods, Molecular Biology and Evolution, 28:2731-2739.

[70] Tekaia, F., Lazcano, A., and Dujon, B., (1999). The genomic tree as revealed from whole proteome comparisons, Genome Res, 9:550-557.

[71] Thomas, C.M., (2000). Paradigms of Plasmid Organization, Mol Microbiol, 37:485-491.

[72] Thomas, C.M., and Nielsen, K.M., (2005). Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria, N Rev Microbiol, 3:711-721.

[73] Wan, Y., Broschat, S.L., Call, D.R., (2007). Validation of mixed-genome microarrays as a method for genetic discrimination, Applied and Enviromental Microbiology, 73(5):1425-1432.

[74] Welch, T.J., Evenhuis, J., White, D.G., McDermott, P.F., Harbottle, H., et al., 2009. Inc AC plasmid-mediated florfenicol resistance3 in the catfish pathogen Edwardsiella ictaluri, Antimicrobial Agents and Chemotherapy, 53:845-846.

[75] Wheelis, M.L., Kandler, O., and Woese, C.R., (1992). On the nature of global classification, Proc Natl Acad Sci, 89:2930-2934.

[76] Wiens, J.J., (2001). Character analysis in morphological phylogenetics: problems and solutions, Syst Biol, 50(5):689-99.

[77] Wiley, O., Brooks, D.R., Siegel-Causey, D., Funk, V.A., (1991). The Compleat Cladist: A Primer of Phylogenetic Procedures.

[78] Woese, C.R., Kandler, O., and Wheelis, M.L., (1990). Proc Natl Acad Sci, 87:4576-579.

[79] Wolf, Y.I., Rogozin, I.B., Grishin, N.V., Tatusov, R.L., and Koonin, E.V., (2001). Genome trees constructed using five different approaches suggest new major bacterial clades, BMC Evol Biol 1(8).

[80] Wolska, K., (2003). Horizontal DNA transfer between bacteria in the environment, Acta Microbiol Pol 52(3):233-43. PMID 14743976.

[81] Wu, C., Kalyanaraman, A., Cannon, W.R., (2012). pGraph: Efficient parallel construction of large-scale protein sequence homology graphs. IEEE Transactions on Parallel and Distributed Systems (TPDS).

[82] Yang, Z., (2007). PAML 4: A program package for phylogenetic analysis by maximum likelihood, Molecular Biology and Evolution, 24:1586-1591.

[83] Yeh, R.F., Lim, L.P. and Burge, C.B., (2001). Computational Inference of Homologous Gene Structures in the Human Genome, Genome Res, 11:803-816.

[84] Zhang, H., Raji, A., Theisen, M., Hansen, P.R., Marconi, R.T., (2005). bdrF2 of Lyme disease Spirochetes is coexpressed with a series of cytop lasmic proteins and is produced specifically during early infection, Journal of Bacteriology, 187:175-184.

[85] Zhou, Y., Call, D.R., and Broschat, S.L., (2012). “Genetic Relationships among 527 Gram-Negative Bacterial Plasmids,” to appear in Plasmid.

Appendices A-F are supplementary figures for chapter 2.

Appendix A

Fig. A.1. List of 527 Gram-negative plasmids used in the study.

organism	name	accession	length	proteins	RNAs	genes
Acaryochloris marina MBIC11017	pREB8	NC_009933	120693 nt	117	0	120
Acaryochloris marina MBIC11017	pREB7	NC_009932	155110 nt	173	0	173
Acaryochloris marina MBIC11017	pREB6	NC_009931	172728 nt	188	0	192
Acaryochloris marina MBIC11017	pREB5	NC_009930	177162 nt	222	0	223
Acaryochloris marina MBIC11017	pREB4	NC_009929	226680 nt	273	0	278
Acaryochloris marina MBIC11017	pREB3	NC_009928	273121 nt	362	0	380
Acaryochloris marina MBIC11017	pREB1	NC_009926	374161 nt	382	0	390
Acaryochloris marina MBIC11017	pREB2	NC_009927	356087 nt	408	0	415
Acetobacter pasteurianus IFO 3283-01	pAPA01-030	NC_013212	49961 nt	63	0	63
Acetobacter pasteurianus IFO 3283-01	pAPA01-020	NC_013211	182940 nt	174	0	174
Acetobacter pasteurianus IFO 3283-01	pAPA01-011	NC_013210	191799 nt	178	0	178
Achromobacter denitrificans	pEST4011	NC_005793	76958 nt	61	0	62
Achromobacter xylosoxidans A8	pA81	NC_006830	98192 nt	100	0	103
Acidiphilium cryptum JF-5	pACRY03	NC_009469	88953 nt	77	0	81
Acidiphilium cryptum JF-5	pACRY01	NC_009467	203589 nt	171	0	187
Acidiphilium cryptum JF-5	pACRY02	NC_009468	187422 nt	169	0	179
Acidovorax sp. JS42	pAOVO02	NC_008766	63609 nt	64	0	73
Acidovorax sp. JS42	pAOVO01	NC_008765	72689 nt	84	0	85
Acinetobacter baumannii ACICU	pACICU2	NC_010606	64366 nt	64	0	64
Acinetobacter baumannii AYE	p3ABAYE	NC_010404	94413 nt	82	0	107
Aeromonas hydrophila	pRA1	NC_012885	143963 nt	158	0	158
Aeromonas punctata	pFBAOT6	NC_006143	84749 nt	92	0	95
Aeromonas salmonicida subsp. salmonicida A449	5	NC_009350	155098 nt	154	0	164
Aeromonas salmonicida subsp. salmonicida A449	4	NC_009349	166749 nt	173	0	178
Agrobacterium radiobacter K84	pAtK84b	NC_011990	184668 nt	176	0	177
Agrobacterium radiobacter K84	pAtK84c	NC_011987	388169 nt	366	0	373
Agrobacterium rhizogenes	pRi2659	NC_010841	185462 nt	146	0	146
Agrobacterium rhizogenes	pRi1724	NC_002575	217594 nt	173	0	173
Agrobacterium tumefaciens	Ti	NC_002377	194140 nt	157	0	157
Agrobacterium tumefaciens	pTi-SAKURA	NC_002147	206479 nt	195	0	196
Agrobacterium tumefaciens	Ti plasmid pTiBo542	NC_010929	244978 nt	223	0	223
Agrobacterium tumefaciens str. C58	Ti	NC_003065	214233 nt	197	0	199
Agrobacterium tumefaciens str. C58	At	NC_003064	542868 nt	542	0	557
Agrobacterium vitis S4	pAtS4a	NC_011986	78730 nt	82	0	82
Agrobacterium vitis S4	pAtS4b	NC_011991	130435 nt	107	0	109
Agrobacterium vitis S4	pAtS4c	NC_011984	211620 nt	178	0	197
Agrobacterium vitis S4	pTiS4	NC_011982	258824 nt	205	1	238
Agrobacterium vitis S4	pAtS4e	NC_011981	631775 nt	529	0	543
Aliivibrio salmonicida LFI1238	pVSAL840	NC_011311	83540 nt	72	0	72
Allochromatium vinosum DSM 180	pALVIN01	NC_013852	102242 nt	119	0	127
Anabaena variabilis ATCC 29413	C	NC_007412	300758 nt	243	1	248
Anabaena variabilis ATCC 29413	A	NC_007410	366354 nt	344	0	350
Aromatoleum aromaticum EbN1	2	NC_006824	223670 nt	194	0	196
Aromatoleum aromaticum EbN1	1	NC_006823	207355 nt	272	0	274
Azospirillum sp. B510	pAB510f	NC_013860	261596 nt	187	0	187
Azospirillum sp. B510	pAB510e	NC_013859	537299 nt	415	12	427
Azospirillum sp. B510	pAB510d	NC_013858	628837 nt	519	6	525
Azospirillum sp. B510	pAB510c	NC_013857	681723 nt	533	6	539
Azospirillum sp. B510	pAB510b	NC_013856	723779 nt	631	5	636
Azospirillum sp. B510	pAB510a	NC_013855	1455109 nt	1131	25	1156
Beijerinckia indica subsp. indica ATCC 9039	pBIND01	NC_010580	181736 nt	176	0	199
Borrelia afzelii ACA-1	lp54	NC_011786	60749 nt	76	0	76
Borrelia afzelii PKo	lp32	NC_008567	32368 nt	55	0	55
Borrelia afzelii PKo	lp60-2	NC_008565	59804 nt	59	0	59
Borrelia afzelii PKo	lp60	NC_008564	59958 nt	75	0	75
Borrelia burgdorferi 118a	118a_cp32-7-9	NC_012253	60942 nt	86	0	86
Borrelia burgdorferi 118a	118a_lp54	NC_012241	54027 nt	66	0	66
Borrelia burgdorferi 156a	156a_lp54	NC_011876	53885 nt	66	0	66
Borrelia burgdorferi 297	297_lp54	NC_013128	48220 nt	57	0	58
Borrelia burgdorferi 29805	29805_lp54	NC_012505	53849 nt	69	0	69
Borrelia burgdorferi 64b	64b_lp36	NC_012184	36826 nt	50	0	50
Borrelia burgdorferi 64b	64b_lp38	NC_012182	39007 nt	56	0	56
Borrelia burgdorferi 64b	64b_cp32-3-8	NC_012106	61062 nt	86	0	86
Borrelia burgdorferi 64b	64b_lp54	NC_012199	53263 nt	68	0	68
Borrelia burgdorferi 72a	72a_lp54	NC_012193	53686 nt	69	0	69
Borrelia burgdorferi 94a	94a_lp54	NC_012244	53756 nt	72	0	72
Borrelia burgdorferi B31	lp36	NC_001855	36849 nt	51	0	54
Borrelia burgdorferi B31	lp56	NC_000956	52971 nt	72	0	82
Borrelia burgdorferi B31	lp54	NC_001857	53561 nt	76	0	76
Borrelia burgdorferi Bol26	Bol26_lp54	NC_012504	53655 nt	65	0	65
Borrelia burgdorferi CA-11.2a	CA-11.2A_lp54	NC_012194	54021 nt	71	0	71
Borrelia burgdorferi CA-11.2a	CA-11.2A_lp36-28-4	NC_012202	54623 nt	81	0	81
Borrelia burgdorferi JD1	JD1_lp54	NC_013129	52916 nt	63	0	63
Borrelia burgdorferi N40	N40_lp54	NC_013130	53786 nt	64	0	64
Borrelia burgdorferi WI91-23	WI91-23_lp38	NC_012167	38893 nt	62	0	62
Borrelia burgdorferi WI91-23	WI91-23_lp54	NC_012175	53783 nt	69	0	69
Borrelia burgdorferi ZS7	ZS7_lp54	NC_011784	53615 nt	55	0	66
Borrelia burgdorferi ZS7	ZS7_cp32-3+10	NC_011720	48168 nt	60	0	67
Borrelia duttonii Ly	pl165	NC_011247	163884 nt	134	0	144
Borrelia garinii Far04	Far04_lp36	NC_011867	35446 nt	54	0	54
Borrelia garinii Far04	Far04_lp54	NC_011877	58069 nt	79	0	79
Borrelia garinii Far04	Far04_lp32-10	NC_011869	41906 nt	84	0	84
Borrelia garinii PBi	lp54	NC_006129	55560 nt	74	0	74
Borrelia garinii PBr	PBr_lp25	NC_011856	32543 nt	53	0	53
Borrelia garinii PBr	PBr_lp36	NC_011857	36709 nt	53	0	53
Borrelia garinii PBr	PBr_lp28-4	NC_011860	40280 nt	69	0	69
Borrelia garinii PBr	PBr_cp32-10	NC_011842	34837 nt	75	0	75
Borrelia garinii PBr	PBr_lp54	NC_011859	56869 nt	75	0	75

Borrelia recurrentis A1	pl124	NC_011246	123937 nt	91	0	110
Borrelia sp. SV1	SV1_lp32-12	NC_012247	52543 nt	71	0	71
Borrelia sp. SV1	SV1_lp54	NC_012234	83377 nt	117	0	117
Borrelia spielmanii A14S	A14S_lp54	NC_012164	56055 nt	70	0	70
Borrelia valaisiana VS116	VS116_lp25	NC_012166	32795 nt	62	0	62
Borrelia valaisiana VS116	VS116_lp54	NC_012177	53953 nt	70	0	70
Borrelia valaisiana VS116	VS116_lp28-3	NC_012185	81995 nt	142	0	142
Bradyrhizobium sp. BTAi1	pBBta01	NC_009475	228826 nt	228	0	257
Burkholderia ambifaria MC40-6	pBMC401	NC_010553	301592 nt	274	0	298
Burkholderia cenocepacia HI2424	1	NC_008545	164857 nt	156	0	159
Burkholderia cenocepacia J2315	pBCJ2315	NC_011003	92661 nt	93	1	132
Burkholderia cepacia	pIJB1	NC_013666	99448 nt	68	0	68
Burkholderia glumae BGR1	bglu_1p	NC_012723	133591 nt	102	0	144
Burkholderia glumae BGR1	bglu_2p	NC_012718	141792 nt	97	0	121
Burkholderia glumae BGR1	bglu_4p	NC_012725	134349 nt	102	1	115
Burkholderia glumae BGR1	bglu_3p	NC_012720	141067 nt	106	1	143
Burkholderia multivorans ATCC 17616	pTGL1	NC_010802	167422 nt	131	0	138
Burkholderia multivorans ATCC 17616	pBMUL01	NC_010070	167422 nt	138	0	139
Burkholderia phymatum STM815	pBPHY02	NC_010627	595108 nt	449	0	575
Burkholderia phymatum STM815	pBPHY01	NC_010625	1904893 nt	1626	1	1704
Burkholderia phytofirmans PsJN	pBPHYT01	NC_010679	121122 nt	167	0	168
Burkholderia sp. CCGE1002	pBC201	NC_014120	489136 nt	406	0	495
Burkholderia vietnamiensis G4	pBVIE04	NC_009228	107231 nt	107	0	112
Burkholderia vietnamiensis G4	pBVIE05	NC_009226	88096 nt	111	0	111
Burkholderia vietnamiensis G4	pBVIE03	NC_009229	226679 nt	249	0	251
Burkholderia vietnamiensis G4	pBVIE02	NC_009227	265616 nt	263	0	272
Burkholderia vietnamiensis G4	pBVIE01	NC_009230	397868 nt	403	0	409
Caedibacter taeniospiralis	pKAP298	NC_005915	49112 nt	57	0	57
Campylobacter coli	pCC31	NC_006134	44707 nt	50	0	50
Campylobacter jejuni subsp. jejuni 81-176	pVir	NC_005012	37468 nt	52	0	54
Campylobacter jejuni subsp. jejuni 81-176	pTet	NC_008790	45025 nt	52	0	52
Campylobacter jejuni subsp. jejuni 81-176	pVir	NC_008770	37473 nt	53	0	53
Candidatus Accumulibacter phosphatis clade IIA str.	pAph01	NC_013193	167595 nt	156	0	168
Candidatus Hamiltonella defensa 5AT (Acyrthosipho	pHD5AT	NC_012752	59032 nt	54	0	55
Caulobacter sp. K31	pCAUL02	NC_010333	177878 nt	165	0	170
Caulobacter sp. K31	pCAUL01	NC_010335	233649 nt	212	0	212
Chelativorans sp. BNC1	2	NC_008243	131247 nt	113	0	122
Chelativorans sp. BNC1	1	NC_008242	343931 nt	317	4	343
Citrobacter freundii	pCTX-M3	NC_004464	89468 nt	105	0	105
Citrobacter rodentium ICC168	pCROD2	NC_013718	39265 nt	55	0	56
Citrobacter rodentium ICC168	pCROD1	NC_013717	54449 nt	55	1	60
Comamonas sp. CNB-1	pCNB	NC_010935	91181 nt	92	0	92
Coxiella burnetii Dugway 5J108-111	pQpDG	NC_009726	54179 nt	52	0	59
Coxiella burnetii 'MSU Goat Q177'	QpRS	NC_010258	39281 nt	54	0	57
Cronobacter sakazakii ATCC BAA-894	pESA3	NC_009780	131196 nt	127	0	127
Cronobacter turicensis	pCTU3	NC_013285	53842 nt	74	0	74
Cronobacter turicensis	pCTU1	NC_013283	138339 nt	136	0	136
Cupriavidus metallidurans CH34	pMOL28	NC_006525	171461 nt	112	0	112
Cupriavidus metallidurans CH34	pMOL28	NC_007972	171459 nt	165	0	174
Cupriavidus metallidurans CH34	pMOL30	NC_006466	233755 nt	177	0	177
Cupriavidus metallidurans CH34	pMOL30	NC_007971	233720 nt	200	0	283
Cupriavidus metallidurans CH34	megaplasmid	NC_007974	2580084 nt	2386	14	2483
Cupriavidus taiwanensis	pRALTA	NC_010529	557200 nt	519	0	576
Cyanothece sp. PCC 7424	pP742402	NC_011737	197705 nt	175	0	191
Cyanothece sp. PCC 7424	pP742401	NC_011738	328635 nt	236	0	264
Cyanothece sp. PCC 7425	pP742501	NC_011880	196837 nt	160	0	168
Cyanothece sp. PCC 7425	pP742502	NC_011885	179973 nt	172	0	197
Cyanothece sp. PCC 8802	pP880201	NC_013160	75678 nt	58	0	69
Delftia acidovorans	pUO1	NC_005088	67066 nt	69	0	69
Desulfobacterium autotrophicum HRM2	pHRM2a	NC_012109	68709 nt	76	0	76
Desulfotalea psychrophila LSv54	large	NC_006139	121587 nt	101	0	101
Desulfovibrio magneticus RS-1	pDMC1	NC_012797	58704 nt	65	0	65
Desulfovibrio vulgaris DP4	pDVUL01	NC_008741	198504 nt	150	0	150
Desulfovibrio vulgaris str. Hildenborough	pDV	NC_005863	202301 nt	152	1	153
Dinoroseobacter shibae DFL 12	pDSHI04	NC_009958	86208 nt	68	0	68
Dinoroseobacter shibae DFL 12	pDSHI05	NC_009959	72296 nt	73	0	73
Dinoroseobacter shibae DFL 12	pDSHI03	NC_009957	126304 nt	135	0	136
Dinoroseobacter shibae DFL 12	pDSHI02	NC_009956	152970 nt	137	0	137
Dinoroseobacter shibae DFL 12	pDSHI01	NC_009955	190506 nt	195	0	196
Edwardsiella tarda EIB202	pEIB202	NC_013509	43703 nt	53	0	53
Enterobacter aerogenes	R751	NC_001735	53423 nt	67	0	69
Enterobacter cloacae	pEC-IMP	NC_012555	318782 nt	309	0	315
Enterobacter cloacae	pEC-IMPQ	NC_012556	324503 nt	313	0	320
Enterobacter cloacae subsp. cloacae ATCC 13047	pECL_B	NC_014108	84653 nt	124	0	124
Enterobacter cloacae subsp. cloacae ATCC 13047	pECL_A	NC_014107	199562 nt	274	0	274
Enterobacter sp. 638	pENTE01	NC_009425	157749 nt	125	0	125
Erwinia amylovora	pEL60	NC_005246	60145 nt	68	0	68
Erwinia amylovora ATCC 49946	2	NC_013973	71487 nt	87	0	87
Erwinia billingiae Eb661	pEB102	NC_014304	102323 nt	114	0	115
Erwinia billingiae Eb661	pEB170	NC_014305	169778 nt	220	0	220
Erwinia tasmaniensis Et1/99	pET49	NC_010697	48751 nt	61	0	61
Escherichia coli	pKC394	NC_014231	53207 nt	55	0	57
Escherichia coli	pMUR050	NC_007682	56634 nt	60	0	63
Escherichia coli	pMAS2027	NC_013503	42644 nt	58	0	58
Escherichia coli	pOLA52	NC_010378	51602 nt	68	0	68
Escherichia coli	pRAx	NC_012886	52637 nt	70	0	70
Escherichia coli	pLEW517	NC_009132	63946 nt	71	0	71
Escherichia coli	pB171	NC_002142	68817 nt	80	0	80
Escherichia coli	pEK516	NC_013121	64471 nt	79	0	79

Escherichia coli	pCoo	NC_007635	98396 nt	94	2	131
Escherichia coli	R721	NC_002525	75582 nt	92	0	93
Escherichia coli	pSFO157	NC_009602	121239 nt	96	0	96
Escherichia coli	pEK204	NC_013120	93732 nt	95	1	96
Escherichia coli	pC15-1a	NC_005327	92353 nt	100	0	102
Escherichia coli	pMAR7	NC_010862	101558 nt	100	1	116
Escherichia coli	pO26I	NC_011812	72946 nt	101	0	101
Escherichia coli	pAPEC-O2-R	NC_006671	101375 nt	119	0	119
Escherichia coli	pEC14_114	NC_013175	114222 nt	117	0	117
Escherichia coli	NR1	NC_009133	94289 nt	123	0	123
Escherichia coli	pAPEC-O103-ColBM	NC_011964	124705 nt	126	0	131
Escherichia coli	p1658/97	NC_004998	125491 nt	141	2	143
Escherichia coli	pEK499	NC_013122	117536 nt	141	0	141
Escherichia coli	pVM01	NC_010409	151002 nt	150	0	150
Escherichia coli	pO113	NC_007365	165548 nt	155	0	191
Escherichia coli	pO86A1	NC_008460	120730 nt	163	1	164
Escherichia coli	pAPEC-1	NC_011980	103275 nt	164	0	167
Escherichia coli	peH4H	NC_012690	148105 nt	173	0	173
Escherichia coli	pO26-Vir	NC_012487	168100 nt	195	0	195
Escherichia coli	pAR060302	NC_012692	166530 nt	189	0	189
Escherichia coli	pAPEC-O2-ColV	NC_007675	184501 nt	209	0	210
Escherichia coli 1520	pIP1206	NC_010558	168113 nt	176	0	208
Escherichia coli 53638	p53638_75	NC_010720	75089 nt	93	2	105
Escherichia coli 53638	p53638_226	NC_010719	225683 nt	276	2	331
Escherichia coli 55989	55989p	NC_011752	72482 nt	103	3	106
Escherichia coli APEC O1	pAPEC-O1-ColBM	NC_009837	174241 nt	199	0	199
Escherichia coli APEC O1	pAPEC-O1-R	NC_009838	241387 nt	224	0	224
Escherichia coli E24377A	pETEC_74	NC_009790	74224 nt	69	0	77
Escherichia coli E24377A	pETEC_80	NC_009786	79237 nt	68	0	87
Escherichia coli E24377A	pETEC_73	NC_009788	70609 nt	68	0	76
Escherichia coli ED1a	pECOED	NC_011754	119594 nt	138	1	151
Escherichia coli ETEC 1392/75	p557	NC_014233	55709 nt	52	0	59
Escherichia coli ETEC 1392/75	p746	NC_014234	74575 nt	73	0	90
Escherichia coli ETEC 1392/75	p1081	NC_014232	101857 nt	109	0	133
Escherichia coli ETEC H10407	pEntH10407	NC_013507	67094 nt	99	0	99
Escherichia coli O103:H2 str. 12009	pO103	NC_013354	75546 nt	67	0	90
Escherichia coli O111:H- str. 11128	pO111_3	NC_013366	77690 nt	72	0	90
Escherichia coli O111:H- str. 11128	pO111_2	NC_013370	97897 nt	121	3	128
Escherichia coli O111:H- str. 11128	pO111_1	NC_013365	204604 nt	222	0	233
Escherichia coli O127:H6 str. E2348/69	pMAR2	NC_011603	97978 nt	90	3	115
Escherichia coli O157:H7 EDL933	pO157	NC_007414	92077 nt	99	1	101
Escherichia coli O157:H7 str. EC4115	pEC4115	NC_011351	37452 nt	54	0	55
Escherichia coli O157:H7 str. EC4115	pO157	NC_011350	94644 nt	108	1	120
Escherichia coli O157:H7 str. Sakai	pO157	NC_002128	92721 nt	85	0	85
Escherichia coli O157:H7 str. TW14359	pO157	NC_013010	94601 nt	110	0	110
Escherichia coli O26:H-	pO26-CRL	NC_013728	111481 nt	110	0	123
Escherichia coli O26:H11 str. 11368	pO26_1	NC_013369	85167 nt	65	0	93
Escherichia coli O26:H11 str. 11368	pO26_2	NC_013362	63365 nt	81	0	84
Escherichia coli O55:H7 str. CB9615	pO55	NC_013942	66001 nt	107	0	108
Escherichia coli S88	pECOS88	NC_011747	133853 nt	135	3	146
Escherichia coli SE11	pSE11-3	NC_011416	60555 nt	67	0	67
Escherichia coli SE11	pSE11-2	NC_011413	91158 nt	112	0	112
Escherichia coli SE11	pSE11-1	NC_011419	100021 nt	124	0	124
Escherichia coli SMS-3-5	pSMS35_130	NC_010488	130440 nt	153	0	166
Escherichia coli UMN026	p1ESCUM	NC_011749	122301 nt	142	5	156
Escherichia coli UTI89	pUTI89	NC_007941	114230 nt	145	0	145
Escherichia coli Vir68	pVir68	NC_012944	138362 nt	119	0	119
Escherichia fergusonii ATCC 35469	pEFER	NC_011743	55150 nt	55	1	57
Fluoribacter dumoffii	pLD-TEX-KL	NC_009966	66512 nt	57	0	57
Geobacter lovleyi SZ	pGLOV01	NC_010815	77113 nt	79	0	81
Gluconacetobacter diazotrophicus PAl 5	pGDIPal5I	NC_010124	38818 nt	53	0	53
Gluconobacter oxydans 621H	pGOX1	NC_006672	163186 nt	163	0	163
Haemophilus influenzae	ICEhin1056	NC_011409	59393 nt	62	0	64
Herpetosiphon aurantiacus ATCC 23779	pHAU02	NC_009974	99204 nt	71	0	71
Herpetosiphon aurantiacus ATCC 23779	pHAU01	NC_009973	339639 nt	231	0	233
Hirschia baltica ATCC 49814	pHbal01	NC_012983	84492 nt	68	0	68
Jannaschia sp. CCS1	plasmid1	NC_007801	86072 nt	71	0	71
Klebsiella oxytoca KOX105	pKOX105	NC_014208	54641 nt	70	0	70
Klebsiella pneumoniae	pKP96	NC_011617	67850 nt	64	0	72
Klebsiella pneumoniae	pNL194	NC_014368	79307 nt	83	0	91
Klebsiella pneumoniae	pK245	NC_010886	98264 nt	90	0	90
Klebsiella pneumoniae	pCTXM360	NC_011641	68018 nt	76	0	80
Klebsiella pneumoniae	9	NC_011383	70655 nt	87	0	87
Klebsiella pneumoniae	12	NC_011385	75617 nt	92	0	92
Klebsiella pneumoniae	pKF3-70	NC_013542	70057 nt	95	0	95
Klebsiella pneumoniae	pKF3-94	NC_013950	94219 nt	115	0	115
Klebsiella pneumoniae	pKpQIL	NC_014016	113637 nt	121	0	121
Klebsiella pneumoniae	pKP048	NC_014312	151188 nt	156	0	156
Klebsiella pneumoniae	pKF3-140	NC_013951	147416 nt	196	0	196
Klebsiella pneumoniae	pLVPK	NC_005249	219385 nt	251	0	260
Klebsiella pneumoniae	pK29	NC_010870	269674 nt	310	0	310
Klebsiella pneumoniae 342	pKP91	NC_011281	91096 nt	113	0	121
Klebsiella pneumoniae 342	pKP187	NC_011282	187922 nt	230	0	248
Klebsiella pneumoniae NTUH-K2044	pK2044	NC_006625	224152 nt	270	0	270
Klebsiella pneumoniae subsp. pneumoniae MGH 78	pKPN5	NC_009651	88582 nt	98	0	98
Klebsiella pneumoniae subsp. pneumoniae MGH 78	pKPN4	NC_009650	107576 nt	123	0	124
Klebsiella pneumoniae subsp. pneumoniae MGH 78	pKPN3	NC_009649	175879 nt	178	0	179
Lawsonia intracellularis PHE/MN1-00	3	NC_008014	194553 nt	104	0	104
Legionella pneumophila str. Lens	pLPL	NC_006366	59832 nt	56	0	57

Legionella pneumophila str. Paris	pLPP	NC_006365	131885 nt	139	0	142
Leptospira biflexa serovar Patoc strain 'Patoc 1 (Ame	p74	NC_010846	74117 nt	57	0	57
Leptospira biflexa serovar Patoc strain 'Patoc 1 (Paris	p74	NC_010844	74116 nt	59	0	59
Listonella anguillarum	pJM1	NC_005250	65009 nt	57	2	59
Marinobacter aquaeolei VT8	pMAQU02	NC_008739	213290 nt	201	0	213
Marinobacter aquaeolei VT8	pMAQU01	NC_008738	239623 nt	213	0	217
Mesorhizobium loti MAFF303099	pMLb	NC_002682	208315 nt	209	0	209
Mesorhizobium loti MAFF303099	pMLa	NC_002679	351911 nt	320	0	320
Methylibium petroleiphilum PM1	RPME01	NC_008826	599444 nt	630	27	673
Methylobacterium chloromethanicum CM4	pMCHL01	NC_011758	380207 nt	306	0	351
Methylobacterium extorquens AM1	megaplasmid	NC_012811	1261460 nt	1162	6	1177
Methylobacterium extorquens DM4	p1METDI	NC_012987	141504 nt	105	0	137
Methylobacterium nodulans ORS 2060	pMNOD03	NC_011893	40463 nt	53	0	60
Methylobacterium nodulans ORS 2060	pMNOD01	NC_011892	487734 nt	402	2	506
Methylobacterium nodulans ORS 2060	pMNOD02	NC_011887	458070 nt	409	0	482
Methylobacterium radiotolerans JCM 2831	pMRAD02	NC_010509	47003 nt	52	0	52
Methylobacterium radiotolerans JCM 2831	pMRAD01	NC_010510	586164 nt	507	7	536
Methylobacterium sp. 4-46	pM44601	NC_010373	57951 nt	62	0	77
Methylovorus sp. SIP3-4	pMsip01	NC_012970	76680 nt	79	0	79
Microscilla sp. PRE1	pSD15	NC_002806	101648 nt	64	0	64
Nitrobacter hamburgensis X14	3	NC_007961	121408 nt	111	0	135
Nitrobacter hamburgensis X14	2	NC_007960	188318 nt	172	0	185
Nitrobacter hamburgensis X14	1	NC_007959	294829 nt	239	0	314
Nitrosococcus halophilus Nc4	pNHAL01	NC_013958	65833 nt	68	0	72
Nitrosomonas eutropha C91	Plasmid2	NC_008342	55635 nt	52	0	53
Nitrosomonas eutropha C91	Plasmid1	NC_008341	65132 nt	55	0	56
'Nostoc azollae' 0708	pAzo01	NC_014249	109570 nt	51	0	114
Nostoc punctiforme PCC 73102	pNPUN03	NC_010630	123028 nt	73	0	83
Nostoc punctiforme PCC 73102	pNPUN02	NC_010632	254918 nt	197	0	223
Nostoc punctiforme PCC 73102	pNPUN01	NC_010631	354564 nt	260	0	284
Nostoc sp. PCC 7120	pCC7120delta	NC_003273	55414 nt	66	19	85
Nostoc sp. PCC 7120	pCC7120gamma	NC_003267	101965 nt	90	0	90
Nostoc sp. PCC 7120	pCC7120beta	NC_003240	186614 nt	186	0	186
Nostoc sp. PCC 7120	pCC7120alpha	NC_003276	408101 nt	386	0	386
Novosphingobium aromaticivorans	pNL1	NC_002033	184457 nt	186	0	186
Novosphingobium aromaticivorans DSM 12444	pNL1	NC_009426	184462 nt	182	0	187
Novosphingobium aromaticivorans DSM 12444	pNL2	NC_009427	487268 nt	431	0	432
Ochrobactrum anthropi ATCC 49188	pOANT03	NC_009671	93589 nt	86	0	91
Ochrobactrum anthropi ATCC 49188	pOANT02	NC_009670	101491 nt	91	0	96
Ochrobactrum anthropi ATCC 49188	pOANT01	NC_009669	170351 nt	161	0	180
Oligotropha carboxidovorans OM5	pHCG3	NC_005873	133058 nt	125	0	125
Pantoea vagans C9-1	pPag3	NC_014258	529676 nt	535	0	535
Paracoccus denitrificans PD1222	1	NC_008688	653815 nt	616	0	620
Pelobacter propionicus DSM 2379	pPRO1	NC_008607	202397 nt	195	0	195
Persephonella marina EX-H1	unnamed	NC_012439	53682 nt	70	0	70
Phenylobacterium zucineum HLK1	unnamed	NC_011143	382976 nt	325	0	327
Photobacterium damselae subsp. piscicida	pP9014	NC_012919	55851 nt	61	0	61
Photobacterium damselae subsp. piscicida	pP91278	NC_008613	131520 nt	161	0	161
Photobacterium damselae subsp. piscicida	pP99-018	NC_008612	150157 nt	187	0	187
Photobacterium profundum SS9	pPBPR1	NC_005871	80033 nt	67	0	67
Polaromonas naphthalenivorans CJ2	pPNAP05	NC_008761	58808 nt	52	0	53
Polaromonas naphthalenivorans CJ2	pPNAP04	NC_008760	143747 nt	136	1	149
Polaromonas naphthalenivorans CJ2	pPNAP03	NC_008759	171866 nt	155	1	161
Polaromonas naphthalenivorans CJ2	pPNAP02	NC_008758	190172 nt	161	1	171
Polaromonas naphthalenivorans CJ2	pPNAP01	NC_008757	353291 nt	305	0	317
Polaromonas sp. JS666	2	NC_007950	338007 nt	310	0	323
Polaromonas sp. JS666	1	NC_007949	360405 nt	326	0	341
Prosthecochloris aestuarii DSM 271	pPAES01	NC_011061	66772 nt	64	0	67
Proteus mirabilis	pHI4320	NC_010555	36289 nt	55	0	55
Proteus vulgaris	Rts1	NC_003905	217182 nt	300	0	300
Pseudomonas aeruginosa	Rms149	NC_007100	57121 nt	52	0	56
Pseudomonas aeruginosa	pBS228	NC_008357	89147 nt	98	0	104
Pseudomonas fluorescens	pNAH20	NC_012674	83042 nt	93	0	93
Pseudomonas fluorescens SBW25	pQBR103	NC_009444	425094 nt	474	0	478
Pseudomonas putida	pW2	NC_013176	76494 nt	70	0	70
Pseudomonas putida	pWW53	NC_008275	107929 nt	86	0	87
Pseudomonas putida	NAH7	NC_007926	82232 nt	84	0	84
Pseudomonas putida	pDTG1	NC_004999	83042 nt	90	0	90
Pseudomonas putida	pDK1	NC_014124	128921 nt	112	0	112
Pseudomonas putida	pWW0	NC_003350	116580 nt	139	0	150
Pseudomonas resinovorans	pCAR1	NC_004444	199035 nt	190	0	190
Pseudomonas resinovorans	pCAR1.2	NC_011838	200231 nt	192	0	192
Pseudomonas sp. ADP	pADP-1	NC_004956	108845 nt	104	0	104
Pseudomonas sp. ND6	pND6-1	NC_005244	101858 nt	102	0	102
Pseudomonas syringae pv. maculicola	pPMA4326A	NC_005918	46697 nt	50	0	53
Pseudomonas syringae pv. phaseolicola 1448A	small plasmid	NC_007275	51711 nt	60	0	60
Pseudomonas syringae pv. phaseolicola 1448A	large plasmid	NC_007274	131950 nt	127	0	149
Pseudomonas syringae pv. syringae	pPSR1	NC_005205	72601 nt	55	0	55
Pseudomonas syringae pv. tomato str. DC3000	pDC3000A	NC_004633	73661 nt	68	0	73
Pseudomonas syringae pv. tomato str. DC3000	pDC3000B	NC_004632	67473 nt	70	0	77
Ralstonia eutropha H16	megaplasmid pHG1	NC_005241	452156 nt	420	0	423
Ralstonia eutropha JMP134	pJP4	NC_005912	87688 nt	87	0	88
Ralstonia eutropha JMP134	1	NC_007337	87688 nt	88	0	90
Ralstonia eutropha JMP134	megaplasmid	NC_007336	634917 nt	512	1	555
Ralstonia pickettii 12D	pRp12D03	NC_012851	51398 nt	67	0	68
Ralstonia pickettii 12D	pRp12D02	NC_012849	273136 nt	276	0	287
Ralstonia pickettii 12D	pRp12D01	NC_012855	389779 nt	427	0	432
Ralstonia pickettii 12J	pRPIC01	NC_010683	80934 nt	90	0	91
Ralstonia solanacearum CFBP2957	RCFBPv3_mp	NC_014309	2163376 nt	1790	4	1856

Ralstonia solanacearum GMI1000	pGMI1000MP	NC_003296	2094509 nt	1676	6	1684
Rhizobium etli CFN 42	p42b	NC_007763	184338 nt	163	0	165
Rhizobium etli CFN 42	p42a	NC_007762	194229 nt	175	0	182
Rhizobium etli CFN 42	p42c	NC_007764	250948 nt	232	0	234
Rhizobium etli CFN 42	symbiotic plasmid p4	NC_004041	371254 nt	336	0	354
Rhizobium etli CFN 42	p42e	NC_007765	505334 nt	455	0	459
Rhizobium etli CFN 42	p42f	NC_007766	642517 nt	567	0	573
Rhizobium etli CIAT 652	pB	NC_010996	429111 nt	356	0	381
Rhizobium etli CIAT 652	pA	NC_010998	414090 nt	381	0	383
Rhizobium etli CIAT 652	pC	NC_010997	1091523 nt	976	0	989
Rhizobium leguminosarum bv. trifolii WSM1325	pR132505	NC_012854	294782 nt	275	0	287
Rhizobium leguminosarum bv. trifolii WSM1325	pR132504	NC_012852	350312 nt	306	0	315
Rhizobium leguminosarum bv. trifolii WSM1325	pR132503	NC_012853	516088 nt	494	0	560
Rhizobium leguminosarum bv. trifolii WSM1325	pR132502	NC_012858	660973 nt	624	0	661
Rhizobium leguminosarum bv. trifolii WSM1325	pR132501	NC_012848	828924 nt	737	0	769
Rhizobium leguminosarum bv. trifolii WSM2304	pRLG204	NC_011371	257956 nt	223	0	228
Rhizobium leguminosarum bv. trifolii WSM2304	pRLG203	NC_011370	308747 nt	256	0	282
Rhizobium leguminosarum bv. trifolii WSM2304	pRLG202	NC_011366	501946 nt	450	0	458
Rhizobium leguminosarum bv. trifolii WSM2304	pRLG201	NC_011368	1266105 nt	1161	0	1244
Rhizobium leguminosarum bv. viciae 3841	pRL8	NC_008383	147463 nt	132	0	142
Rhizobium leguminosarum bv. viciae 3841	pRL7	NC_008382	151564 nt	156	0	188
Rhizobium leguminosarum bv. viciae 3841	pRL9	NC_008379	352782 nt	305	0	313
Rhizobium leguminosarum bv. viciae 3841	pRL10	NC_008381	488135 nt	442	0	471
Rhizobium leguminosarum bv. viciae 3841	pRL11	NC_008384	684202 nt	634	0	644
Rhizobium leguminosarum bv. viciae 3841	pRL12	NC_008378	870021 nt	780	0	790
Rhizobium sp. NGR234	pNGR234a	NC_000914	536165 nt	405	4	422
Rhizobium sp. NGR234	pNGR234b	NC_012586	2430033 nt	2328	7	2351
Rhodobacter capsulatus SB 1003	pRCB133	NC_014035	132962 nt	149	0	153
Rhodobacter sphaeroides 2.4.1	C	NC_007489	105284 nt	82	0	90
Rhodobacter sphaeroides 2.4.1	D	NC_007490	100828 nt	87	0	98
Rhodobacter sphaeroides 2.4.1	A	NC_009007	114045 nt	87	0	90
Rhodobacter sphaeroides 2.4.1	B	NC_007488	114178 nt	100	0	102
Rhodobacter sphaeroides ATCC 17025	pRSPA03	NC_009431	121962 nt	109	0	114
Rhodobacter sphaeroides ATCC 17025	pRSPA02	NC_009430	289489 nt	268	6	278
Rhodobacter sphaeroides ATCC 17025	pRSPA01	NC_009429	877879 nt	802	17	849
Rhodobacter sphaeroides ATCC 17029	pRSPH01	NC_009040	122606 nt	107	0	118
Rhodobacter sphaeroides KD131	pRSKD131B	NC_011960	103355 nt	102	0	102
Rhodobacter sphaeroides KD131	pRSKD131A	NC_011962	157345 nt	142	0	142
Rhodoferax ferrireducens T118	plasmid1	NC_007901	257447 nt	248	1	251
Rhodothermus marinus DSM 4252	pRMAR01	NC_013502	125133 nt	97	0	107
Rickettsia felis URRWXCal2	pRF	NC_007110	62829 nt	68	0	68
Roseobacter denitrificans OCh 114	pTB2	NC_008387	69269 nt	56	0	56
Roseobacter denitrificans OCh 114	pTB1	NC_008386	106469 nt	105	0	106
Ruegeria pomeroyi DSS-3	megaplasmid	NC_006569	491611 nt	442	2	447
Ruegeria sp. PR1b	pSD20	NC_004929	76093 nt	52	0	52
Ruegeria sp. PR1b	pSD25	NC_004574	148650 nt	133	0	133
Ruegeria sp. TM1040	unnamed	NC_008042	130973 nt	106	7	116
Ruegeria sp. TM1040	mega plasmid	NC_008043	821788 nt	728	29	758
Salinibacter ruber M8	pSR84	NC_014157	84340 nt	70	0	70
Salmonella enterica	pOU1113	NC_007208	80156 nt	83	0	89
Salmonella enterica	pAM04528	NC_012693	158213 nt	180	0	180
Salmonella enterica subsp. enterica serovar Agona s	unnamed	NC_011148	37978 nt	52	0	53
Salmonella enterica subsp. enterica serovar Cholera	pOU7519	NC_010119	127212 nt	127	0	143
Salmonella enterica subsp. enterica serovar Cholera	pMAK1	NC_009981	208409 nt	222	0	236
Salmonella enterica subsp. enterica serovar Cholera	pSCV50	NC_006855	49558 nt	51	0	51
Salmonella enterica subsp. enterica serovar Cholera	pSC138	NC_006856	138742 nt	170	0	173
Salmonella enterica subsp. enterica serovar Dublin	pMAK2	NC_009980	61571 nt	52	0	54
Salmonella enterica subsp. enterica serovar Dublin	pOU1115	NC_010422	74589 nt	88	0	89
Salmonella enterica subsp. enterica serovar Dublin s	pCT02021853_74	NC_011204	74551 nt	103	1	122
Salmonella enterica subsp. enterica serovar Enteritic	pSE34	NC_010860	32950 nt	53	0	53
Salmonella enterica subsp. enterica serovar Heidelb	pSL476_91	NC_011081	91374 nt	125	0	131
Salmonella enterica subsp. enterica serovar Kentuck	pCVM29188_101	NC_011077	101461 nt	136	0	141
Salmonella enterica subsp. enterica serovar Kentuck	pCVM29188_146	NC_011076	146811 nt	170	4	204
Salmonella enterica subsp. enterica serovar Newpor	pSN254	NC_009140	176473 nt	190	0	198
Salmonella enterica subsp. enterica serovar Paratypl	pSPCV	NC_012124	55414 nt	62	0	65
Salmonella enterica subsp. enterica serovar Schwarz	pCVM19633_110	NC_011092	110227 nt	122	0	129
Salmonella enterica subsp. enterica serovar Typhi	R27	NC_002305	180461 nt	207	0	207
Salmonella enterica subsp. enterica serovar Typhi st	pHCM2	NC_003385	106516 nt	128	1	129
Salmonella enterica subsp. enterica serovar Typhi st	pHCM1	NC_003384	218160 nt	235	0	246
Salmonella enterica subsp. enterica serovar Typhimu	pU302L	NC_006816	84514 nt	103	0	104
Salmonella enterica subsp. enterica serovar Typhimu	pSLT-BT	NC_013437	117047 nt	110	0	120
Salmonella enterica subsp. enterica serovar Typhimu	R64	NC_005014	120826 nt	134	0	137
Salmonella enterica subsp. enterica serovar Typhimu	pSLT	NC_003277	93939 nt	102	1	112
Sebaldella termitidis ATCC 33386	pSTERM01	NC_013518	54160 nt	51	0	55
Serratia entomophila	pADAP	NC_002523	153404 nt	141	0	148
Serratia marcescens	R478	NC_005211	274762 nt	291	0	295
Serratia proteamaculans 568	pSPRO01	NC_009829	46804 nt	51	0	51
Shewanella baltica OS155	pSbal02	NC_009036	74000 nt	71	0	79
Shewanella baltica OS155	pSbal01	NC_009035	116763 nt	92	10	102
Shewanella baltica OS185	pS18501	NC_009661	83224 nt	71	0	75
Shewanella baltica OS195	pS19501	NC_009998	75605 nt	67	0	68
Shewanella baltica OS195	pS19502	NC_009999	75508 nt	73	0	74
Shewanella baltica OS223	pS22303	NC_011665	59223 nt	51	0	51
Shewanella baltica OS223	pS22302	NC_011668	65448 nt	61	0	64
Shewanella baltica OS223	pS22301	NC_011664	88311 nt	79	0	88
Shewanella oneidensis MR-1	megaplasmid	NC_004349	161613 nt	149	0	184
Shewanella sp. ANA-3	1	NC_008573	278942 nt	249	0	251
Shigella boydii CDC 3083-94	pBS512_33	NC_010657	33103 nt	52	0	52
Shigella boydii CDC 3083-94	pBS512_211	NC_010660	210919 nt	242	1	309

Shigella boydii Sb227	pSB4_227	NC_007608	126697 nt	148	0	150
Shigella dysenteriae Sd197	pSD1_197	NC_007607	182726 nt	223	0	224
Shigella flexneri 2a str. 301	pCP301	NC_004851	221618 nt	263	0	269
Shigella flexneri 5a	virulence plasmid pW	NC_002698	221851 nt	293	1	294
Shigella sonnei	pEG356	NC_013727	70275 nt	88	0	102
Shigella sonnei Ss046	pSS_046	NC_007385	214396 nt	238	0	242
Sinorhizobium medicae WSM419	pSMED03	NC_009622	219313 nt	149	0	196
Sinorhizobium medicae WSM419	pSMED02	NC_009621	1245408 nt	1094	0	1278
Sinorhizobium medicae WSM419	pSMED01	NC_009620	1570951 nt	1441	1	1478
Sinorhizobium meliloti	pSmeSM11a	NC_013545	144170 nt	160	0	160
Sinorhizobium meliloti 1021	pSymA	NC_003037	1354226 nt	1290	3	1293
Sinorhizobium meliloti 1021	pSymB	NC_003078	1683333 nt	1569	1	1570
Sinorhizobium meliloti SM11	pSmeSM11b	NC_010865	181251 nt	166	0	166
Sodalis glossinidius	pSG1	NC_007182	81553 nt	64	0	91
Sodalis glossinidius	pSG1	NC_007183	81553 nt	64	0	91
Sodalis glossinidius str. 'morsitans'	pSG1	NC_007713	83306 nt	54	0	54
Sphingobium japonicum UT26S	pCHQ1	NC_014007	190974 nt	224	0	224
Sphingomonas sp. KA1	pCAR3	NC_008308	254797 nt	263	0	263
Sphingomonas wittichii RW1	pSWIT02	NC_009508	222757 nt	210	0	226
Sphingomonas wittichii RW1	pSWIT01	NC_009507	310228 nt	285	0	288
Spirosoma linguale DSM 74	pSLIN02	NC_013732	146936 nt	142	0	147
Spirosoma linguale DSM 74	pSLIN01	NC_013731	189452 nt	177	0	185
Synechococcus elongatus PCC 7942	1	NC_007595	46366 nt	50	0	50
Synechococcus elongatus PCC 7942	pANL	NC_004073	46366 nt	58	0	58
Synechococcus sp. PCC 7002	pAQ6	NC_010480	124030 nt	109	0	109
Synechococcus sp. PCC 7002	pAQ7	NC_010474	186459 nt	165	0	165
Synechocystis sp. PCC 6803	pSYSX	NC_005232	106004 nt	110	0	110
Synechocystis sp. PCC 6803	pSYSA	NC_005230	103307 nt	106	0	106
Synechocystis sp. PCC 6803	pSYSM	NC_005229	119895 nt	132	0	132
Thauera sp. MZ1T	pTha01	NC_011667	78374 nt	75	0	79
Thermomicrobium roseum DSM 5159	unnamed	NC_011961	917738 nt	937	0	937
Thermus thermophilus HB27	pTT27	NC_005838	232605 nt	228	0	228
Thermus thermophilus HB8	pTT27	NC_006462	256992 nt	251	0	251
Thioalkalivibrio sp. K90mix	pTK9001	NC_013930	240256 nt	280	0	285
Thiomonas sp. 3As	pTHI	NC_014144	46756 nt	68	0	68
Vibrio fischeri ES114	pES100	NC_006842	45849 nt	57	0	57
Vibrio fischeri MJ11	pMJ100	NC_011185	179459 nt	195	0	195
Vibrio harveyi ATCC BAA-1116	pVIBHAR	NC_009777	89008 nt	120	0	120
Vibrio sp. 0908	p0908	NC_010113	81413 nt	95	0	99
Vibrio sp. 23023	p23023	NC_010112	52527 nt	64	0	64
Vibrio tapetis	pVT1	NC_010614	82266 nt	85	0	85
Vibrio vulnificus	pC4602-2	NC_009703	66946 nt	67	0	67
Vibrio vulnificus	pR99	NC_009701	68446 nt	71	0	71
Vibrio vulnificus	pC4602-1	NC_009702	56628 nt	69	0	69
Vibrio vulnificus YJ016	pYJ016	NC_005128	48508 nt	69	0	69
Xanthobacter autotrophicus Py2	pXAUT01	NC_009717	316164 nt	289	0	308
Xanthomonas axonopodis pv. citri str. 306	pXAC64	NC_003922	64920 nt	73	0	73
Xanthomonas campestris pv. vesicatoria str. 85-10	pXCV183	NC_007507	182572 nt	172	1	179
Xenorhabdus nematophila ATCC 19061	XNC1_p	NC_014170	155327 nt	175	0	187
Xylella fastidiosa 9a5c	pXF51	NC_002490	51158 nt	64	0	65
Yersinia enterocolitica	pYVe8081	NC_005017	67720 nt	67	0	68
Yersinia enterocolitica	pYVe227	NC_002120	69673 nt	69	0	69
Yersinia enterocolitica	pYVa127/90	NC_004564	66591 nt	74	1	75
Yersinia enterocolitica	pYE854	NC_010377	95499 nt	232	0	232
Yersinia enterocolitica subsp. enterocolitica 8081	pYVe8081	NC_008791	67721 nt	72	0	89
Yersinia pestis	pG8786	NC_006323	137036 nt	146	1	148
Yersinia pestis Angola	new_pCD	NC_010157	68190 nt	88	1	98
Yersinia pestis Angola	pMT-pPCP	NC_010158	114570 nt	120	2	141
Yersinia pestis Antiqua	pCD	NC_008122	70299 nt	89	0	89
Yersinia pestis Antiqua	pMT	NC_008120	96471 nt	99	0	99
Yersinia pestis biovar Microtus str. 91001	pCD1	NC_005813	70159 nt	85	1	99
Yersinia pestis biovar Microtus str. 91001	pMT1	NC_005815	106642 nt	122	0	128
Yersinia pestis biovar Orientalis str. IP275	pIP1202	NC_009141	182913 nt	212	0	221
Yersinia pestis CA88-4125	pCD1	NC_009595	70305 nt	73	1	87
Yersinia pestis CA88-4125	pMT1	NC_009596	96210 nt	99	0	100
Yersinia pestis CO92	pCD1	NC_003131	70305 nt	71	1	102
Yersinia pestis CO92	pMT1	NC_003134	96210 nt	101	0	107
Yersinia pestis KIM 10	pCD1	NC_004836	70504 nt	68	1	91
Yersinia pestis KIM 10	pCD1	NC_004839	70559 nt	68	1	90
Yersinia pestis KIM 10	pMT1	NC_004835	100984 nt	102	0	102
Yersinia pestis KIM 10	pMT-1	NC_004838	100990 nt	116	0	119
Yersinia pestis Nepal516	pMT	NC_008118	100918 nt	104	0	104
Yersinia pestis Pestoides F	CD	NC_009377	71507 nt	88	0	93
Yersinia pestis Pestoides F	MT	NC_009378	137010 nt	131	0	131
Yersinia pestis Z176003	pCD1	NC_014017	68342 nt	67	0	67
Yersinia pestis Z176003	pMT1	NC_014022	94251 nt	76	0	76
Yersinia pseudotuberculosis	pGDT4	NC_011759	94967 nt	101	0	101
Yersinia pseudotuberculosis IP 31758	plasmid_59kb	NC_009704	58679 nt	64	0	65
Yersinia pseudotuberculosis IP 31758	plasmid_153kb	NC_009705	153140 nt	136	0	136
Yersinia pseudotuberculosis IP 32953	pYV	NC_006153	68525 nt	95	0	99
Yersinia pseudotuberculosis PB1/+	pYPTS01	NC_010635	69812 nt	87	0	87
Yersinia ruckeri	pYR1	NC_009139	158038 nt	185	0	185
Zymomonas mobilis subsp. mobilis ZM4	pZZM401	NC_013784	37066 nt	51	0	53

Appendix B

Fig. B.1. Newick standard file for 527 Gram-negative plasmids using majority voting. Because an unrooted consensus tree is order dependent, we used human X and cattle X reproductive proteins as an out group to force clustering of the plasmids.

```
(((((((((((((((((((((NC_011964_Escherichi:100.0,(NC_011092_Salmonella:100.0,(NC_010119_Salmonella:100.0,NC_006856_Salmonella:
100.0):77.0):85.0):59.0,((NC_006625_Klebsiella:100.0,NC_005249_Klebsiella:100.0):100.0,NC_014107_Enterobact:100.0):99.0):51.0,
(((NC_010870_Klebsiella:100.0,((NC_012555_Enterobact:100.0,NC_012556_Enterobact:100.0):100.0,NC_005211_Serratia:100.0):97.
0):83.0,NC_009838_Escherichi:100.0):100.0,(NC_002305_Salmonella:100.0,((NC_009981_Salmonella:100.0,NC_013365_Escherichi:1
00.0):74.0,NC_003384_Salmonella:100.0):100.0):100.0):100.0):51.0,(((((((NC_002128_Escherichi:100.0,NC_007414_Escherichi:100.
0):93.0,(NC_011350_Escherichi:100.0,NC_013010_Escherichi:100.0):100.0):100.0,NC_006816_Salmonella:100.0):42.0,NC_011413_E
scherichi:100.0):41.0,(((NC_013950_Klebsiella:100.0,((NC_009650_Klebsiella:100.0,NC_014016_Klebsiella:100.0):100.0,NC_014312_
Klebsiella:100.0):83.0):100.0,NC_009649_Klebsiella:100.0):96.0,((NC_003277_Salmonella:100.0,NC_013437_Salmonella:100.0):100.
0,NC_007208_Salmonella:100.0):100.0):81.0):18.0,(NC_013354_Escherichi:100.0,NC_013942_Escherichi:100.0):91.0):7.00,((((((((((N
C_010720_Escherichi:100.0,(((NC_013362_Escherichi:100.0,((NC_013542_Klebsiella:100.0,NC_013727_Shigella:100.0):100.0,NC_01
1812_Escherichi:100.0):96.0):59.0,(NC_005327_Escherichi:100.0,NC_009133_Escherichi:100.0):94.0):45.0,NC_006671_Escherichi:10
0.0):95.0):73.0,((NC_004998_Escherichi:100.0,((NC_011076_Salmonella:100.0,(NC_011747_Escherichi:100.0,(NC_010409_Escherichi
:100.0,NC_009837_Escherichi:100.0):51.0):69.0):96.0,NC_010488_Escherichi:100.0):75.0):88.0,((((NC_013175_Escherichi:100.0,(NC_
007941_Escherichi:100.0,NC_011749_Escherichi:100.0):90.0):100.0,NC_008460_Escherichi:100.0):96.0,NC_013951_Klebsiella:100.0
):80.0,NC_010558_Escherichi:100.0):86.0):81.0):85.0,(NC_009602_Escherichi:100.0,NC_012944_Escherichi:100.0):100.0):86.0,(NC_0
11603_Escherichi:100.0,NC_010862_Escherichia:100.0):100.0):86.0,((NC_010719_Escherichi:100.0,NC_007675_Escherichi:100.0):84
.0,NC_011980_Escherichi:100.0):68.0):60.0,(((NC_013121_Escherichi:100.0,NC_013122_Escherichi:100.0):98.0,NC_013507_Escheric
hi:100.0):94.0,NC_013366_Escherichi:100.0):61.0):36.0,NC_011752_Escherichi:100.0):26.0,((((NC_010660_Shigella:100.0,(NC_00760
7_Shigella:100.0,(NC_004851_Shigella:100.0,NC_002698_Shigella:100.0):100.0):98.0):77.0,NC_007385_Shigella:100.0):72.0,NC_007
608_Shigella:100.0):100.0,(NC_009786_Escherichi:100.0,NC_014232_Escherichi:100.0):75.0):18.0):18.0,(NC_002142_Escherichi:100.
0,(NC_009790_Escherichi:100.0,NC_014233_Escherichi:100.0):93.0):48.0):14.0,(((((((NC_011419_Escherichi:100.0,(NC_011077_Salm
onella:100.0,NC_011081_Salmonella:100.0):100.0):93.0,NC_005014_Salmonella:100.0):98.0,NC_013120_Escherichi:100.0):100.0,NC
_007365_Escherichi:100.0):56.0,NC_009788_Escherichi:100.0):41.0,((NC_013728_Escherichi:100.0,NC_013369_Escherichi:100.0):68
.0,NC_012487_Escherichi:100.0):57.0):87.0,(NC_014234_Escherichi:100.0,NC_007635_Escherichi:100.0):73.0):17.0):3.00):37.0,NC_0
11754_Escherichi:100.0):36.0):61.0,((NC_009651_Klebsiella:100.0,NC_010886_Klebsiella:100.0):97.0,NC_011743_Escherichi:100.0):
98.0):69.0,NC_003905_Proteus:100.0):38.0,(((NC_006143_Aeromonas:100.0,NC_007100_Pseudomona:100.0):100.0,(((((((((NC_012
692_Escherichi:100.0,NC_009140_Salmonella:100.0):100.0,NC_012693_Salmonella:100.0):98.0,NC_009141_Yersinia:100.0):68.0,NC
_012690_Escherichi:100.0):100.0,((NC_008613_Photobacte:100.0,NC_008612_Photobacte:100.0):100.0,NC_009139_Yersinia:100.0)
:98.0):66.0,NC_012885_Aeromonas:100.0):100.0,NC_014170_Xenorhabdu:100.0):93.0,NC_009349_Aeromonas:100.0):100.0,NC_01
2886_Escherichi:100.0):99.0):45.0,((((((NC_007682_Escherichi:100.0,NC_014231_Escherichi:100.0):93.0,NC_011617_Klebsiella:100.
0):66.0,NC_014208_Klebsiella:100.0):55.0,(NC_011385_Klebsiella:100.0,(NC_014368_Klebsiella:100.0,NC_009132_Escherichi:100.0):
99.0):100.0):96.0,NC_009980_Salmonella:100.0):98.0,NC_011383_Klebsiella:100.0):100.0):20.0):41.0,((((((NC_006466_Cupriavidu:1
00.0,NC_007971_Cupriavidu:100.0):100.0,NC_008573_Shewanella:100.0):58.0,(NC_006525_Cupriavidu:100.0,NC_007972_Cupriavi
du:100.0):100.0):64.0,NC_008341_Nitrosomon:100.0):100.0,NC_008739_Marinobact:100.0):100.0,NC_009444_Pseudomona:100.0):
94.0):14.0,((((((((NC_005017_Yersinia:100.0,NC_004564_Yersinia:100.0):59.0,NC_008791_Yersinia:100.0):100.0,NC_002120_Yersinia
:100.0):99.0,((((NC_009377_Yersinia:100.0,NC_008122_Yersinia:100.0):100.0,((NC_009595_Yersinia:100.0,(NC_005813_Yersinia:100
.0,NC_003131_Yersinia:100.0):36.0):65.0,(NC_004836_Yersinia:100.0,NC_004839_Yersinia:100.0):69.0):100.0):95.0,(NC_010635_Yer
sinia:100.0,(NC_010157_Yersinia:100.0,NC_006153_Yersinia:100.0):67.0):96.0):100.0,NC_014017_Yersinia:100.0):100.0):100.0,NC_0
09350_Aeromonas:100.0):100.0,(NC_011282_Klebsiella:100.0,(NC_014305_Erwinia:100.0,NC_009425_Enterobact:100.0):78.0):78.0
):73.0,(((((((NC_009596_Yersinia:100.0,NC_003134_Yersinia:100.0):57.0,(NC_005815_Yersinia:100.0,NC_004838_Yersinia:100.0):41
.0):64.0,(NC_008118_Yersinia:100.0,NC_008120_Yersinia:100.0):71.0):65.0,NC_004835_Yersinia:100.0):100.0,(NC_009378_Yersinia:
100.0,NC_006323_Yersinia:100.0):100.0):69.0,NC_010158_Yersinia:100.0):73.0,NC_014022_Yersinia:100.0):100.0,NC_003385_Salm
onella:100.0):100.0):32.0,((((NC_010422_Salmonella:100.0,NC_011204_Salmonella:100.0):100.0,(NC_006855_Salmonella:100.0,NC_
012124_Salmonella:100.0):100.0):98.0,NC_011281_Klebsiella:100.0):62.0,(NC_013717_Citrobacte:100.0,NC_011416_Escherichi:100
.0):91.0):30.0):10.0):4.00,(((((((((NC_008379_Rhizobium:100.0,NC_012852_Rhizobium:100.0):100.0,NC_011371_Rhizobium:100.0):
100.0,NC_007763_Rhizobium:100.0):100.0,NC_011987_Agrobacter:100.0):39.0,(((((NC_008381_Rhizobium:100.0,NC_012854_Rhizo
bium:100.0):100.0,NC_007764_Rhizobium:100.0):100.0,NC_011981_Agrobacter:100.0):71.0,(NC_003064_Agrobacter:100.0,NC_012
853_Rhizobium:100.0):48.0):48.0,((NC_004041_Rhizobium:100.0,NC_010996_Rhizobium:100.0):100.0,NC_000914_Rhizobium:100.0
):79.0):36.0):17.0,NC_011370_Rhizobium:100.0):22.0,((((((((((((NC_003296_Ralstonia:100.0,NC_014309_Ralstonia:100.0):100.0,NC_
007974_Cupriavidu:100.0):100.0,NC_010625_Burkholder:100.0):77.0,(((NC_009621_Sinorhizob:100.0,NC_003037_Sinorhizob:100.0
):100.0,(((((NC_008378_Rhizobium:100.0,NC_012848_Rhizobium:100.0):100.0,NC_011368_Rhizobium:100.0):73.0,NC_010997_Rhiz
obium:100.0):85.0,NC_007766_Rhizobium:100.0):100.0,(NC_012586_Rhizobium:100.0,(NC_009620_Sinorhizob:100.0,NC_003078_S
inorhizob:100.0):100.0):100.0):99.0):100.0,((NC_008384_Rhizobium:100.0,NC_012858_Rhizobium:100.0):100.0,((NC_007765_Rhizo
bium:100.0,NC_010998_Rhizobium:100.0):77.0,NC_011366_Rhizobium:100.0):95.0):100.0):64.0):75.0,NC_013855_Azospirill:100.0):
94.0,NC_013856_Azospirill:100.0):54.0,NC_009429_Rhodobacte:100.0):92.0,(NC_013857_Azospirill:100.0,NC_008688_Paracoccus:1
00.0):55.0):32.0,NC_008043_Ruegeria:100.0):51.0,NC_006569_Ruegeria:100.0):23.0,NC_013858_Azospirill:100.0):25.0,NC_014258_
Pantoea:100.0):15.0,(NC_010510_Methylobac:100.0,(NC_012811_Methylobac:100.0,NC_011887_Methylobac:100.0):79.0):35.0):13.
0):9.00,(NC_011961_Thermomicr:100.0,NC_013859_Azospirill:100.0):24.0):3.00,(NC_007336_Ralstonia:100.0,((NC_012849_Ralstoni
```

```
a:100.0,(NC_010553_Burkholder:100.0,NC_009227_Burkholder:100.0):99.0):91.0,(((NC_014120_Burkholder:100.0,NC_010529_Cupr
iavidu:100.0):99.0,NC_010627_Burkholder:100.0):63.0,NC_005241_Ralstonia:100.0):69.0):42.0):17.0):7.00,(NC_008826_Methylibiu:
100.0,NC_009230_Burkholder:100.0):82.0):6.00):2.00,(((((((((((NC_008242_Chelativor:100.0,NC_009717_Xanthobact:100.0):89.0,N
C_009475_Bradyrhizo:100.0):100.0,NC_007960_Nitrobacte:100.0):99.0,NC_009670_Ochrobactr:100.0):51.0,NC_002679_Mesorhizo
b:100.0):100.0,((((((((NC_003065_Agrobacter:100.0,NC_002147_Agrobacter:100.0):99.0,NC_011990_Agrobacter:100.0):100.0,NC_0
11984_Agrobacter:100.0):58.0,(NC_011982_Agrobacter:100.0,(NC_010929_Agrobacter:100.0,NC_002377_Agrobacter:100.0):99.0):
100.0):49.0,((NC_002575_Agrobacter:100.0,NC_010841_Agrobacter:100.0):100.0,NC_007762_Rhizobium:100.0):66.0):91.0,NC_008
383_Rhizobium:100.0):93.0,NC_011991_Agrobacter:100.0):61.0,(NC_007961_Nitrobacte:100.0,NC_005873_Oligotroph:100.0):83.0)
:52.0,NC_010865_Sinorhizob:100.0):78.0):63.0,NC_002682_Mesorhizob:100.0):28.0,((NC_008382_Rhizobium:100.0,NC_011892_Me
thylobac:100.0):65.0,((NC_009622_Sinorhizob:100.0,NC_009669_Ochrobactr:100.0):92.0,NC_009671_Ochrobactr:100.0):76.0):22.0)
:27.0,(NC_011758_Methylobac:100.0,NC_012987_Methylobac:100.0):84.0):16.0,((NC_011986_Agrobacter:100.0,NC_013545_Sinorh
izob:100.0):84.0,NC_007959_Nitrobacte:100.0):62.0):17.0,(NC_004574_Ruegeria:100.0,(NC_009957_Dinoroseob:100.0,NC_009955
_Dinoroseob:100.0):100.0):100.0):8.00,(((((NC_013211_Acetobacte:100.0,NC_013210_Acetobacte:100.0):100.0,NC_013212_Acetob
acte:100.0):52.0,NC_006672_Gluconobac:100.0):70.0,(NC_007274_Pseudomona:100.0,NC_013370_Escherichi:100.0):62.0):22.0,(NC
_007949_Polaromona:100.0,NC_013860_Azospirill:100.0):23.0):3.00):1.00,((((((((((NC_011786_Borrelia:100.0,NC_008564_Borrelia:
100.0):58.0,NC_012164_Borrelia:100.0):100.0,(NC_011877_Borrelia:100.0,NC_011859_Borrelia:100.0):100.0):97.0,NC_012177_Borr
elia:100.0):96.0,((((((((NC_012244_Borrelia:100.0,NC_012194_Borrelia:100.0):71.0,NC_012193_Borrelia:100.0):46.0,(NC_012175_B
orrelia:100.0,NC_012199_Borrelia:100.0):56.0):78.0,NC_012505_Borrelia:100.0):49.0,NC_001857_Borrelia:100.0):29.0,((NC_013129
_Borrelia:100.0,NC_013130_Borrelia:100.0):100.0,NC_012504_Borrelia:100.0):40.0):79.0,(NC_012241_Borrelia:100.0,NC_011876_B
orrelia:100.0):74.0):100.0,NC_012234_Borrelia:100.0):91.0,NC_013128_Borrelia:100.0):79.0):55.0,NC_011784_Borrelia:100.0):100.0
,NC_006129_Borrelia:100.0):100.0,(NC_011246_Borrelia:100.0,NC_011247_Borrelia:100.0):100.0):100.0,(((((NC_000956_Borrelia:10
0.0,((NC_012253_Borrelia:100.0,NC_012106_Borrelia:100.0):53.0,NC_011720_Borrelia:100.0):100.0):100.0,NC_012247_Borrelia:10
0.0):100.0,(NC_011869_Borrelia:100.0,NC_011842_Borrelia:100.0):71.0):89.0,NC_012185_Borrelia:100.0):50.0,NC_008567_Borrelia
:100.0):93.0):67.0,((((NC_011867_Borrelia:100.0,NC_011857_Borrelia:100.0):100.0,((NC_012202_Borrelia:100.0,NC_008565_Borreli
a:100.0):89.0,(NC_001855_Borrelia:100.0,NC_012184_Borrelia:100.0):100.0):80.0):91.0,(NC_012167_Borrelia:100.0,NC_012182_Bo
rrelia:100.0):100.0):61.0,((NC_011860_Borrelia:100.0,NC_012166_Borrelia:100.0):86.0,NC_011856_Borrelia:100.0):45.0):25.0):68.0,
(NC_010844_Leptospira:100.0,NC_010846_Leptospira:100.0):100.0):3.00,(((((((((NC_009999_Shewanella:100.0,NC_009661_Shewan
ella:100.0):100.0,NC_011664_Shewanella:100.0):100.0,NC_009998_Shewanella:100.0):67.0,NC_011668_Shewanella:100.0):86.0,NC
_011665_Shewanella:100.0):100.0,((NC_010614_Vibrio:100.0,NC_011311_Aliivibrio:100.0):91.0,NC_005871_Photobacte:100.0):78.
0):84.0,((NC_005128_Vibrio:100.0,NC_009702_Vibrio:100.0):99.0,(NC_009703_Vibrio:100.0,NC_009701_Vibrio:100.0):100.0):82.0):
98.0,((NC_007182_Sodalis:100.0,NC_007183_Sodalis:100.0):100.0,NC_007713_Sodalis:100.0):100.0):40.0,((NC_009966_Fluoribact:1
00.0,NC_006366_Legionella:100.0):100.0,NC_006365_Legionella:100.0):93.0):32.0,((((NC_002525_Escherichi:100.0,((NC_013718_Ci
trobacte:100.0,NC_010657_Shigella:100.0):100.0,((NC_013503_Escherichi:100.0,NC_010378_Escherichi:100.0):53.0,NC_010860_Sal
monella:100.0):100.0):92.0):69.0,(NC_010697_Erwinia:100.0,NC_011351_Escherichi:100.0):68.0):76.0,NC_010555_Proteus:100.0):5
9.0,(NC_005250_Listonella:100.0,NC_010404_Acinetobac:100.0):30.0):20.0,((((NC_004633_Pseudomona:100.0,NC_004632_Pseudo
mona:100.0):100.0,(NC_011759_Yersinia:100.0,NC_012752_Candidatus:100.0):95.0):38.0,NC_002523_Serratia:100.0):17.0,((((NC_0
04349_Shewanella:100.0,NC_009036_Shewanella:100.0):95.0,NC_010377_Yersinia:100.0):52.0,((NC_011880_Cyanothece:100.0,NC
_011667_Thauera:100.0):52.0,NC_008607_Pelobacter:100.0):68.0):28.0,NC_012797_Desulfovib:100.0):27.0):5.00,(((((NC_014157_S
alinibact:100.0,NC_012983_Hirschia:100.0):65.0,((NC_004929_Ruegeria:100.0,NC_008387_Roseobacte:100.0):78.0,NC_008042_Rue
geria:100.0):67.0):40.0,NC_007801_Jannaschia:100.0):37.0,NC_008014_Lawsonia:100.0):29.0,(((NC_010630_Nostoc:100.0,NC_0032
67_Nostoc:100.0):97.0,NC_003273_Nostoc:100.0):31.0,((NC_007595_Synechococ:100.0,NC_004073_Synechococ:100.0):100.0,NC_0
09959_Dinoroseob:100.0):35.0):15.0):5.00,((NC_012970_Methylovor:100.0,(((((((((((NC_005912_Ralstonia:100.0,NC_007337_Ralst
onia:100.0):100.0,(NC_004956_Pseudomona:100.0,NC_005088_Delftia:100.0):100.0):100.0,NC_013176_Pseudomona:100.0):89.0,N
C_013666_Burkholder:100.0):90.0,((NC_006830_Achromobac:100.0,(NC_008766_Acidovorax:100.0,NC_001735_Enterobact:100.0):
50.0):50.0,NC_010935_Comamonas:100.0):96.0):93.0,NC_008357_Pseudomona:100.0):99.0,NC_012919_Photobacte:100.0):43.0,N
C_006824_Aromatoleu:100.0):44.0,NC_013193_Candidatus:100.0):83.0,NC_005793_Achromobac:100.0):99.0,NC_009704_Yersinia:
100.0):100.0,NC_006139_Desulfotal:100.0):64.0):50.0,((NC_011061_Prosthecoc:100.0,NC_007110_Rickettsia:100.0):63.0,NC_00977
7_Vibrio:100.0):19.0):5.00,(NC_010679_Burkholder:100.0,(((NC_008545_Burkholder:100.0,(NC_012725_Burkholder:100.0,NC_0127
20_Burkholder:100.0):96.0):51.0,NC_011003_Burkholder:100.0):98.0,NC_008765_Acidovorax:100.0):66.0):7.00,(((NC_011143_Phen
ylobac:100.0,(NC_010070_Burkholder:100.0,NC_010802_Burkholder:100.0):100.0):56.0,((NC_009469_Acidiphili:100.0,(NC_009468_
Acidiphili:100.0,NC_009467_Acidiphili:100.0):77.0):66.0,NC_008738_Marinobact:100.0):28.0):9.00,((NC_009229_Burkholder:100.0,
NC_009228_Burkholder:100.0):42.0,(NC_012723_Burkholder:100.0,NC_012718_Burkholder:100.0):97.0):19.0):2.00,((NC_005838_T
hermus:100.0,NC_006462_Thermus:100.0):100.0,(NC_005863_Desulfovib:100.0,NC_008741_Desulfovib:100.0):100.0):7.00,((((((NC_
004444_Pseudomona:100.0,NC_011838_Pseudomona:100.0):100.0,NC_014124_Pseudomona:100.0):86.0,NC_008275_Pseudomon
a:100.0):99.0,((((NC_004999_Pseudomona:100.0,NC_012674_Pseudomona:100.0):100.0,NC_007926_Pseudomona:100.0):100.0,NC
_003350_Pseudomona:100.0):97.0,NC_005244_Pseudomona:100.0):83.0):100.0,NC_008757_Polaromona:100.0):35.0,(((((NC_0094
26_Novosphing:100.0,NC_002033_Novosphing:100.0):100.0,NC_008308_Sphingomon:100.0):65.0,NC_009508_Sphingomon:100.0):
86.0,(NC_009507_Sphingomon:100.0,NC_014007_Sphingobiu:100.0):74.0):41.0,NC_009427_Novosphing:100.0):21.0):2.00,((((NC_0
```

```
09040_Rhodobacte:100.0,NC_009007_Rhodobacte:100.0):100.0,NC_011960_Rhodobacte:100.0):100.0,NC_009430_Rhodobacte:10
0.0):98.0,((NC_014035_Rhodobacte:100.0,(NC_008243_Chelativor:100.0,(NC_011962_Rhodobacte:100.0,NC_007488_Rhodobacte:1
00.0):100.0):60.0):44.0,NC_010580_Beijerinck:100.0):30.0):3.00,(((NC_010474_Synechococ:100.0,NC_010480_Synechococ:100.0):4
9.0,((NC_005229_Synechocys:100.0,NC_005232_Synechocys:100.0):100.0,NC_005230_Synechocys:100.0):44.0):36.0,(((((NC_003922
_Xanthomona:100.0,NC_007507_Xanthomona:100.0):100.0,NC_014144_Thiomonas:100.0):58.0,NC_002490_Xylella:100.0):84.0,(NC
_010509_Methylobac:100.0,NC_008342_Nitrosomon:100.0):85.0):62.0,(NC_009973_Herpetosip:100.0,NC_009974_Herpetosip:100.
0):54.0):7.00):2.00,((((NC_012851_Ralstonia:100.0,NC_010683_Ralstonia:100.0):100.0,NC_013852_Allochroma:100.0):60.0,(((NC_00
5918_Pseudomona:100.0,(NC_005205_Pseudomona:100.0,NC_007275_Pseudomona:100.0):98.0):99.0,NC_011409_Haemophilu:10
0.0):39.0,NC_013930_Thioalkali:100.0):18.0):24.0,(NC_013283_Cronobacte:100.0,NC_009780_Cronobacte:100.0):100.0):5.00,(((((N
C_003276_Nostoc:100.0,NC_007410_Anabaena:100.0):100.0,(NC_010632_Nostoc:100.0,NC_010631_Nostoc:100.0):64.0):100.0,(NC
_003240_Nostoc:100.0,NC_007412_Anabaena:100.0):100.0):100.0,((((NC_009928_Acaryochlo:100.0,NC_009926_Acaryochlo:100.0)
:100.0,NC_009929_Acaryochlo:100.0):66.0,NC_009931_Acaryochlo:100.0):53.0,((NC_009932_Acaryochlo:100.0,NC_009933_Acaryo
chlo:100.0):95.0,(NC_009930_Acaryochlo:100.0,NC_009927_Acaryochlo:100.0):71.0):41.0):100.0):32.0,((NC_008758_Polaromona:1
00.0,NC_008760_Polaromona:100.0):72.0,NC_006823_Aromatoleu:100.0):44.0):2.00,((NC_009226_Burkholder:100.0,NC_010606_A
cinetobac:100.0):63.0,NC_013509_Edwardsiel:100.0):13.0,((NC_009829_Serratia:100.0,NC_012109_Desulfobac:100.0):44.0,((NC_00
9726_Coxiella:100.0,NC_010258_Coxiella:100.0):100.0,(NC_010124_Gluconacet:100.0,NC_013958_Nitrosococ:100.0):68.0):36.0):3.
00,(((((NC_005246_Erwinia:100.0,(NC_004464_Citrobacte:100.0,NC_011641_Klebsiella:100.0):97.0):100.0,(NC_014304_Erwinia:100.
0,NC_013973_Erwinia:100.0):62.0):61.0,(NC_009705_Yersinia:100.0,NC_014108_Enterobact:100.0):41.0):22.0,NC_007901_Rhodofe
rax:100.0):17.0,NC_013285_Cronobacte:100.0):7.00,((NC_011737_Cyanothece:100.0,NC_011738_Cyanothece:100.0):100.0,NC_011
885_Cyanothece:100.0):26.0,((NC_010815_Geobacter:100.0,NC_008761_Polaromona:100.0):87.0,(NC_013731_Spirosoma:100.0,NC
_013732_Spirosoma:100.0):100.0):21.0,((((NC_009431_Rhodobacte:100.0,NC_007490_Rhodobacte:100.0):90.0,(NC_009958_Dinor
oseob:100.0,NC_009956_Dinoroseob:100.0):57.0):38.0,NC_007489_Rhodobacte:100.0):37.0,(NC_008386_Roseobacte:100.0,NC_01
0373_Methylobac:100.0):75.0):10.0,(NC_011893_Methylobac:100.0,NC_009035_Shewanella:100.0):6.00,NC_013502_Rhodotherm:
100.0,((NC_010335_Caulobacte:100.0,NC_011185_Vibrio:100.0):35.0,NC_008759_Polaromona:100.0):3.00,((NC_010333_Caulobact
e:100.0,NC_007950_Polaromona:100.0):88.0,NC_012855_Ralstonia:100.0):1.00,(((NC_006134_Campylobac:100.0,NC_008790_Cam
pylobac:100.0):100.0,(NC_008770_Campylobac:100.0,NC_005012_Campylobac:100.0):100.0):71.0,(NC_010112_Vibrio:100.0,NC_00
6842_Vibrio:100.0):95.0):31.0,NC_010113_Vibrio:100.0):1.00,NC_011148_Salmonella:100.0):1.00,NC_002806_Microscill:100.0):3.0
0,NC_012439_Persephone:100.0):2.00,NC_013160_Cyanothece:100.0):4.00,NC_014249_Nostoc:100.0):5.00,NC_013784_Zymomon
as:100.0):9.00,NC_005915_Caedibacte:100.0):30.0,NC_013518_Sebaldella:100.0):100.0,cattle_X:100.0):100.0,human_X:100.0);
```

Appendix C

Fig. C.1. Newick standard file for 527 Gram-negative plasmids using averaged distance.

```
(((((NC_008826_Methylibium:7.8938,NC_009230_Burkholderia:7.2174)Branch
338:7.9586,((NC_007336_Ralstonia:8.896,((NC_005241_Ralstonia:8.3553,(NC_010627_Burkholderia:8.493,(NC_010529_Cupriavidus
:8.4162,NC_014120_Burkholderia:7.4558)Branch 229:8.5788)Branch 263:9.1075)Branch
289:9.1639,(NC_012849_Ralstonia:6.6713,(NC_009227_Burkholderia:7.9145,NC_010553_Burkholderia:5.8082)Branch
272:7.4103)Branch 298:7.2866)Branch 321:8.5579)Branch
353:8.832,(((NC_011887_Methylobacterium:8.1063,NC_012811_Methylobacterium:11.2127)Branch
322:9.903,(NC_006569_Ruegeria:8.2788,((NC_014258_Pantoea:9.1542,(((NC_008043_Ruegeria:10.0267,((NC_009429_Rhodobacter
:10.8876,(NC_013856_Azospirillum:9.2727,(NC_013855_Azospirillum:12.7813,((((NC_008384_Rhizobium:6.1274,NC_012858_Rhizob
ium:5.9125)Branch 8:8.1161,(NC_011366_Rhizobium:6.0516,(NC_007765_Rhizobium:5.6477,NC_010998_Rhizobium:4.8258)Branch
28:5.4577)Branch 36:6.3705)Branch 50:12.8452,((NC_003037_Sinorhizobium:11.9628,NC_009621_Sinorhizobium:11.1469)Branch
68:15.0068,((NC_012586_Rhizobium:16.7419,(NC_003078_Sinorhizobium:8.0393,NC_009620_Sinorhizobium:7.696)Branch
2:12.9158)Branch
12:20.0449,(NC_007766_Rhizobium:8.3156,(NC_010997_Rhizobium:12.5133,(NC_011368_Rhizobium:13.3854,(NC_008378_Rhizobi
um:7.4336,NC_012848_Rhizobium:7.6185)Branch 5:10.2555)Branch 47:12.3154)Branch 59:12.6572)Branch 61:13.3912)Branch
127:17.5045)Branch 141:17.7629)Branch
178:15.8469,(NC_010625_Burkholderia:14.1074,(NC_007974_Cupriavidus:16.7606,(NC_003296_Ralstonia:8.5412,NC_014309_Ralst
onia:8.6265)Branch 6:15.7108)Branch 151:17.4735)Branch 180:16.3024)Branch 196:16.3247)Branch 205:15.1179)Branch
232:12.4285)Branch 241:12.105,(NC_008688_Paracoccus:10.2719,NC_013857_Azospirillum:8.9668)Branch 261:9.6971)Branch
264:11.063)Branch
275:10.9653,((NC_000914_Rhizobium:9.6137,(NC_004041_Rhizobium:5.1159,NC_010996_Rhizobium:4.2)Branch
124:8.3278)Branch
242:9.7053,((NC_011981_Agrobacterium:10.3221,(NC_007764_Rhizobium:5.2841,(NC_008381_Rhizobium:7.3582,NC_012854_Rhiz
obium:4.7264)Branch 110:7.4889)Branch 139:9.4644)Branch
244:10.3054,(NC_003064_Agrobacterium:10.3884,NC_012853_Rhizobium:9.1826)Branch 246:10.109)Branch 265:10.4686)Branch
281:10.3753)Branch
292:10.7694,(NC_011987_Agrobacterium:7.9537,(NC_007763_Rhizobium:4.7771,(NC_011371_Rhizobium:4.2819,(NC_008379_Rhiz
obium:4.8417,NC_012852_Rhizobium:4.4815)Branch 102:5.3795)Branch 118:6.3465)Branch 152:9.0659)Branch 293:8.5493)Branch
296:9.8826)Branch 314:9.697,(NC_010510_Methylobacterium:8.4571,NC_013858_Azospirillum:8.8414)Branch 325:8.7193)Branch
332:9.3092)Branch 339:8.85)Branch 344:9.512,(NC_011961_Thermomicrobium:9.7244,NC_013859_Azospirillum:8.2204)Branch
350:9.0382)Branch 356:9.3441)Branch 363:9.2604)Branch
374:8.7949,((NC_009444_Pseudomonas:7.6294,(NC_008739_Marinobacter:5.7877,(NC_008341_Nitrosomonas:4.4994,((NC_006525
_Cupriavidus:1.7744,NC_007972_Cupriavidus:2.2967)Branch
117:6.2903,(NC_008573_Shewanella:6.5855,(NC_006466_Cupriavidus:2.8753,NC_007971_Cupriavidus:3.2163)Branch
132:6.6534)Branch 253:6.7067)Branch 258:6.6616)Branch 270:6.1125)Branch 295:6.4399)Branch
346:7.4617,((NC_011383_Klebsiella:5.0752,(NC_009980_Salmonella:4.975,((NC_011385_Klebsiella:5.6279,(NC_009132_Escherichia:
5.1936,NC_014368_Klebsiella:4.8173)Branch 130:6.3183)Branch
157:6.9598,(NC_014208_Klebsiella:4.2913,(NC_011617_Klebsiella:5.1746,(NC_007682_Escherichia:3.4744,NC_014231_Escherichia:
2.8996)Branch 160:3.6981)Branch 173:4.6412)Branch 179:4.5552)Branch 182:6.3002)Branch 202:6.2706)Branch
231:7.5206,((NC_006143_Aeromonas:5.362,NC_007100_Pseudomonas:3.7822)Branch
247:5.8716,((NC_012886_Escherichia:5.8091,(NC_009349_Aeromonas:7.3539,(NC_014170_Xenorhabdus:5.8891,(NC_012885_Aero
monas:5.1576,((NC_012690_Escherichia:5.2843,(NC_009141_Yersinia:6.255,(NC_012693_Salmonella:3.3429,(NC_009140_Salmonell
a:2.8957,NC_012692_Escherichia:3.0317)Branch 3:4.4169)Branch 10:5.0201)Branch 30:6.1279)Branch
45:7.9542,(NC_009139_Yersinia:5.2926,(NC_008612_Photobacterium:3.2769,NC_008613_Photobacterium:2.7919)Branch
18:5.0057)Branch 69:6.2604)Branch 95:7.6034)Branch 105:8.6688)Branch 155:8.7816)Branch 199:9.3166)Branch
257:8.5643,(NC_003905_Proteus:7.7523,((NC_011743_Escherichia:6.3099,(NC_009651_Klebsiella:7.5448,NC_010886_Klebsiella:6.6
605)Branch 223:7.4311)Branch
235:7.8292,((((NC_009838_Escherichia:6.9737,(NC_010870_Klebsiella:7.9332,(NC_005211_Serratia:7.0184,(NC_012555_Enterobact
er:0.25659,NC_012556_Enterobacter:0.26256)Branch 1:8.2729)Branch 64:8.4713)Branch 86:8.5749)Branch
93:10.3919,(NC_002305_Salmonella:5.4248,(NC_003384_Salmonella:4.6253,(NC_009981_Salmonella:4.1322,NC_013365_Escherichi
a:3.8994)Branch 21:4.3478)Branch 29:7.6851)Branch 104:8.5317)Branch
147:11.9751,((NC_014107_Enterobacter:10.7021,(NC_005249_Klebsiella:3.1655,NC_006625_Klebsiella:3.4216)Branch
33:9.362)Branch
169:11.2882,(NC_011964_Escherichia:10.5319,(NC_011092_Salmonella:6.3402,(NC_006856_Salmonella:9.6369,NC_010119_Salmo
nella:7.8232)Branch 138:9.1183)Branch 150:9.3708)Branch 191:10.5589)Branch 212:11.227)Branch
227:12.5224,(NC_011754_Escherichia:9.4561,((((NC_013369_Escherichia:4.4638,NC_013728_Escherichia:7.473)Branch
121:8.4977,(NC_009788_Escherichia:7.1842,((NC_013120_Escherichia:5.5376,(NC_005014_Salmonella:6.6907,(NC_011419_Escheri
chia:5.4586,(NC_011077_Salmonella:5.4948,NC_011081_Salmonella:3.9667)Branch 76:5.8797)Branch 101:6.0194)Branch
111:6.9886)Branch 122:7.1749,(NC_007365_Escherichia:7.153,NC_012487_Escherichia:8.059)Branch 125:8.3752)Branch
```

```
140:8.6854)Branch 161:8.6139)Branch 176:9.7195,((NC_013354_Escherichia:6.0698,NC_013942_Escherichia:8.0772)Branch
158:8.8626,(((NC_007208_Salmonella:4.8793,(NC_003277_Salmonella:4.9659,NC_013437_Salmonella:6.4228)Branch
85:7.4691)Branch
119:8.3405,(NC_009649_Klebsiella:8.5239,(NC_013950_Klebsiella:6.5055,(NC_014312_Klebsiella:7.8148,(NC_009650_Klebsiella:4.1
506,NC_014016_Klebsiella:4.1743)Branch 56:6.633)Branch 107:8.2455)Branch 126:8.2698)Branch 145:9.3159)Branch
164:9.7176,(NC_011413_Escherichia:8.0471,(NC_006816_Salmonella:9.7589,((NC_011350_Escherichia:2.9228,NC_013010_Escheric
hia:3.2621)Branch 11:4.8406,(NC_002128_Escherichia:3.4338,NC_007414_Escherichia:3.5509)Branch 38:3.9432)Branch
49:9.4091)Branch 153:10.1418)Branch 166:9.9459)Branch 188:10.5312)Branch 215:9.8467)Branch
222:10.2229,((NC_007635_Escherichia:7.8876,NC_014234_Escherichia:5.7084)Branch
168:8.3538,(((NC_007608_Shigella:5.4955,(NC_007385_Shigella:6.4358,(NC_010660_Shigella:6.0153,(NC_007607_Shigella:5.1588,(
NC_002698_Shigella:5.259,NC_004851_Shigella:4.519)Branch 17:6.9688)Branch 67:6.6285)Branch 78:6.7495)Branch
89:6.9748)Branch 98:10.32,((NC_009786_Escherichia:4.7439,NC_014232_Escherichia:8.1505)Branch
129:8.7334,(NC_011752_Escherichia:8.1696,((((NC_010862_Escherichia:3.6392,NC_011603_Escherichia:2.865)Branch
15:7.829,((NC_009602_Escherichia:6.5557,NC_012944_Escherichia:7.3055)Branch
77:8.1794,(((NC_004998_Escherichia:8.7944,(NC_010488_Escherichia:8.8474,(NC_011076_Salmonella:7.6909,(NC_011747_Escheric
hia:5.8925,(NC_009837_Escherichia:7.206,NC_010409_Escherichia:6.3419)Branch 14:7.1308)Branch 20:7.017)Branch
35:8.1242)Branch 51:9.4407)Branch
71:9.6212,(NC_010558_Escherichia:9.8566,(NC_013951_Klebsiella:9.3273,(NC_008460_Escherichia:7.2527,(NC_013175_Escherichia
:5.0102,(NC_007941_Escherichia:5.4418,NC_011749_Escherichia:5.701)Branch 4:6.2543)Branch 7:7.4798)Branch 34:8.4987)Branch
58:9.5871)Branch 72:10.2239)Branch
79:10.7992,(NC_010720_Escherichia:7.1026,((NC_013362_Escherichia:5.1576,(NC_011812_Escherichia:5.9652,(NC_013542_Klebsiel
la:3.8726,NC_013727_Shigella:3.5149)Branch 13:4.848)Branch 37:6.3044)Branch
57:6.3653,(NC_006671_Escherichia:8.7785,(NC_005327_Escherichia:6.7121,NC_009133_Escherichia:8.3038)Branch
48:8.3398)Branch 65:8.8166)Branch 70:8.482)Branch 88:8.2939)Branch 99:9.854)Branch 103:9.6662)Branch
115:10.1746,(NC_011980_Escherichia:8.7345,(NC_007675_Escherichia:8.8747,NC_010719_Escherichia:10.3405)Branch
91:11.0712)Branch 120:11.1643)Branch
144:11.1958,(NC_013366_Escherichia:7.8559,(NC_013507_Escherichia:8.0438,(NC_013121_Escherichia:5.7293,NC_013122_Escheri
chia:8.5016)Branch 75:9.5619)Branch 123:9.9067)Branch 148:9.2444)Branch 154:10.9745)Branch 170:9.8755)Branch
181:9.588)Branch
190:10.7015,(NC_002142_Escherichia:7.5808,(NC_009790_Escherichia:7.3976,NC_014233_Escherichia:6.5777)Branch
177:7.9212)Branch 211:7.8893)Branch 219:9.4609)Branch 226:9.2446)Branch 237:10.0373)Branch 252:9.9902)Branch
267:11.6449)Branch 286:10.113)Branch 311:9.0473)Branch 320:9.1011)Branch 349:7.5874)Branch 358:7.8615)Branch
384:7.7467)Branch
396:8.3681,((NC_009427_Novosphingobium:7.4598,(NC_009508_Sphingomonas:6.3764,(NC_008308_Sphingomonas:6.6444,(NC_00
2033_Novosphingobium:1.0738,NC_009426_Novosphingobium:1.0627)Branch 113:5.9252)Branch 288:6.8402)Branch
337:7.0804)Branch
382:7.4317,(NC_008757_Polaromonas:6.7041,((NC_008275_Pseudomonas:4.2933,(NC_014124_Pseudomonas:4.7505,(NC_004444_
Pseudomonas:1.1629,NC_011838_Pseudomonas:1.3027)Branch 80:6.3854)Branch 224:6.2736)Branch
259:6.2332,(NC_005244_Pseudomonas:5.47,(NC_003350_Pseudomonas:5.6302,(NC_007926_Pseudomonas:2.4467,(NC_004999_Ps
eudomonas:0.45708,NC_012674_Pseudomonas:0.49396)Branch 108:3.2016)Branch 175:4.528)Branch 254:5.9467)Branch
310:5.8574)Branch 324:6.6751)Branch 389:6.7921)Branch 401:7.152)Branch
405:1.2029,((NC_007950_Polaromonas:6.696,NC_010333_Caulobacter:5.2517)Branch
364:6.4131,((NC_009507_Sphingomonas:6.4698,NC_014007_Sphingobium:5.6952)Branch
351:6.6821,(NC_012855_Ralstonia:7.1909,((NC_010580_Beijerinckia:5.619,(NC_014035_Rhodobacter:4.7057,(NC_011370_Rhizobiu
m:7.0802,(NC_008243_Chelativorans:4.7885,(NC_007488_Rhodobacter:2.4099,NC_011962_Rhodobacter:2.944)Branch
174:5.2706)Branch 304:5.2649)Branch 330:6.3592)Branch 347:6.0322)Branch
400:6.1277,(((NC_006672_Gluconobacter:5.18,(NC_013212_Acetobacter:3.4553,(NC_013210_Acetobacter:4.4972,NC_013211_Acet
obacter:4.9254)Branch 302:5.2577)Branch 366:4.503)Branch
378:5.168,(((NC_007274_Pseudomonas:5.105,NC_013370_Escherichia:4.8041)Branch
376:5.0646,(NC_007949_Polaromonas:6.9991,NC_013860_Azospirillum:6.3107)Branch 388:6.6942)Branch
394:6.0808,((NC_004574_Ruegeria:5.1523,(NC_009955_Dinoroseobacter:4.4509,NC_009957_Dinoroseobacter:3.2233)Branch
198:5.9211)Branch
326:6.294,((NC_007959_Nitrobacter:6.3868,(NC_011986_Agrobacterium:4.8753,NC_013545_Sinorhizobium:5.9044)Branch
348:5.6857)Branch 373:6.1496,((NC_011758_Methylobacterium:7.0506,NC_012987_Methylobacterium:4.16)Branch
317:6.2169,(NC_002682_Mesorhizobium:6.199,((NC_009671_Ochrobactrum:5.1261,(NC_009622_Sinorhizobium:6.3658,NC_009669
_Ochrobactrum:5.9803)Branch 305:6.4982)Branch
334:6.0849,((NC_008382_Rhizobium:6.2918,NC_011892_Methylobacterium:8.2123)Branch
318:7.6152,((NC_010865_Sinorhizobium:6.4738,((NC_005873_Oligotropha:4.3329,NC_007961_Nitrobacter:4.7804)Branch
```

```
249:4.6937,(NC_011991_Agrobacterium:4.3893,(NC_008383_Rhizobium:5.282,((NC_007762_Rhizobium:6.0212,(NC_002575_Agrob
acterium:3.7934,NC_010841_Agrobacterium:3.308)Branch 136:6.0634)Branch
209:6.3629,(NC_011984_Agrobacterium:6.1333,((NC_011990_Agrobacterium:4.5324,(NC_002147_Agrobacterium:3.528,NC_00306
5_Agrobacterium:4.1387)Branch 128:5.2316)Branch
159:6.4201,(NC_011982_Agrobacterium:6.3,(NC_002377_Agrobacterium:4.1577,NC_010929_Agrobacterium:5.7711)Branch
163:5.8208)Branch 189:6.6055)Branch 208:6.6322)Branch 216:6.6312)Branch 228:6.7353)Branch 236:6.3396)Branch
250:5.5031)Branch 262:5.3305)Branch
278:6.6817,(NC_002679_Mesorhizobium:7.704,(NC_009670_Ochrobactrum:5.1466,(NC_007960_Nitrobacter:5.1379,(NC_009475_B
radyrhizobium:5.7401,(NC_008242_Chelativorans:7.2479,NC_009717_Xanthobacter:6.9068)Branch 251:7.3846)Branch
273:6.8233)Branch 284:6.3115)Branch 303:5.7563)Branch 306:6.9627)Branch 327:7.0894)Branch 352:7.4554)Branch
361:6.8516)Branch 368:6.6597)Branch 379:6.5093)Branch 386:6.4673)Branch 402:6.5002)Branch 416:6.3416)Branch
422:5.9145,((NC_009430_Rhodobacter:6.1071,(NC_011960_Rhodobacter:3.0148,(NC_009007_Rhodobacter:1.9933,NC_009040_Rh
odobacter:2.2186)Branch 187:2.9828)Branch 225:4.9066)Branch
375:6.0309,(((NC_003240_Nostoc:3.8015,NC_007412_Anabaena:4.7549)Branch
271:5.7426,((NC_003276_Nostoc:4.8223,NC_007410_Anabaena:4.8787)Branch
221:6.4173,(NC_010631_Nostoc:4.7855,NC_010632_Nostoc:4.4032)Branch 290:4.8567)Branch 312:6.5374)Branch
417:6.4108,(((NC_009780_Cronobacter:2.5685,NC_013283_Cronobacter:2.8392)Branch
195:5.7185,((NC_013285_Cronobacter:5.8149,(NC_002523_Serratia:5.7511,(((NC_011416_Escherichia:5.5226,NC_013717_Citrobact
er:5.7558)Branch
268:6.6434,(NC_011281_Klebsiella:6.6489,((NC_006855_Salmonella:1.8589,NC_012124_Salmonella:2.1115)Branch
114:5.503,(NC_010422_Salmonella:3.3073,NC_011204_Salmonella:2.7586)Branch 131:5.8284)Branch 206:7.4723)Branch
301:7.4805)Branch
341:7.3615,((NC_003385_Salmonella:5.0404,(NC_014022_Yersinia:4.2114,(NC_010158_Yersinia:4.2125,((NC_006323_Yersinia:2.957
8,NC_009378_Yersinia:2.8948)Branch
9:6.4135,(NC_004835_Yersinia:2.7344,((NC_008118_Yersinia:2.1404,NC_008120_Yersinia:1.8268)Branch
25:2.4646,((NC_003134_Yersinia:1.6802,NC_009596_Yersinia:1.562)Branch
16:2.1633,(NC_004838_Yersinia:2.1289,NC_005815_Yersinia:2.2318)Branch 19:2.5398)Branch 26:2.8302)Branch 40:2.9232)Branch
46:4.2753)Branch 81:6.0372)Branch 97:5.5944)Branch 106:7.8526)Branch
171:9.9141,((NC_009350_Aeromonas:6.6936,((NC_014017_Yersinia:3.4875,((NC_010635_Yersinia:2.9018,(NC_006153_Yersinia:2.70
54,NC_010157_Yersinia:2.4299)Branch 74:2.8096)Branch
82:3.4816,((NC_008122_Yersinia:1.2018,NC_009377_Yersinia:1.1406)Branch
32:3.3405,((NC_009595_Yersinia:1.7603,(NC_003131_Yersinia:1.7744,NC_005813_Yersinia:1.9964)Branch 52:1.9676)Branch
54:2.2672,(NC_004836_Yersinia:1.6381,NC_004839_Yersinia:1.7603)Branch 55:2.0737)Branch 60:3.0228)Branch 83:3.7782)Branch
94:4.6741)Branch
116:5.308,(NC_002120_Yersinia:4.043,(NC_008791_Yersinia:1.9281,(NC_004564_Yersinia:1.5737,NC_005017_Yersinia:1.4248)Branc
h 84:1.8178)Branch 90:3.8127)Branch 133:4.4703)Branch 142:8.0217)Branch
256:8.5642,(NC_011282_Klebsiella:8.1682,(NC_009425_Enterobacter:6.3217,NC_014305_Erwinia:6.3419)Branch
282:6.7502)Branch 309:7.8035)Branch 340:8.4155)Branch 362:9.264)Branch 370:8.4974)Branch 385:7.2835)Branch
406:6.6721,(NC_014108_Enterobacter:4.1345,((NC_005246_Erwinia:2.8553,(NC_004464_Citrobacter:5.0937,NC_011641_Klebsiella:
3.2646)Branch 184:4.9146)Branch 214:6.1935,(NC_013973_Erwinia:5.1063,NC_014304_Erwinia:4.5602)Branch 367:4.9873)Branch
381:5.796)Branch 409:5.0422)Branch 418:6.0786)Branch
442:5.9682,((NC_005838_Thermus:2.5672,NC_006462_Thermus:3.0179)Branch
200:5.7643,(((NC_009932_Acaryochloris:4.3574,NC_009933_Acaryochloris:3.6069)Branch
313:4.6136,(NC_009931_Acaryochloris:4.7073,((NC_009927_Acaryochloris:7.049,NC_009930_Acaryochloris:4.7097)Branch
297:6.3315,(NC_009929_Acaryochloris:5.5474,(NC_009926_Acaryochloris:6.2167,NC_009928_Acaryochloris:5.6281)Branch
274:6.6353)Branch 319:6.3196)Branch 345:6.4318)Branch 357:5.894)Branch
395:5.7568,(NC_010335_Caulobacter:5.6193,(((NC_006823_Aromatoleum:5.6191,(NC_008758_Polaromonas:5.0362,NC_008760_P
olaromonas:4.8053)Branch 392:5.3078)Branch
436:5.5889,((NC_008738_Marinobacter:4.9642,((NC_011143_Phenylobacterium:6.6354,(NC_010070_Burkholderia:1.2388,NC_0108
02_Burkholderia:1.1856)Branch 134:6.1502)Branch
398:6.6346,(NC_009469_Acidiphilium:4.3022,(NC_009467_Acidiphilium:5.4751,NC_009468_Acidiphilium:5.2758)Branch
377:5.5836)Branch 404:5.1045)Branch 423:5.9621)Branch
429:5.5723,((NC_012718_Burkholderia:4.4607,NC_012723_Burkholderia:4.2253)Branch
354:5.0465,(NC_009228_Burkholderia:4.6502,NC_009229_Burkholderia:5.4458)Branch 403:5.3382)Branch 437:5.2442)Branch
445:5.4838)Branch
456:5.574,((NC_008765_Acidovorax:3.6353,(NC_011003_Burkholderia:4.0458,(NC_008545_Burkholderia:5.6562,(NC_012720_Burkh
olderia:4.0693,NC_012725_Burkholderia:3.5231)Branch 277:4.2969)Branch 308:5.1914)Branch 331:5.2018)Branch
399:4.9017,((NC_005863_Desulfovibrio:1.365,NC_008741_Desulfovibrio:1.3456)Branch
```

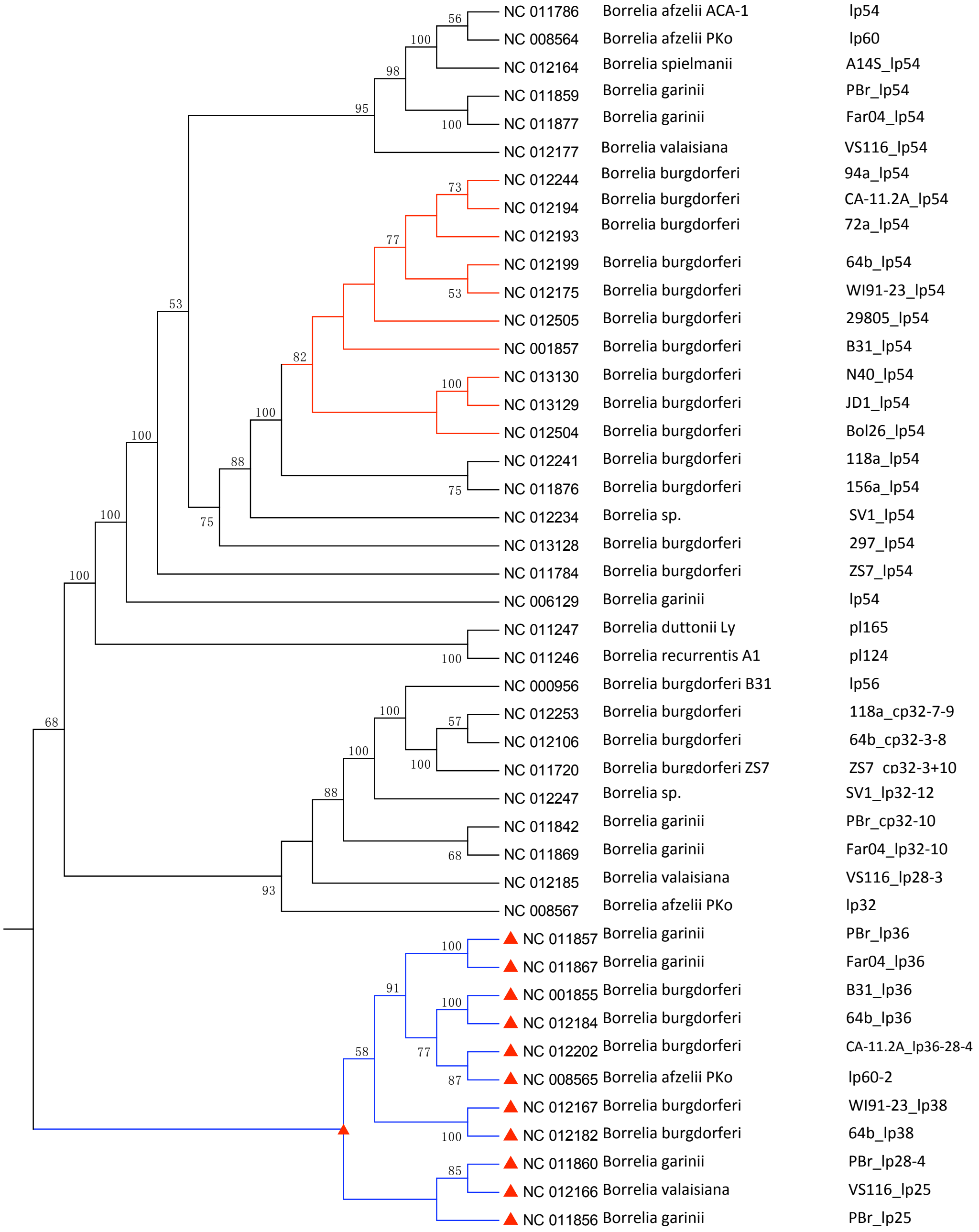
```
167:5.2353,((NC_007901_Rhodoferax:5.2236,NC_009705_Yersinia:4.2595)Branch
446:4.938,(((NC_006365_Legionella:4.5043,(NC_006366_Legionella:2.6077,NC_009966_Fluoribacter:2.7375)Branch
335:3.282)Branch 438:4.0772,((NC_007713_Sodalis:2.05,(NC_007182_Sodalis:0.0033478,NC_007183_Sodalis:0.0034429)Branch
137:2.45)Branch 220:4.7962,(((NC_009701_Vibrio:1.8637,NC_009703_Vibrio:1.5312)Branch
204:3.3279,(NC_005128_Vibrio:2.3186,NC_009702_Vibrio:2.7781)Branch 243:3.4487)Branch
307:4.1329,((NC_011665_Shewanella:3.0362,(NC_011668_Shewanella:2.8387,(NC_009998_Shewanella:2.8862,(NC_011664_Shewa
nella:2.8797,(NC_009661_Shewanella:1.9595,NC_009999_Shewanella:1.7428)Branch 156:2.7961)Branch 186:3.9251)Branch
234:3.5509)Branch 240:3.3866)Branch
248:4.3295,(NC_005871_Photobacterium:2.7981,(NC_010614_Vibrio:3.1513,NC_011311_Aliivibrio:2.9228)Branch 291:3.299)Branch
315:3.2289)Branch 342:4.2525)Branch 413:4.5528)Branch 463:4.7154)Branch
469:4.4516,(((NC_011759_Yersinia:3.9671,NC_012752_Candidatus:2.8295)Branch
369:3.9867,(((NC_004632_Pseudomonas:3.1141,NC_004633_Pseudomonas:3.3601)Branch
255:4.8808,(NC_010377_Yersinia:5.9597,(NC_004349_Shewanella:5.1904,NC_009036_Shewanella:3.5173)Branch
343:4.8598)Branch 408:5.5281)Branch
421:5.3518,(NC_012797_Desulfovibrio:3.4516,(NC_008607_Pelobacter:5.1022,(NC_011667_Thauera:3.8739,NC_011880_Cyanothec
e:4.3884)Branch 380:4.2382)Branch 397:4.9408)Branch 426:4.2567)Branch 431:5.0039)Branch
462:4.6061,((NC_011885_Cyanothece:4.4738,(NC_011737_Cyanothece:3.9702,NC_011738_Cyanothece:4.8393)Branch
410:4.8709)Branch
475:4.7236,((NC_010679_Burkholderia:4.4949,(NC_013509_Edwardsiella:4.762,((NC_013852_Allochromatium:3.557,(NC_010683_R
alstonia:1.3922,NC_012851_Ralstonia:1.1348)Branch 194:4.0019)Branch
449:3.8221,(NC_013930_Thioalkalivibrio:5.3691,(NC_011409_Haemophilus:4.8699,(NC_005918_Pseudomonas:3.1561,(NC_005205
_Pseudomonas:3.7759,NC_007275_Pseudomonas:3.6258)Branch 316:3.9659)Branch 355:4.1842)Branch 440:4.5955)Branch
450:5.0353)Branch 457:4.5726)Branch 477:4.6946)Branch
480:4.6138,(NC_011185_Vibrio:4.495,(((((NC_006129_Borrelia:2.7634,(NC_011784_Borrelia:3.4426,((NC_013128_Borrelia:2.7998,(
NC_012234_Borrelia:3.5454,(((NC_012504_Borrelia:1.091,(NC_013129_Borrelia:0.3478,NC_013130_Borrelia:0.37214)Branch
23:1.108)Branch
41:1.2282,(NC_001857_Borrelia:1.3636,(NC_012505_Borrelia:1.163,((NC_012175_Borrelia:0.67003,NC_012199_Borrelia:0.62593)Br
anch 24:0.89245,(NC_012193_Borrelia:0.79193,(NC_012194_Borrelia:0.49259,NC_012244_Borrelia:0.52013)Branch
22:0.84946)Branch 27:0.93933)Branch 31:1.1886)Branch 39:1.3409)Branch 42:1.432)Branch
43:1.8358,(NC_011876_Borrelia:1.2827,NC_012241_Borrelia:1.1888)Branch 44:1.7172)Branch 53:2.7624)Branch 73:3.8277)Branch
92:3.8728,(NC_012177_Borrelia:1.9757,((NC_011859_Borrelia:1.022,NC_011877_Borrelia:0.86498)Branch
63:1.8246,(NC_012164_Borrelia:1.1533,(NC_008564_Borrelia:1.1539,NC_011786_Borrelia:1.2167)Branch 62:1.4022)Branch
66:1.9811)Branch 87:2.4068)Branch 96:2.5645)Branch 100:3.5391)Branch 109:4.6343)Branch
135:7.6386,(NC_011246_Borrelia:1.5867,NC_011247_Borrelia:2.0818)Branch 165:4.368)Branch
285:7.1285,(NC_008567_Borrelia:2.8518,(NC_012185_Borrelia:4.2593,((NC_012247_Borrelia:2.6543,(NC_000956_Borrelia:2.4732,(
NC_011720_Borrelia:1.1789,(NC_012106_Borrelia:1.2599,NC_012253_Borrelia:1.1327)Branch 143:1.3328)Branch
146:2.2129)Branch 172:3.1375)Branch 201:4.2275,(NC_011842_Borrelia:2.5473,NC_011869_Borrelia:2.8879)Branch
260:2.9635)Branch 279:4.011)Branch 300:4.3556)Branch 328:4.0584)Branch
393:6.0715,(((NC_012167_Borrelia:1.0117,NC_012182_Borrelia:1.0837)Branch
207:3.1284,((NC_011857_Borrelia:2.0963,NC_011867_Borrelia:1.9359)Branch
269:2.7126,((NC_001855_Borrelia:1.3428,NC_012184_Borrelia:1.2867)Branch
217:2.6321,(NC_008565_Borrelia:2.5034,NC_012202_Borrelia:3.0564)Branch 276:3.2406)Branch 299:3.1447)Branch
329:3.4173)Branch 411:3.4915,(NC_011856_Borrelia:2.5976,(NC_011860_Borrelia:2.8669,NC_012166_Borrelia:2.2661)Branch
336:2.9691)Branch 419:2.9406)Branch 444:3.2478)Branch
453:4.8758,((NC_013731_Spirosoma:3.8559,NC_013732_Spirosoma:3.4607)Branch
383:4.2734,(NC_008759_Polaromonas:4.1698,(((NC_008761_Polaromonas:2.9009,NC_010815_Geobacter:3.6546)Branch
443:3.5138,(NC_008386_Roseobacter:4.1943,NC_010373_Methylobacterium:3.4083)Branch 466:3.9558)Branch
492:3.7487,(((NC_012970_Methylovorus:3.189,(NC_006139_Desulfotalea:3.7595,(NC_009704_Yersinia:3.416,(NC_005793_Achrom
obacter:3.7532,(NC_012919_Photobacterium:4.7058,(NC_006824_Aromatoleum:5.5991,(NC_013193_Candidatus:5.442,(NC_00835
7_Pseudomonas:6.9516,((NC_010935_Comamonas:4.391,(NC_006830_Achromobacter:3.8639,(NC_001735_Enterobacter:3.7849,N
C_008766_Acidovorax:2.9306)Branch 192:3.5999)Branch 197:3.8503)Branch
203:4.7392,(NC_013666_Burkholderia:3.6807,(NC_013176_Pseudomonas:3.4935,((NC_005912_Ralstonia:1.5964,NC_007337_Ralsto
nia:1.6034)Branch 112:3.8267,(NC_004956_Pseudomonas:3.1011,NC_005088_Delftia:3.1519)Branch 149:3.7329)Branch
162:4.4971)Branch 183:4.769)Branch 213:4.6272)Branch 233:4.9944)Branch 245:6.5079)Branch 280:6.0858)Branch
283:5.9157)Branch 287:5.4901)Branch 294:4.947)Branch 333:5.0542)Branch 447:4.5209)Branch
464:4.0316,(NC_009777_Vibrio:4.0789,(NC_007110_Rickettsia:2.8273,NC_011061_Prosthecochloris:3.0561)Branch
441:3.1155)Branch 481:3.6613)Branch 493:3.8758,(((NC_009973_Herpetosiphon:4.7423,NC_009974_Herpetosiphon:2.552)Branch
467:3.7883,(((NC_010474_Synechococcus:4.4589,NC_010480_Synechococcus:3.5437)Branch
```

```
461:4.1366,(NC_005230_Synechocystis:3.5797,(NC_005229_Synechocystis:3.5557,NC_005232_Synechocystis:2.7844)Branch
372:3.6761)Branch 468:3.7023)Branch
479:3.9922,((NC_007489_Rhodobacter:3.4666,((NC_007490_Rhodobacter:3.9023,NC_009431_Rhodobacter:3.9828)Branch
387:4.312,(NC_009956_Dinoroseobacter:4.6736,NC_009958_Dinoroseobacter:4.222)Branch 415:4.6501)Branch 435:4.5744)Branch
452:4.2442,((NC_002490_Xylella:3.4505,(NC_014144_Thiomonas:3.6917,(NC_003922_Xanthomonas:3.0274,NC_007507_Xanthomo
nas:4.2586)Branch 323:4.2811)Branch 412:4.1254)Branch
425:3.9713,(NC_008342_Nitrosomonas:3.123,NC_010509_Methylobacterium:2.4009)Branch 428:2.8636)Branch 454:3.5979)Branch
487:3.9514)Branch 491:3.9901)Branch
494:3.9119,(((NC_010555_Proteus:2.9048,((NC_010697_Erwinia:2.6367,NC_011351_Escherichia:2.8085)Branch
371:3.0219,(NC_002525_Escherichia:4.7332,((NC_010657_Shigella:2.016,NC_013718_Citrobacter:2.6291)Branch
238:3.6462,(NC_010860_Salmonella:2.7139,(NC_010378_Escherichia:4.1957,NC_013503_Escherichia:3.4034)Branch
230:4.0396)Branch 239:4.8151)Branch 359:4.4944)Branch 391:4.8793)Branch 424:4.2395)Branch
472:3.6851,(NC_005250_Listonella:3.3197,NC_010404_Acinetobacter:3.9342)Branch 489:3.6749)Branch
497:3.7037,((NC_010844_Leptospira:0.33435,NC_010846_Leptospira:0.31078)Branch
185:3.3892,((NC_008014_Lawsonia:3.2009,(NC_007801_Jannaschia:3.0147,((NC_008042_Ruegeria:3.7597,(NC_004929_Ruegeria:2.
5321,NC_008387_Roseobacter:2.3768)Branch 360:2.6528)Branch
390:3.4834,(NC_012983_Hirschia:2.7461,NC_014157_Salinibacter:2.4511)Branch 427:2.648)Branch 433:3.2163)Branch
473:3.1983)Branch 500:3.2542,(((NC_003273_Nostoc:2.6899,(NC_003267_Nostoc:3.0309,NC_010630_Nostoc:2.9424)Branch
365:3.7371)Branch
496:3.2709,(NC_009959_Dinoroseobacter:2.9426,(NC_004073_Synechococcus:0.56322,NC_007595_Synechococcus:0.48291)Branch
193:3.2239)Branch 504:3.1169)Branch
507:3.2055,(NC_009035_Shewanella:3.237,((NC_009226_Burkholderia:3.2187,NC_010606_Acinetobacter:2.7254)Branch
486:3.0776,(NC_010113_Vibrio:3.0508,(NC_013502_Rhodothermus:2.9697,(((NC_006842_Vibrio:2.6843,NC_010112_Vibrio:2.4831)
Branch 432:2.8069,((NC_005012_Campylobacter:0.63315,NC_008770_Campylobacter:0.65333)Branch
210:2.8601,(NC_006134_Campylobacter:0.70748,NC_008790_Campylobacter:0.71846)Branch 218:2.8453)Branch
503:2.9103)Branch 516:2.8684,(NC_002806_Microscilla:2.8013,(((NC_009726_Coxiella:1.5207,NC_010258_Coxiella:1.4967)Branch
266:2.9512,(NC_010124_Gluconacetobacter:2.4022,NC_013958_Nitrosococcus:2.7192)Branch 474:2.6431)Branch
510:2.825,((NC_009829_Serratia:2.3431,NC_012109_Desulfobacterium:2.7416)Branch
502:2.604,(NC_011893_Methylobacterium:2.555,(NC_011148_Salmonella:2.568,(NC_012439_Persephonella:2.5028,(NC_013160_C
yanothece:2.4921,(NC_005915_Caedibacter:2.3508,(NC_014249_Nostoc:2.3097,(NC_013518_Sebaldella:2.1713,NC_013784_Zymo
monas:2.2454)Branch 506:2.213)Branch 512:2.2665)Branch 513:2.3136)Branch 522:2.4043)Branch 523:2.4554)Branch
524:5.1586)Branch 525:2.737)Branch 521:2.8596)Branch 520:2.886)Branch 519:2.9653)Branch 518:3.055)Branch 517:3.1259)Branch
515:3.1905)Branch 514:3.2954)Branch 511:3.3374)Branch 509:3.4563)Branch 508:3.6436)Branch 505:3.8675)Branch
501:3.9664)Branch 499:3.9842)Branch 498:4.0863)Branch 495:4.3833)Branch 490:4.5302)Branch 488:4.7573)Branch
485:4.5927)Branch 484:4.6865)Branch 483:4.7374)Branch 482:4.7424)Branch 478:4.8148)Branch 476:5.1517)Branch
471:5.3201)Branch 470:5.3527)Branch 465:5.7205)Branch 460:5.7624)Branch 459:5.8899)Branch 458:6.0005)Branch
455:6.1903)Branch 451:6.352)Branch 448:6.2403)Branch 439:6.3418)Branch 434:6.7053)Branch 430:7.2353)Branch
420:6.9763)Branch 414:7.1678)Branch 407:6.5822)Root;
```

Appendix D

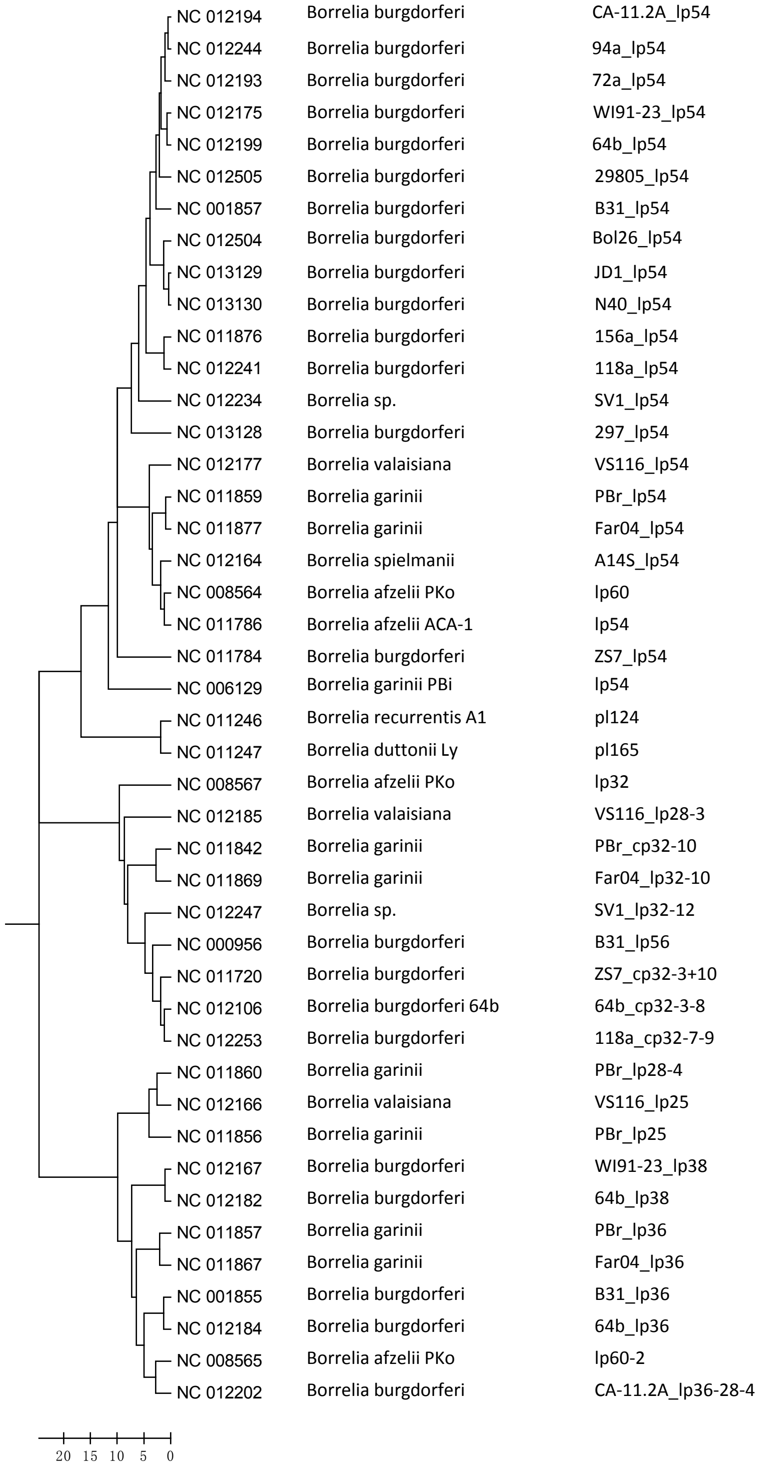
Fig. D.1. *Borrelia* sub-group, majority consensus tree.

Majority Voting consensus tree:



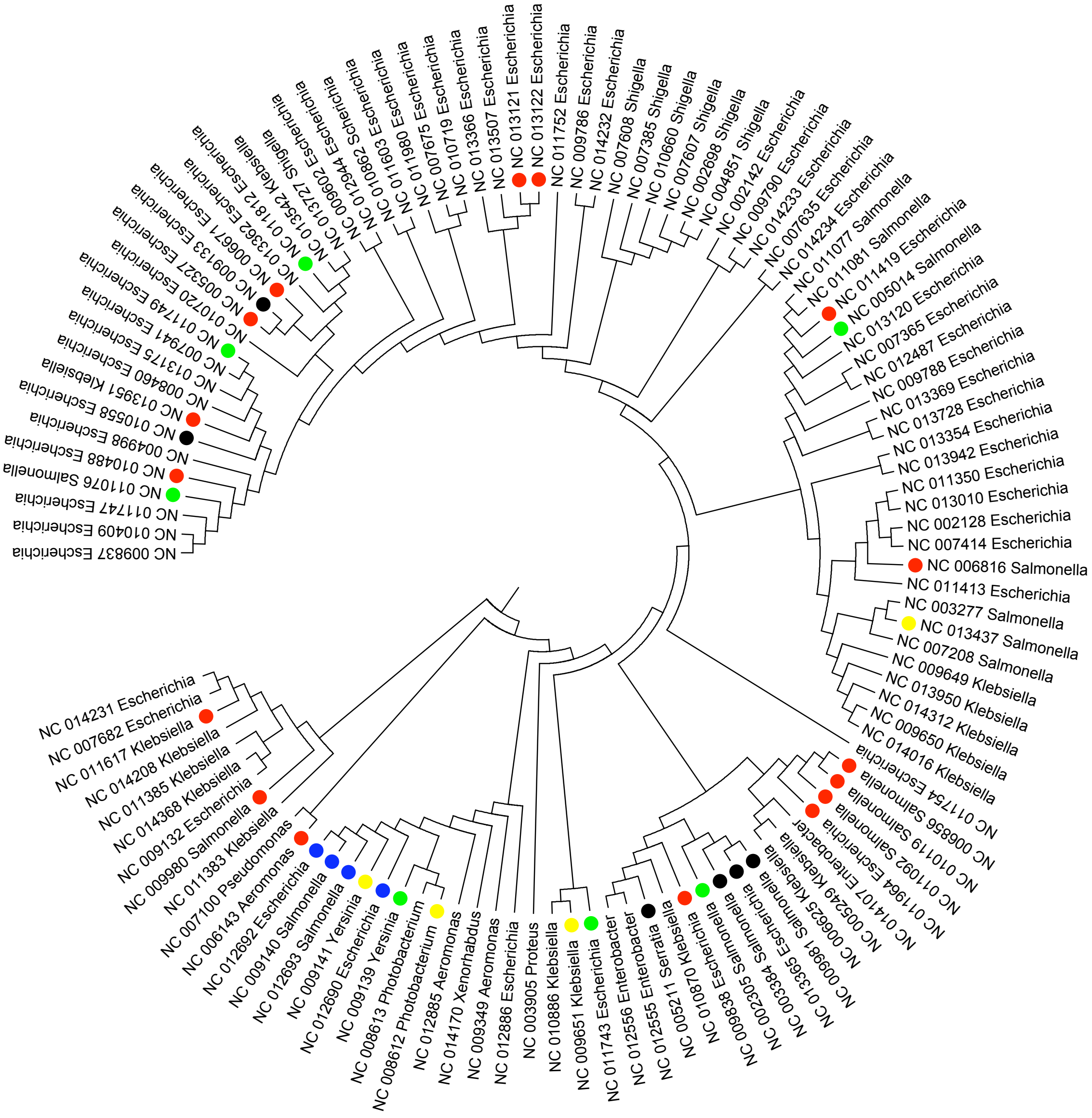
Appendix E

Fig. E.1. *Borrelia* sub-group, averaged distance consensus tree.


NC 012194 Borrelia burgdorferi CA-11.2A_lp54
NC 012244 Borrelia burgdorferi 94a_lp54
NC 012193 Borrelia burgdorferi 72a_lp54
NC 012175 Borrelia burgdorferi WI91-23_lp54
NC 012199 Borrelia burgdorferi 64b_lp54
NC 012505 Borrelia burgdorferi 29805_lp54
NC 001857 Borrelia burgdorferi B31_lp54
NC 012504 Borrelia burgdorferi Bol26_lp54
NC 013129 Borrelia burgdorferi JD1_lp54
NC 013130 Borrelia burgdorferi N40_lp54
NC 011876 Borrelia burgdorferi 156a_lp54
NC 012241 Borrelia burgdorferi 118a_lp54
NC 012234 Borrelia sp. SV1_lp54
NC 013128 Borrelia burgdorferi 297_lp54
NC 012177 Borrelia valaisiana VS116_lp54
NC 011859 Borrelia garinii PBr_lp54
NC 011877 Borrelia garinii Far04_lp54
NC 012164 Borrelia spielmanii A14S_lp54
NC 008564 Borrelia afzelii PKo lp60
NC 011786 Borrelia afzelii ACA-1 lp54
NC 011784 Borrelia burgdorferi ZS7_lp54
NC 006129 Borrelia garinii PBi lp54
NC 011246 Borrelia recurrentis A1 pl124
NC 011247 Borrelia duttonii Ly pl165
NC 008567 Borrelia afzelii PKo lp32
NC 012185 Borrelia valaisiana VS116_lp28-3
NC 011842 Borrelia garinii PBr_cp32-10
NC 011869 Borrelia garinii Far04_lp32-10
NC 012247 Borrelia sp. SV1_lp32-12
NC 000956 Borrelia burgdorferi B31_lp56
NC 011720 Borrelia burgdorferi ZS7_cp32-3+10
NC 012106 Borrelia burgdorferi 64b 64b_cp32-3-8
NC 012253 Borrelia burgdorferi 118a_cp32-7-9
NC 011860 Borrelia garinii PBr_lp28-4
NC 012166 Borrelia valaisiana VS116_lp25
NC 011856 Borrelia garinii PBr_lp25
NC 012167 Borrelia burgdorferi WI91-23_lp38
NC 012182 Borrelia burgdorferi 64b_lp38
NC 011857 Borrelia garinii PBr_lp36
NC 011867 Borrelia garinii Far04_lp36
NC 001855 Borrelia burgdorferi B31_lp36
NC 012184 Borrelia burgdorferi 64b_lp36
NC 008565 Borrelia afzelii PKo lp60-2
NC 012202 Borrelia burgdorferi CA-11.2A_lp36-28-4
20
15
10
5
0

Appendix F

Fig. F.1. 115-plasmid cluster with 37 AR genes.


NC 009837 Escherichia
NC 010409 Escherichia
NC 011747 Escherichia
NC 011076 Salmonella
NC 010488 Escherichia
NC 004998 Escherichia
NC 010558 Escherichia
NC 013951 Klebsiella
NC 008460 Escherichia
NC 013175 Escherichia
NC 007941 Escherichia
NC 011749 Escherichia
NC 010720 Escherichia
NC 005327 Escherichia
NC 009133 Escherichia
NC 006671 Escherichia
NC 013362 Escherichia
NC 011812 Escherichia
NC 013542 Klebsiella
NC 013727 Shigella
NC 009602 Escherichia
NC 012944 Escherichia
NC 010862 Scherichia
NC 011603 Escherichia
NC 011980 Escherichia
NC 007675 Escherichia
NC 010719 Escherichia
NC 013366 Escherichia
NC 013507 Escherichia
NC 013121 Escherichia
NC 013122 Escherichia
NC 011752 Escherichia
NC 009786 Escherichia
NC 014232 Escherichia
NC 007608 Shigella
NC 007385 Shigella
NC 010660 Shigella
NC 007607 Shigella
NC 002698 Shigella
NC 004851 Shigella
NC 002142 Escherichia
NC 009790 Escherichia
NC 014233 Escherichia
NC 007635 Escherichia
NC 014234 Escherichia
NC 011077 Salmonella
NC 011081 Salmonella
NC 011419 Escherichia
NC 005014 Salmonella
NC 013120 Escherichia
NC 007365 Escherichia
NC 012487 Escherichia
NC 009788 Escherichia
NC 013369 Escherichia
NC 013728 Escherichia
NC 013354 Escherichia
NC 013942 Escherichia
NC 011350 Escherichia
NC 013010 Escherichia
NC 002128 Escherichia
NC 007414 Escherichia
NC 006816 Salmonella
NC 011413 Escherichia
NC 003277 Salmonella
NC 013437 Salmonella
NC 007208 Salmonella
NC 009649 Klebsiella
NC 013950 Klebsiella
NC 014312 Klebsiella
NC 009650 Klebsiella
NC 014016 Klebsiella
NC 011754 Escherichia
NC 006856 Salmonella
NC 010119 Salmonella
NC 011092 Salmonella
NC 011964 Escherichia
NC 014107 Enterobacter
NC 005249 Klebsiella
NC 006625 Klebsiella
NC 009981 Salmonella
NC 013365 Escherichia
NC 003384 Salmonella
NC 002305 Salmonella
NC 009838 Escherichia
NC 010870 Klebsiella
NC 005211 Serratia
NC 012555 Enterobacter
NC 012556 Enterobacter
NC 011743 Escherichia
NC 009651 Klebsiella
NC 010886 Klebsiella
NC 003905 Proteus
NC 012886 Escherichia
NC 009349 Aeromonas
NC 014170 Xenorhabdus
NC 012885 Aeromonas
NC 008612 Photobacterium
NC 008613 Photobacterium
NC 009139 Yersinia
NC 012690 Escherichia
NC 009141 Yersinia
NC 012693 Salmonella
NC 009140 Salmonella
NC 012692 Escherichia
NC 006143 Aeromonas
NC 007100 Pseudomonas
NC 011383 Klebsiella
NC 009980 Salmonella
NC 009132 Escherichia
NC 014368 Klebsiella
NC 011385 Klebsiella
NC 014208 Klebsiella
NC 011617 Klebsiella
NC 007682 Escherichia
NC 014231 Escherichia

Appendices G and H are supplementary figures for chapter 3.

Appendix G

Fig. G.1. Newick standard file of Jaccard distance tree for 527 Gram-negative plasmids.

```
((((NC_006365_Legionella:0.39628,(NC_009726_Coxiella:0.049956,NC_010258_Coxiella:0.050044)Branch 17:0.40848)Branch
299:0.49072,(NC_010373_Methylobacterium:0.46896,NC_011893_Methylobacterium:0.47841)Branch 447:0.49143)Branch
501:0.49505,((((NC_009350_Aeromonas:0.3479,((NC_002120_Yersinia:0.15016,(NC_008791_Yersinia:0.054469,(NC_004564_Yersinia:0.022578,NC_005017_Yersinia:
0.022877)Branch 13:0.053935)Branch 34:0.153)Branch
95:0.21576,(NC_014017_Yersinia:0.15614,(((NC_004836_Yersinia:0.036385,NC_004839_Yersinia:0.036079)Branch
25:0.061173,(NC_005813_Yersinia:0.049369,(NC_003131_Yersinia:0.034948,NC_009595_Yersinia:0.035475)Branch 24:0.049929)Branch 38:0.060503)Branch
43:0.12474,((NC_006153_Yersinia:0.080154,(NC_008122_Yersinia:0.025586,NC_009377_Yersinia:0.025696)Branch 18:0.079361)Branch
52:0.10887,(NC_010157_Yersinia:0.094358,NC_010635_Yersinia:0.093142)Branch 61:0.10836)Branch 69:0.12504)Branch 79:0.15519)Branch 98:0.21216)Branch
135:0.3591)Branch 254:0.4812,((NC_003922_Xanthomonas:0.35013,NC_007507_Xanthomonas:0.34432)Branch
251:0.4392,(NC_007274_Pseudomonas:0.42566,NC_013212_Acetobacter:0.42851)Branch 345:0.43812)Branch 365:0.47622)Branch
476:0.48921,((((NC_010509_Methylobacterium:0.42562,(NC_002490_Xylella:0.4121,NC_014144_Thiomonas:0.40841)Branch 313:0.42296)Branch
333:0.46442,(NC_007100_Pseudomonas:0.43578,(NC_006143_Aeromonas:0.36046,NC_013509_Edwardsiella:0.36503)Branch 263:0.44079)Branch
362:0.46365)Branch 426:0.48139,((((NC_006134_Campylobacter:0.083135,NC_008790_Campylobacter:0.083532)Branch
47:0.39004,(NC_005012_Campylobacter:0.25,NC_008770_Campylobacter:0.25)Branch 156:0.39567)Branch
284:0.46688,(NC_006842_Vibrio:0.40701,NC_010112_Vibrio:0.41117)Branch 307:0.46317)Branch
404:0.48795,((NC_007275_Pseudomonas:0.29777,(NC_005205_Pseudomonas:0.25325,NC_005918_Pseudomonas:0.26026)Branch 166:0.29689)Branch
196:0.47493,((NC_007713_Sodalis:0.098334,(NC_007182_Sodalis:0,NC_007183_Sodalis:0)Branch 3:0.096788)Branch
58:0.47055,((((NC_010422_Salmonella:0.11194,NC_011204_Salmonella:0.11306)Branch
74:0.32194,(NC_010378_Escherichia:0.1817,(NC_010860_Salmonella:0.16014,NC_013503_Escherichia:0.15901)Branch 100:0.18349)Branch 117:0.32617)Branch
222:0.39541,(NC_010555_Proteus:0.33273,(NC_010657_Shigella:0.18787,NC_013718_Citrobacter:0.18713)Branch 119:0.32977)Branch 227:0.39748)Branch
293:0.44908,(NC_011148_Salmonella:0.43821,(NC_002525_Escherichia:0.31947,(NC_010697_Erwinia:0.29124,NC_011351_Escherichia:0.29209)Branch
185:0.32095)Branch 216:0.43806)Branch 347:0.45065)Branch 371:0.47071)Branch 425:0.48285)Branch 467:0.48273)Branch 478:0.48763)Branch
483:0.49088,(((NC_010480_Synechococcus:0.40111,(NC_009829_Serratia:0.30419,NC_012109_Desulfobacterium:0.29581)Branch 191:0.40658)Branch
301:0.47424,(NC_009973_Herpetosiphon:0.44496,NC_009974_Herpetosiphon:0.45504)Branch 369:0.47667)Branch
448:0.48766,((NC_007110_Rickettsia:0.43099,(NC_010124_Gluconacetobacter:0.41356,NC_013958_Nitrosococcus:0.41978)Branch 315:0.43508)Branch
344:0.47469,(NC_005250_Listonella:0.44859,NC_012797_Desulfovibrio:0.44029)Branch 372:0.47537)Branch 461:0.48342)Branch 484:0.49089)Branch
502:0.49149)Branch
506:0.4921,(((NC_009226_Burkholderia:0.47455,(NC_010606_Acinetobacter:0.43259,(NC_008765_Acidovorax:0.37755,(NC_011003_Burkholderia:0.33129,(NC_0085
45_Burkholderia:0.2844,(NC_012720_Burkholderia:0.22388,NC_012725_Burkholderia:0.23066)Branch 152:0.28777)Branch 184:0.33074)Branch 232:0.37521)Branch
271:0.42948)Branch 336:0.46989)Branch 429:0.49064,((NC_009036_Shewanella:0.45897,NC_010377_Yersinia:0.45532)Branch
408:0.47629,((((NC_009701_Vibrio:0.10387,NC_009703_Vibrio:0.10665)Branch 64:0.32731,(NC_005128_Vibrio:0.21284,NC_009702_Vibrio:0.21089)Branch
137:0.33117)Branch
228:0.42144,(((NC_009998_Shewanella:0.20872,(NC_011664_Shewanella:0.15738,(NC_009661_Shewanella:0.11794,NC_009999_Shewanella:0.11643)Branch
76:0.158)Branch 105:0.2085)Branch 138:0.25477,(NC_011665_Shewanella:0.23261,NC_011668_Shewanella:0.23677)Branch 154:0.2574)Branch
165:0.34185,(NC_005871_Photobacterium:0.30713,(NC_010614_Vibrio:0.26293,NC_011311_Aliivibrio:0.26207)Branch 171:0.30281)Branch 204:0.34336)Branch
241:0.42172)Branch 325:0.47173,(NC_004349_Shewanella:0.42331,NC_009777_Vibrio:0.42575)Branch 340:0.46394)Branch 438:0.4786)Branch 473:0.48318)Branch
490:0.49265,(((NC_012886_Escherichia:0.42493,(NC_009349_Aeromonas:0.3329,(NC_014170_Xenorhabdus:0.29736,((NC_012690_Escherichia:0.12554,(NC_009140
_Salmonella:0.066798,(NC_012692_Escherichia:0.049062,NC_012693_Salmonella:0.050938)Branch 37:0.067475)Branch 45:0.12625)Branch
85:0.21522,((NC_009141_Yersinia:0.1385,(NC_008612_Photobacterium:0.06952,NC_008613_Photobacterium:0.069208)Branch 44:0.14405)Branch
90:0.16504,(NC_009139_Yersinia:0.14727,NC_012885_Aeromonas:0.14944)Branch 94:0.16665)Branch 107:0.21827)Branch 140:0.29742)Branch 192:0.33594)Branch
235:0.42244)Branch
328:0.48886,(((NC_002305_Salmonella:0.14995,(NC_009981_Salmonella:0.06381,(NC_003384_Salmonella:0.05649,NC_013365_Escherichia:0.058343)Branch
40:0.062671)Branch 42:0.14657)Branch
93:0.29745,(NC_009838_Escherichia:0.20678,(NC_005211_Serratia:0.17857,(NC_010870_Klebsiella:0.16001,(NC_012555_Enterobacter:0.0019149,NC_012556_Enter
obacter:0.0019461)Branch 6:0.15747)Branch 104:0.177)Branch 114:0.20829)Branch 131:0.29383)Branch
188:0.4843,((NC_008573_Shewanella:0.43951,(NC_006466_Cupriavidus:0.037914,NC_007971_Cupriavidus:0.040738)Branch 20:0.44303)Branch
375:0.46446,(NC_013285_Cronobacter:0.44818,((NC_011282_Klebsiella:0.39131,(NC_005249_Klebsiella:0.052442,NC_006625_Klebsiella:0.052453)Branch
39:0.38525)Branch 292:0.43177,(NC_014107_Enterobacter:0.3829,(NC_009425_Enterobacter:0.32292,NC_014305_Erwinia:0.32265)Branch 231:0.38941)Branch
295:0.42832)Branch 349:0.44236)Branch 377:0.46859)Branch 436:0.47758)Branch 469:0.48815)Branch
494:0.49186,((((NC_004464_Citrobacter:0.13942,(NC_005246_Erwinia:0.088728,NC_011641_Klebsiella:0.085185)Branch 54:0.14102)Branch
88:0.4512,(NC_002523_Serratia:0.34611,((NC_004632_Pseudomonas:0.21423,NC_004633_Pseudomonas:0.2113)Branch
142:0.33851,(NC_011759_Yersinia:0.32818,(NC_012752_Candidatus:0.27675,NC_013973_Erwinia:0.27087)Branch 180:0.3254)Branch 230:0.33889)Branch
240:0.34805)Branch 247:0.44997)Branch
382:0.48447,((NC_011281_Klebsiella:0.44797,((NC_007208_Salmonella:0.23875,(NC_003277_Salmonella:0.14533,NC_013437_Salmonella:0.1467)Branch
110:0.23209)Branch 162:0.36377,(NC_010119_Salmonella:0.29419,(NC_006855_Salmonella:0.11136,NC_012124_Salmonella:0.11086)Branch 75:0.30135)Branch
210:0.36355)Branch 268:0.44867)Branch 385:0.47032,((NC_011416_Escherichia:0.41413,NC_013717_Citrobacter:0.41514)Branch
321:0.46962,((((NC_011754_Escherichia:0.32832,((NC_005014_Salmonella:0.22695,(NC_013120_Escherichia:0.1962,(NC_011419_Escherichia:0.18658,(NC_011077_S
almonella:0.094652,NC_011081_Salmonella:0.098626)Branch 68:0.18629)Branch 128:0.1972)Branch 134:0.22772)Branch
158:0.30229,(NC_007365_Escherichia:0.24801,NC_012487_Escherichia:0.24854)Branch 169:0.29942)Branch 212:0.32597)Branch
233:0.41602,((NC_013369_Escherichia:0.19741,NC_013728_Escherichia:0.19009)Branch
133:0.39681,(NC_009788_Escherichia:0.3487,(NC_007635_Escherichia:0.28906,NC_014234_Escherichia:0.2968)Branch 197:0.35178)Branch 255:0.40145)Branch
305:0.41361)Branch 323:0.45098,(NC_006816_Salmonella:0.42298,(NC_006856_Salmonella:0.3171,NC_011092_Salmonella:0.31367)Branch 226:0.41809)Branch
332:0.44807)Branch
388:0.4626,((NC_011964_Escherichia:0.38816,(NC_011980_Escherichia:0.28634,(NC_007675_Escherichia:0.27051,((NC_009837_Escherichia:0.19364,NC_010409_Es
cherichia:0.19098)Branch 144:0.23385,(NC_011076_Salmonella:0.18905,NC_011747_Escherichia:0.19407)Branch 145:0.23202)Branch 164:0.27024)Branch
187:0.27974)Branch 194:0.3851)Branch
291:0.45236,((NC_010719_Escherichia:0.31216,(NC_007608_Shigella:0.29608,(NC_010660_Shigella:0.20634,(NC_007385_Shigella:0.18986,(NC_007607_Shigella:0.1
7044,(NC_002698_Shigella:0.11558,NC_004851_Shigella:0.11834)Branch 81:0.16685)Branch 113:0.18971)Branch 124:0.20525)Branch 136:0.29888)Branch
201:0.32218)Branch
223:0.4442,((((NC_009602_Escherichia:0.28271,(NC_011350_Escherichia:0.14403,(NC_002128_Escherichia:0.11213,(NC_007414_Escherichia:0.095932,NC_013010_
Escherichia:0.092187)Branch 70:0.11123)Branch 82:0.1481)Branch 106:0.28838)Branch
199:0.33348,(NC_013354_Escherichia:0.29361,NC_013942_Escherichia:0.28664)Branch 200:0.33644)Branch
243:0.42576,((NC_009649_Klebsiella:0.34833,(NC_014312_Klebsiella:0.29749,(NC_013950_Klebsiella:0.26075,(NC_009650_Klebsiella:0.10442,NC_014016_Klebsiella
```

```
:0.10247)Branch 84:0.25863)Branch 181:0.30004)Branch 217:0.34696)Branch
259:0.39851,(NC_011413_Escherichia:0.38487,(((NC_004998_Escherichia:0.29358,NC_010488_Escherichia:0.29122)Branch
214:0.33488,((NC_010558_Escherichia:0.30056,(NC_013951_Klebsiella:0.23728,(NC_011749_Escherichia:0.14754,(NC_007941_Escherichia:0.13651,NC_013175_Esc
herichia:0.13621)Branch 111:0.14704)Branch 116:0.23338)Branch 167:0.30116)Branch
218:0.32933,(NC_012944_Escherichia:0.31102,((NC_008460_Escherichia:0.26883,NC_010720_Escherichia:0.27448)Branch
193:0.30151,(NC_006671_Escherichia:0.27806,((NC_005327_Escherichia:0.18926,NC_009133_Escherichia:0.19242)Branch
148:0.23403,(NC_013362_Escherichia:0.21935,(NC_011812_Escherichia:0.18829,(NC_013542_Klebsiella:0.11742,NC_013727_Shigella:0.1197)Branch
92:0.19109)Branch 143:0.21887)Branch 159:0.2394)Branch 168:0.28341)Branch 206:0.30118)Branch 219:0.31225)Branch 229:0.32837)Branch 244:0.33429)Branch
248:0.3751,((NC_013121_Escherichia:0.28558,NC_013122_Escherichia:0.28352)Branch
207:0.36329,(NC_013366_Escherichia:0.3419,NC_013507_Escherichia:0.33895)Branch 249:0.3693)Branch 273:0.37694)Branch 285:0.38472)Branch
294:0.40031)Branch 308:0.42046)Branch 334:0.43749,((NC_002142_Escherichia:0.3012,(NC_010862_scherichia:0.060486,NC_011603_Escherichia:0.064514)Branch
51:0.29175)Branch 209:0.43719,((NC_009790_Escherichia:0.36978,NC_014233_Escherichia:0.36707)Branch
275:0.41548,(NC_011752_Escherichia:0.38772,(NC_009786_Escherichia:0.26458,NC_014232_Escherichia:0.25981)Branch 177:0.39523)Branch 300:0.41502)Branch
326:0.43388)Branch 355:0.43818)Branch 361:0.44277)Branch 370:0.45286)Branch 394:0.46387)Branch 418:0.46747)Branch 434:0.47142)Branch 439:0.48222)Branch
479:0.48992,((((NC_009705_Yersinia:0.43441,NC_012970_Methylovorus:0.43515)Branch
342:0.46197,(NC_009444_Pseudomonas:0.43699,(NC_007901_Rhodoferax:0.42945,(NC_008341_Nitrosomonas:0.37247,NC_008342_Nitrosomonas:0.37753)Branch
280:0.42238)Branch 337:0.43918)Branch 359:0.45668)Branch 403:0.46795,((NC_006366_Legionella:0.29973,NC_009966_Fluoribacter:0.30027)Branch
208:0.45577,(NC_006139_Desulfotalea:0.42259,NC_011061_Prosthecochloris:0.41952)Branch 329:0.45901)Branch 407:0.46396)Branch
427:0.48215,((NC_011383_Klebsiella:0.29208,((NC_009132_Escherichia:0.18138,NC_014368_Klebsiella:0.18112)Branch
122:0.25975,(NC_011385_Klebsiella:0.22837,(NC_009980_Salmonella:0.22149,(NC_014208_Klebsiella:0.20094,(NC_011617_Klebsiella:0.17829,(NC_007682_Escheric
hia:0.13811,NC_014231_Escherichia:0.13775)Branch 91:0.18143)Branch 120:0.19931)Branch 130:0.22315)Branch 151:0.23333)Branch 155:0.26134)Branch
173:0.28583)Branch 186:0.47363,((NC_014108_Enterobacter:0.43645,NC_014304_Erwinia:0.43197)Branch
352:0.4657,((NC_011409_Haemophilus:0.41232,NC_011743_Escherichia:0.40586)Branch
317:0.45747,(NC_003905_Proteus:0.4308,(NC_009651_Klebsiella:0.41753,NC_010886_Klebsiella:0.4158)Branch 330:0.43242)Branch 356:0.45567)Branch
412:0.46161)Branch 428:0.47053)Branch 449:0.48138)Branch 482:0.48692)Branch 499:0.4909)Branch 505:0.49217)Branch 508:0.49357)Branch 510:0.49398)Branch
513:0.25693,(((((NC_013731_Spirosoma:0.32666,NC_013732_Spirosoma:0.32719)Branch
220:0.48468,((NC_008758_Polaromonas:0.4552,NC_011143_Phenylobacterium:0.45313)Branch
399:0.47293,(NC_012811_Methylobacterium:0.46634,(NC_010070_Burkholderia:0.032776,NC_010802_Burkholderia:0.032798)Branch 14:0.46563)Branch
435:0.47307)Branch 456:0.48025)Branch
471:0.49169,((NC_005793_Achromobacter:0.40881,(NC_013193_Candidatus:0.36513,(NC_006824_Aromatoleum:0.35013,(NC_009704_Yersinia:0.32776,(NC_01317
6_Pseudomonas:0.30334,(((NC_006830_Achromobacter:0.22974,((NC_005912_Ralstonia:0.048837,NC_007337_Ralstonia:0.049755)Branch
33:0.21753,(NC_008766_Acidovorax:0.15965,NC_010935_Comamonas:0.15853)Branch 109:0.21598)Branch 149:0.22813)Branch
153:0.26814,(NC_013666_Burkholderia:0.24677,(NC_004956_Pseudomonas:0.20426,(NC_001735_Enterobacter:0.16465,NC_005088_Delftia:0.16371)Branch
112:0.20363)Branch 139:0.24332)Branch 161:0.26811)Branch 175:0.29691,(NC_008357_Pseudomonas:0.27427,NC_012919_Photobacterium:0.27836)Branch
183:0.29319)Branch 195:0.30428)Branch 211:0.32461)Branch 225:0.35573)Branch 258:0.36787)Branch 269:0.39981)Branch
302:0.4835,(NC_008607_Pelobacter:0.4352,NC_011667_Thauera:0.44285)Branch 376:0.47362)Branch 472:0.48888)Branch
503:0.4927,((NC_002806_Microscilla:0.48751,(NC_008761_Polaromonas:0.43731,NC_011185_Vibrio:0.44504)Branch 350:0.48868)Branch
487:0.49355,(((NC_008738_Marinobacter:0.42984,NC_008739_Marinobacter:0.41921)Branch
341:0.47959,((NC_012718_Burkholderia:0.38786,NC_012723_Burkholderia:0.3918)Branch
290:0.47982,(NC_009468_Acidiphilium:0.45546,(NC_009467_Acidiphilium:0.43106,NC_009469_Acidiphilium:0.43005)Branch 360:0.45507)Branch
419:0.47422)Branch 474:0.48149)Branch
488:0.48461,(((NC_008757_Polaromonas:0.43495,(NC_002682_Mesorhizobium:0.42421,NC_014007_Sphingobium:0.42273)Branch 335:0.43873)Branch
364:0.46424,(NC_010333_Caulobacter:0.44358,NC_010335_Caulobacter:0.45297)Branch 389:0.46619)Branch
437:0.4851,((NC_009228_Burkholderia:0.45496,((NC_008275_Pseudomonas:0.37463,(NC_014124_Pseudomonas:0.27909,(NC_004444_Pseudomonas:0.0081286,NC
_011838_Pseudomonas:0.0073753)Branch 5:0.27795)Branch 178:0.37584)Branch
272:0.42576,(NC_005244_Pseudomonas:0.40394,(NC_003350_Pseudomonas:0.28752,(NC_007926_Pseudomonas:0.10427,(NC_004999_Pseudomonas:0,NC_012674
_Pseudomonas:0)Branch 4:0.10305)Branch 63:0.29395)Branch 190:0.40139)Branch 304:0.42474)Branch 331:0.4569)Branch
400:0.47936,(NC_009427_Novosphingobium:0.47007,(NC_009508_Sphingomonas:0.45053,(NC_009507_Sphingomonas:0.41028,(NC_008308_Sphingomonas:0.3416
7,(NC_002033_Novosphingobium:0.018632,NC_009426_Novosphingobium:0.018751)Branch 7:0.35084)Branch 246:0.41379)Branch 316:0.45574)Branch
402:0.46574)Branch 440:0.48135)Branch 475:0.48687)Branch 495:0.48775)Branch 500:0.48827)Branch 504:0.49297)Branch
511:0.4946,(((NC_009956_Dinoroseobacter:0.48522,((NC_008014_Lawsonia:0.47124,(NC_013860_Azospirillum:0.44786,(NC_009958_Dinoroseobacter:0.44033,((NC
_012983_Hirschia:0.34987,NC_014157_Salinibacter:0.35013)Branch
242:0.43741,(NC_007801_Jannaschia:0.37172,(NC_004929_Ruegeria:0.30355,NC_008387_Roseobacter:0.30515)Branch 198:0.36883)Branch 264:0.43404)Branch
339:0.45462)Branch 381:0.45424)Branch 397:0.4596)Branch
416:0.47812,((NC_011960_Rhodobacter:0.12425,(NC_009007_Rhodobacter:0.091113,NC_009040_Rhodobacter:0.090705)Branch 53:0.12322)Branch
73:0.47439,(NC_007489_Rhodobacter:0.43915,NC_009430_Rhodobacter:0.43265)Branch 348:0.47296)Branch 445:0.47933)Branch 460:0.48624)Branch
480:0.49334,((NC_010580_Beijerinckia:0.48068,(NC_008386_Roseobacter:0.4661,(NC_008042_Ruegeria:0.42272,(NC_004574_Ruegeria:0.38514,(NC_009955_Dinor
oseobacter:0.18287,NC_009957_Dinoroseobacter:0.18688)Branch 118:0.38455)Branch 286:0.41485)Branch 322:0.46783)Branch 424:0.48304)Branch
468:0.48949,((((NC_009671_Ochrobactrum:0.44735,(NC_007490_Rhodobacter:0.4449,NC_009431_Rhodobacter:0.43882)Branch 374:0.45635)Branch
405:0.46933,((NC_013545_Sinorhizobium:0.44873,(NC_009622_Sinorhizobium:0.37043,(NC_009669_Ochrobactrum:0.36105,NC_011986_Agrobacterium:0.35739)Br
anch 267:0.36879)Branch 274:0.44864)Branch 396:0.46573,((((NC_007960_Nitrobacter:0.37529,NC_008242_Chelativorans:0.3717)Branch
278:0.40943,(NC_009475_Bradyrhizobium:0.36934,NC_009717_Xanthobacter:0.36816)Branch 279:0.40419)Branch
311:0.43287,(NC_002679_Mesorhizobium:0.4253,NC_009670_Ochrobactrum:0.42742)Branch 351:0.43029)Branch
358:0.45799,(NC_008382_Rhizobium:0.43822,(NC_010865_Sinorhizobium:0.40387,(((NC_011990_Agrobacterium:0.24166,(NC_002147_Agrobacterium:0.11875,NC_
003065_Agrobacterium:0.11934)Branch 78:0.24203)Branch
160:0.35943,((NC_011982_Agrobacterium:0.26174,(NC_002377_Agrobacterium:0.21002,NC_010929_Agrobacterium:0.20956)Branch 146:0.26669)Branch
176:0.32809,(NC_007762_Rhizobium:0.32012,(NC_002575_Agrobacterium:0.11639,NC_010841_Agrobacterium:0.11938)Branch 77:0.32638)Branch
234:0.32987)Branch 239:0.35763)Branch
266:0.39057,((NC_007961_Nitrobacter:0.35563,(NC_005873_Oligotropha:0.34915,NC_011991_Agrobacterium:0.34497)Branch 252:0.35763)Branch
262:0.37406,(NC_008383_Rhizobium:0.37057,NC_011984_Agrobacterium:0.36824)Branch 276:0.37131)Branch 277:0.39136)Branch 296:0.40982)Branch
312:0.44103)Branch 380:0.45774)Branch 420:0.46498)Branch 442:0.46999)Branch
453:0.47478,(NC_007959_Nitrobacter:0.47377,((NC_011758_Methylobacterium:0.40779,NC_012987_Methylobacterium:0.41039)Branch
319:0.46128,(NC_006672_Gluconobacter:0.41306,(NC_013210_Acetobacter:0.38702,NC_013211_Acetobacter:0.38325)Branch 288:0.41692)Branch
```

```
327:0.46144)Branch 431:0.46926)Branch 454:0.47614)Branch
464:0.48358,(((NC_007763_Rhizobium:0.28615,(NC_011371_Rhizobium:0.19348,(NC_008379_Rhizobium:0.1578,NC_012852_Rhizobium:0.15899)Branch
99:0.19226)Branch 121:0.28591)Branch 182:0.48013,((NC_011961_Thermomicrobium:0.45784,NC_013859_Azospirillum:0.45742)Branch
409:0.47184,(NC_008243_Chelativorans:0.46776,(NC_011370_Rhizobium:0.4462,NC_011987_Agrobacterium:0.44328)Branch 387:0.46696)Branch
443:0.47128)Branch 451:0.47532)Branch
459:0.48169,((NC_011887_Methylobacterium:0.45778,(NC_010510_Methylobacterium:0.44079,NC_011892_Methylobacterium:0.44156)Branch 384:0.45694)Branch
417:0.47568,((NC_009429_Rhodobacter:0.44837,(NC_014035_Rhodobacter:0.41788,(NC_007488_Rhodobacter:0.15612,NC_011962_Rhodobacter:0.15571)Branch
97:0.41468)Branch 318:0.4558)Branch
398:0.47568,(((NC_011981_Agrobacterium:0.43712,(NC_007764_Rhizobium:0.26652,(NC_008381_Rhizobium:0.1947,NC_012854_Rhizobium:0.1966)Branch
123:0.26419)Branch 172:0.44174)Branch 367:0.46421,(((NC_003037_Sinorhizobium:0.29117,NC_009621_Sinorhizobium:0.29016)Branch
189:0.43991,((NC_012586_Rhizobium:0.27046,(NC_003078_Sinorhizobium:0.09271,NC_009620_Sinorhizobium:0.092741)Branch 55:0.27281)Branch
174:0.42664,(NC_011368_Rhizobium:0.31795,(NC_010997_Rhizobium:0.30474,(NC_007766_Rhizobium:0.27962,(NC_008378_Rhizobium:0.1605,NC_012848_Rhizob
ium:0.16014)Branch 102:0.27832)Branch 179:0.30522)Branch 205:0.31793)Branch 215:0.42787)Branch 338:0.4439)Branch
373:0.46087,(NC_003064_Agrobacterium:0.4471,NC_012853_Rhizobium:0.45186)Branch 393:0.45958)Branch 411:0.46432)Branch
422:0.47398,(((NC_008688_Paracoccus:0.44562,NC_013857_Azospirillum:0.44383)Branch
386:0.45907,((NC_010625_Burkholderia:0.42826,(NC_007974_Cupriavidus:0.34354,(NC_003296_Ralstonia:0.11217,NC_014309_Ralstonia:0.11325)Branch
67:0.34552)Branch 245:0.43033)Branch 343:0.45639,(NC_013855_Azospirillum:0.44516,NC_013856_Azospirillum:0.44234)Branch 390:0.45102)Branch
401:0.45834)Branch 410:0.4685,(((NC_008384_Rhizobium:0.12932,NC_012858_Rhizobium:0.13017)Branch
83:0.20267,(NC_011366_Rhizobium:0.16291,(NC_007765_Rhizobium:0.16118,NC_010998_Rhizobium:0.16069)Branch 101:0.16272)Branch 103:0.20355)Branch
127:0.46596,(NC_013858_Azospirillum:0.44967,(NC_006569_Ruegeria:0.43954,NC_008043_Ruegeria:0.43914)Branch 368:0.44888)Branch 392:0.46311)Branch
421:0.47098)Branch 444:0.4734)Branch 450:0.47695)Branch 458:0.47914)Branch 463:0.47908)Branch 466:0.48712)Branch 485:0.48902)Branch 498:0.4918)Branch
507:0.49574,(((NC_010404_Acinetobacter:0.49173,(NC_013930_Thioalkalivibrio:0.48342,(NC_010815_Geobacter:0.44714,(NC_005838_Thermus:0.17562,NC_00646
2_Thermus:0.17322)Branch 108:0.45059)Branch 379:0.4821)Branch 477:0.48908)Branch
496:0.49586,(((NC_008759_Polaromonas:0.4459,NC_013852_Allochromatium:0.44299)Branch
354:0.48181,(NC_010679_Burkholderia:0.47369,(NC_010683_Ralstonia:0.10734,NC_012851_Ralstonia:0.10694)Branch 62:0.47335)Branch 433:0.48208)Branch
462:0.49461,((((NC_005241_Ralstonia:0.39212,(NC_012849_Ralstonia:0.35213,(NC_006525_Cupriavidus:0.079735,NC_007972_Cupriavidus:0.079839)Branch
50:0.35304)Branch 257:0.38844)Branch
289:0.4135,(NC_010529_Cupriavidus:0.39768,(NC_014120_Burkholderia:0.35203,(NC_009227_Burkholderia:0.31341,NC_010553_Burkholderia:0.31925)Branch
221:0.34995)Branch 253:0.4001)Branch 303:0.41043)Branch
310:0.46319,(NC_010627_Burkholderia:0.42666,(NC_000914_Rhizobium:0.38506,(NC_004041_Rhizobium:0.13205,NC_010996_Rhizobium:0.13741)Branch
87:0.3953)Branch 298:0.42559)Branch 346:0.4607)Branch 423:0.48556,(((NC_009229_Burkholderia:0.42416,NC_009230_Burkholderia:0.41932)Branch
324:0.46984,(NC_006823_Aromatoleum:0.45346,NC_008826_Methylibium:0.4485)Branch 391:0.46623)Branch
432:0.48392,((NC_008760_Polaromonas:0.47116,NC_012855_Ralstonia:0.47507)Branch
446:0.48052,(NC_007336_Ralstonia:0.47489,(NC_007949_Polaromonas:0.43579,NC_007950_Polaromonas:0.43577)Branch 363:0.472)Branch 457:0.47649)Branch
465:0.48427)Branch 481:0.48668)Branch 489:0.49202)Branch 509:0.496)Branch
515:0.49721,(((NC_014258_Pantoea:0.46332,(NC_009780_Cronobacter:0.13445,NC_013283_Cronobacter:0.13341)Branch 80:0.46832)Branch
414:0.49237,((NC_003385_Salmonella:0.265,((NC_006323_Yersinia:0.086541,NC_009378_Yersinia:0.084887)Branch
59:0.21005,(NC_010158_Yersinia:0.179,(NC_014022_Yersinia:0.14307,((NC_008118_Yersinia:0.070059,NC_008120_Yersinia:0.069941)Branch
49:0.11063,((NC_003134_Yersinia:0.02984,NC_009596_Yersinia:0.03016)Branch
19:0.094487,(NC_005815_Yersinia:0.088925,(NC_004835_Yersinia:0.064719,NC_004838_Yersinia:0.063721)Branch 46:0.088471)Branch 57:0.095724)Branch
60:0.11172)Branch 71:0.14293)Branch 89:0.17899)Branch 115:0.2179)Branch 141:0.2532)Branch
163:0.49199,(NC_010113_Vibrio:0.47235,NC_013370_Escherichia:0.46515)Branch 413:0.49182)Branch 492:0.49343)Branch
497:0.49823,((((NC_009959_Dinoroseobacter:0.47822,((NC_010844_Leptospira:0,NC_010846_Leptospira:0)Branch
1:0.46995,(NC_005863_Desulfovibrio:0.041627,NC_008741_Desulfovibrio:0.041707)Branch 15:0.46939)Branch 415:0.48086)Branch
452:0.49089,((NC_014249_Nostoc:0.45591,(NC_004073_Synechococcus:0,NC_007595_Synechococcus:0)Branch 2:0.45318)Branch
383:0.48633,(NC_010474_Synechococcus:0.46592,(NC_005230_Synechocystis:0.43675,(NC_005229_Synechocystis:0.33879,NC_005232_Synechocystis:0.34303)Bran
ch 238:0.44034)Branch 353:0.46809)Branch 430:0.48026)Branch 470:0.48783)Branch
486:0.49539,(((NC_003273_Nostoc:0.45271,(NC_003267_Nostoc:0.3669,NC_010630_Nostoc:0.36644)Branch 261:0.44846)Branch
366:0.47844,((NC_010631_Nostoc:0.34228,(NC_010632_Nostoc:0.24541,(NC_003276_Nostoc:0.20346,NC_007410_Anabaena:0.20305)Branch 125:0.24642)Branch
157:0.34185)Branch 237:0.46623,((NC_003240_Nostoc:0.38697,NC_007412_Anabaena:0.38576)Branch
282:0.45103,(NC_011737_Cyanothece:0.41396,NC_011738_Cyanothece:0.41531)Branch 309:0.45355)Branch 378:0.46441)Branch 406:0.47633)Branch
441:0.4929,((((NC_009926_Acaryochloris:0.30487,NC_009928_Acaryochloris:0.30565)Branch
203:0.36398,(NC_009929_Acaryochloris:0.35686,NC_009931_Acaryochloris:0.35919)Branch 256:0.36413)Branch
260:0.421,(NC_009927_Acaryochloris:0.36636,(NC_009932_Acaryochloris:0.35219,(NC_009930_Acaryochloris:0.30522,NC_009933_Acaryochloris:0.30262)Branch
202:0.35261)Branch 250:0.36851)Branch 265:0.41912)Branch
320:0.48823,(NC_013160_Cyanothece:0.48271,(NC_011880_Cyanothece:0.38566,NC_011885_Cyanothece:0.39212)Branch 287:0.47265)Branch 455:0.48809)Branch
491:0.48955)Branch 493:0.4951)Branch
512:0.49829,(NC_005915_Caedibacter:0.4996,(NC_013784_Zymomonas:0.49994,(NC_009035_Shewanella:0.49996,(NC_013502_Rhodothermus:0.49999,((((NC_006
129_Borrelia:0.21078,(NC_012234_Borrelia:0.15925,(NC_011784_Borrelia:0.14058,(((NC_011859_Borrelia:0.040485,NC_011877_Borrelia:0.041482)Branch
27:0.078304,(NC_012177_Borrelia:0.0525,(NC_011786_Borrelia:0.026834,(NC_008564_Borrelia:0.0076253,NC_012164_Borrelia:0.0079997)Branch
10:0.026546)Branch 21:0.052318)Branch 32:0.078089)Branch
48:0.1102,(NC_013128_Borrelia:0.096759,((NC_011876_Borrelia:0.031331,NC_012241_Borrelia:0.032161)Branch
23:0.064229,(((NC_013129_Borrelia:0,NC_013130_Borrelia:0)Branch 8:0.046444,(NC_001857_Borrelia:0.045684,NC_012504_Borrelia:0.045225)Branch
28:0.047328)Branch
29:0.055785,(NC_012199_Borrelia:0.05071,(NC_012505_Borrelia:0.040022,(NC_012175_Borrelia:0.028091,(NC_012193_Borrelia:0.01124,(NC_012194_Borrelia:0.00
74517,NC_012244_Borrelia:0.0076998)Branch 11:0.011488)Branch 12:0.028218)Branch 22:0.039913)Branch 26:0.050679)Branch 31:0.056482)Branch
36:0.06399)Branch 41:0.096824)Branch 56:0.11042)Branch 65:0.1401)Branch 86:0.15979)Branch 96:0.21025)Branch
129:0.40672,((NC_011246_Borrelia:0.1211,NC_011247_Borrelia:0.12133)Branch
72:0.38195,(NC_012247_Borrelia:0.21492,(NC_000956_Borrelia:0.11337,(NC_011720_Borrelia:0.036824,(NC_012106_Borrelia:0.024477,NC_012253_Borrelia:0.0243
04)Branch 9:0.036956)Branch 16:0.1139)Branch 66:0.21562)Branch 132:0.38157)Branch 270:0.40583)Branch
297:0.46478,(NC_011856_Borrelia:0.44945,(((NC_011860_Borrelia:0.26879,NC_012166_Borrelia:0.26967)Branch
170:0.39252,(NC_008567_Borrelia:0.34189,(NC_011842_Borrelia:0.22769,NC_011869_Borrelia:0.22686)Branch 147:0.34144)Branch 236:0.39255)Branch
283:0.42425,((NC_012185_Borrelia:0.33245,(NC_012167_Borrelia:0.062921,NC_012182_Borrelia:0.062079)Branch 30:0.33436)Branch
```

```
224:0.41391,((NC_012202_Borrelia:0.20345,(NC_001855_Borrelia:0.065161,NC_012184_Borrelia:0.065274)Branch 35:0.20564)Branch
126:0.38923,(NC_008565_Borrelia:0.31733,(NC_011857_Borrelia:0.23359,NC_011867_Borrelia:0.23307)Branch 150:0.3177)Branch 213:0.38844)Branch
281:0.41355)Branch 306:0.42433)Branch 314:0.44909)Branch 357:0.46488)Branch 395:0.5,(NC_012439_Persephonella:0.5,NC_013518_Sebaldella:0.5)Branch
524:1)Branch 525:0.5)Branch 523:0.49999)Branch 522:0.49995)Branch 521:0.49988)Branch 520:0.49944)Branch 519:0.49842)Branch 518:0.49764)Branch
517:0.49681)Branch 516:0.49584)Branch 514:0.23863)Root;
```

Appendix H

Fig. H.1. Newick standard file of Euclidean distance tree for 527 Gram-negative plasmids.

```
((((NC_003385_Salmonella:3.1606,(NC_010158_Yersinia:3.0183,(NC_014022_Yersinia:2.553,((NC_006323_Yersinia:2.6032,NC_009378_Yersinia:2.2958)Branch
37:3.8852,((NC_008118_Yersinia:2.0349,NC_008120_Yersinia:1.7067)Branch 48:2.2996,((NC_003134_Yersinia:1.2621,NC_009596_Yersinia:1.1874)Branch
35:2.2318,(NC_004835_Yersinia:1.919,(NC_004838_Yersinia:1.8785,NC_005815_Yersinia:2.1215)Branch 38:2.5013)Branch 44:2.4993)Branch 54:2.6019)Branch
60:2.778)Branch 81:3.9869)Branch 112:3.518)Branch 121:4.0114)Branch
161:6.1546,((NC_009350_Aeromonas:4.5192,((NC_002120_Yersinia:2.2918,(NC_008791_Yersinia:1.3993,(NC_004564_Yersinia:0.88698,NC_005017_Yersinia:0.84507
)Branch 57:1.3255)Branch 76:2.476)Branch 129:2.935,(NC_014017_Yersinia:2.1981,(((NC_004836_Yersinia:1.0976,NC_004839_Yersinia:1.1384)Branch
64:1.3793,(NC_005813_Yersinia:1.5085,(NC_003131_Yersinia:1.1369,NC_009595_Yersinia:1.0991)Branch 59:1.2286)Branch 65:1.6059)Branch
77:2.1484,((NC_006153_Yersinia:1.7639,(NC_008122_Yersinia:1.0423,NC_009377_Yersinia:0.95768)Branch 45:1.8416)Branch
79:2.177,(NC_010157_Yersinia:1.9657,NC_010635_Yersinia:1.9073)Branch 93:2.0609)Branch 97:2.3603)Branch 108:2.72)Branch 130:2.9915)Branch
157:4.1871)Branch 248:5.6257,((((NC_006855_Salmonella:1.6315,NC_012124_Salmonella:1.8326)Branch
137:3.1881,(NC_007208_Salmonella:3.2147,(NC_003277_Salmonella:2.7244,NC_013437_Salmonella:3.0202)Branch 98:4.0568)Branch 154:4.9746)Branch
243:5.2196,(NC_009425_Enterobacter:3.6548,NC_014305_Erwinia:3.4866)Branch 253:4.5616)Branch
371:5.0336,(((((NC_002305_Salmonella:3.4382,(NC_013365_Escherichia:2.3697,(NC_003384_Salmonella:2.5663,NC_009981_Salmonella:2.6299)Branch
14:2.6779)Branch 15:4.3487)Branch
28:5.9526,(NC_009838_Escherichia:4.4309,(NC_005211_Serratia:5.1567,(NC_010870_Klebsiella:4.5574,(NC_012555_Enterobacter:0.48614,NC_012556_Enterobacte
r:0.51386)Branch 4:5.1121)Branch 20:5.1749)Branch 22:6.2271)Branch 33:6.9732)Branch
84:9.9436,(NC_011282_Klebsiella:4.2531,(NC_014107_Enterobacter:6.8725,(NC_005249_Klebsiella:1.7789,NC_006625_Klebsiella:2.094)Branch 24:6.6361)Branch
224:7.181)Branch 249:6.2422)Branch 298:8.5258,((NC_011092_Salmonella:3.5947,(NC_006856_Salmonella:5.2895,NC_010119_Salmonella:4.7604)Branch
192:5.2631)Branch
213:5.7948,(NC_006816_Salmonella:5.312,(NC_011964_Escherichia:5.3816,(NC_011980_Escherichia:4.3775,(NC_007675_Escherichia:4.5417,((NC_009837_Escherich
ia:4.4891,NC_010409_Escherichia:3.5731)Branch 85:4.6769,(NC_011076_Salmonella:4.0079,NC_011747_Escherichia:3.6733)Branch 96:4.2274)Branch
114:5.036)Branch 135:5.0804)Branch 152:6.0768)Branch 240:6.5228)Branch 313:6.1009)Branch 337:6.0508)Branch
349:7.5308,(((NC_009788_Escherichia:3.6418,(NC_007635_Escherichia:4.0051,NC_014234_Escherichia:3.6107)Branch 212:4.5157)Branch
266:4.4904,((NC_013369_Escherichia:2.2381,NC_013728_Escherichia:3.3297)Branch
158:4.5418,(NC_011754_Escherichia:4.6494,((NC_005014_Salmonella:4.4377,(NC_013120_Escherichia:3.0273,(NC_011419_Escherichia:3.5256,(NC_011077_Salmon
ella:2.5445,NC_011081_Salmonella:2.2514)Branch 51:3.5028)Branch 103:3.8788)Branch 120:3.5602)Branch
122:4.8468,(NC_007365_Escherichia:4.0073,NC_012487_Escherichia:4.478)Branch 132:4.8392)Branch 167:5.2605)Branch 196:6.2773)Branch 302:5.4633)Branch
306:5.4467,(((NC_013942_Escherichia:3.9393,(NC_013354_Escherichia:3.1155,(NC_009602_Escherichia:3.9171,(NC_011350_Escherichia:2.5125,(NC_002128_Escheri
chia:2.0588,(NC_007414_Escherichia:2.059,NC_013010_Escherichia:2.2999)Branch 70:2.5701)Branch 92:2.7753)Branch 111:4.1195)Branch 190:4.5077)Branch
230:4.0414)Branch
242:4.7227,((NC_011413_Escherichia:4.0869,((NC_009649_Klebsiella:5.925,(NC_013950_Klebsiella:4.1215,(NC_014312_Klebsiella:4.5657,(NC_009650_Klebsiella:2.3
846,NC_014016_Klebsiella:2.5143)Branch 63:4.266)Branch 150:4.9456)Branch 179:4.7788)Branch
203:6.133,((NC_012944_Escherichia:3.9374,(NC_010862_scherichia:1.7883,NC_011603_Escherichia:1.6758)Branch 56:4.6632)Branch
214:4.4838,((NC_010720_Escherichia:3.7406,((NC_008460_Escherichia:4.5942,(NC_013951_Klebsiella:4.6539,(NC_013175_Escherichia:2.7034,(NC_007941_Escheric
hia:3.3126,NC_011749_Escherichia:3.3206)Branch 75:3.3788)Branch 78:4.0831)Branch 126:5.0532)Branch
159:5.3129,(NC_006671_Escherichia:4.7783,(NC_013362_Escherichia:3.1188,((NC_013542_Klebsiella:2.4069,NC_013727_Shigella:2.389)Branch
99:3.476,(NC_005327_Escherichia:3.2236,(NC_009133_Escherichia:4.0132,NC_011812_Escherichia:3.198)Branch 118:3.776)Branch 124:4.0169)Branch
141:3.8949)Branch 155:3.81)Branch 170:4.5527)Branch 185:4.9857)Branch 191:4.6702,((NC_004998_Escherichia:4.8672,NC_010488_Escherichia:5.1328)Branch
156:5.6447,(NC_010558_Escherichia:5.375,NC_013122_Escherichia:4.5749)Branch 160:5.5322)Branch 197:5.8657)Branch 217:5.3975)Branch 225:5.3854)Branch
252:5.9066)Branch 268:5.3469,(NC_013366_Escherichia:3.7951,(NC_013121_Escherichia:3.8717,NC_013507_Escherichia:4.0023)Branch 231:4.506)Branch
276:4.4174)Branch 300:5.0099)Branch
312:5.0906,((NC_007608_Shigella:3.69,(NC_010719_Escherichia:6.0054,(NC_010660_Shigella:3.985,(NC_007385_Shigella:3.9003,(NC_007607_Shigella:3.4318,(NC_0
02698_Shigella:3.211,NC_004851_Shigella:3.1136)Branch 34:4.118)Branch 53:4.1206)Branch 66:4.3002)Branch 83:5.7969)Branch 164:5.93)Branch
166:6.8015,(NC_011752_Escherichia:4.264,(NC_009786_Escherichia:2.4794,NC_014232_Escherichia:4.078)Branch 211:4.4228)Branch 307:4.4998)Branch
329:5.766)Branch 341:5.6113)Branch 367:5.6277)Branch 378:6.689)Branch 390:5.9293)Branch 398:5.877)Branch
411:6.058,((NC_012886_Escherichia:3.7993,(NC_009349_Aeromonas:4.6789,(NC_014170_Xenorhabdus:4.2247,((NC_012690_Escherichia:3.0868,(NC_012693_Salmo
nella:2.2005,(NC_009140_Salmonella:2.2075,NC_012692_Escherichia:2.2646)Branch 16:2.6595)Branch 19:3.7193)Branch
30:4.9313,(NC_012885_Aeromonas:3.4849,(NC_009139_Yersinia:3.6779,(NC_009141_Yersinia:4.3882,(NC_008612_Photobacterium:2.7482,NC_008613_Photobacte
rium:2.1508)Branch 21:3.286)Branch 29:4.0607)Branch 31:4.4611)Branch 40:4.8671)Branch 67:6.3082)Branch 134:5.9741)Branch 176:6.9844)Branch
314:6.139,(((NC_008275_Pseudomonas:3.6287,(NC_014124_Pseudomonas:3.7485,(NC_004444_Pseudomonas:0.67661,NC_011838_Pseudomonas:0.73761)Branch
17:5.3618)Branch 140:6.1786)Branch
263:5.4343,(NC_005244_Pseudomonas:4.7974,(NC_003350_Pseudomonas:4.116,(NC_007926_Pseudomonas:1.869,(NC_004999_Pseudomonas:0,NC_012674_Pseud
omonas:0)Branch 27:2.2541)Branch 94:3.8452)Branch 194:5.1532)Branch 292:5.2023)Branch
316:5.9335,((NC_009427_Novosphingobium:6.134,(NC_009508_Sphingomonas:4.9773,(NC_009507_Sphingomonas:5.435,(NC_008308_Sphingomonas:6.0378,(NC_0
02033_Novosphingobium:0.99913,NC_009426_Novosphingobium:1.0009)Branch 25:5.1424)Branch 182:6.6265)Branch 265:6.8305)Branch 353:6.0188)Branch
366:6.2432,(((NC_011961_Thermomicrobium:5.4868,NC_013859_Azospirillum:7.2411)Branch
335:6.5684,((NC_011370_Rhizobium:6.1086,(NC_002679_Mesorhizobium:5.8373,NC_011987_Agrobacterium:6.9689)Branch 290:6.6563)Branch
318:6.6223,((((NC_009227_Burkholderia:6.1499,NC_010553_Burkholderia:4.9856)Branch
145:6.8295,((NC_006525_Cupriavidus:1.6651,NC_007972_Cupriavidus:2.2079)Branch 58:4.6446,(NC_005241_Ralstonia:8.0931,NC_012849_Ralstonia:5.9426)Branch
205:7.1725)Branch 215:6.213)Branch 237:7.3047,((NC_010627_Burkholderia:6.7367,(NC_010529_Cupriavidus:7.1856,NC_014120_Burkholderia:7.1322)Branch
153:7.7711)Branch 195:8.0315,(NC_000914_Rhizobium:7.124,(NC_004041_Rhizobium:3.6122,NC_010996_Rhizobium:3.096)Branch 39:6.7858)Branch
204:7.6252)Branch 246:8.4699)Branch
301:8.2205,(NC_011887_Methylobacterium:6.9509,(NC_007336_Ralstonia:7.877,((NC_010510_Methylobacterium:7.5544,NC_011892_Methylobacterium:7.4456)Br
anch
261:7.8941,((NC_006569_Ruegeria:7.4407,(NC_013858_Azospirillum:8.5036,(NC_013856_Azospirillum:9.104,(NC_013857_Azospirillum:8.5726,(NC_008043_Ruegeri
a:8.9392,(NC_009429_Rhodobacter:10.1538,(NC_008688_Paracoccus:10.1034,(((NC_008384_Rhizobium:5.4351,NC_012858_Rhizobium:5.6554)Branch
6:7.9803,(NC_010998_Rhizobium:5.02,(NC_007765_Rhizobium:6.1432,NC_011366_Rhizobium:5.8568)Branch 7:6.3816)Branch 9:6.2369)Branch
11:12.6802,(NC_013855_Azospirillum:11.171,(((NC_003037_Sinorhizobium:11.1214,NC_009621_Sinorhizobium:9.7592)Branch
12:13.8007,((NC_012586_Rhizobium:15.2848,(NC_003078_Sinorhizobium:6.5107,NC_009620_Sinorhizobium:6.9429)Branch 1:11.9816)Branch
3:21.2941,(NC_007766_Rhizobium:8.5157,((NC_008378_Rhizobium:6.9705,NC_012848_Rhizobium:7.2422)Branch
5:10.1898,(NC_010997_Rhizobium:12.0743,NC_011368_Rhizobium:13.7133)Branch 8:13.4461)Branch 10:12.2867)Branch 13:12.9313)Branch 36:17.7732)Branch
47:17,(NC_010625_Burkholderia:11.8481,(NC_007974_Cupriavidus:12.734,(NC_003296_Ralstonia:6.0019,NC_014309_Ralstonia:6.2046)Branch 2:12.3344)Branch
18:15.7244)Branch 68:14.4572)Branch 102:16.3364)Branch 127:14.1979)Branch 146:13.9845)Branch 178:12.2433)Branch 199:11.3734)Branch 209:10.4292)Branch
```

```
227:9.6663)Branch 238:9.5913)Branch 250:9.2589)Branch
270:8.6267,(NC_014258_Pantoea:8.3417,((NC_007763_Rhizobium:4.9045,(NC_011371_Rhizobium:4.4317,(NC_008379_Rhizobium:4.7375,NC_012852_Rhizobium:4.
3729)Branch 23:5.437)Branch 32:6.898)Branch
95:8.7735,(NC_012853_Rhizobium:8.6301,(NC_003064_Agrobacterium:10.2286,(NC_011981_Agrobacterium:9.2553,(NC_007764_Rhizobium:5.5007,(NC_008381_Rh
izobium:6.579,NC_012854_Rhizobium:4.2376)Branch 26:6.9772)Branch 52:9.1693)Branch 206:9.4417)Branch 222:10.1598)Branch 239:9.7958)Branch
274:9.4801)Branch 286:8.9727)Branch 291:8.8351)Branch 293:8.4649)Branch 305:8.3176)Branch 319:7.7891)Branch 342:8.0827)Branch 352:7.4264)Branch
360:7.1754,(NC_011143_Phenylobacterium:6.0557,(NC_012811_Methylobacterium:6.5236,(NC_007960_Nitrobacter:4.4248,(NC_009475_Bradyrhizobium:4.5887,(N
C_008242_Chelativorans:6.9128,NC_009717_Xanthobacter:6.2781)Branch 223:7.0877)Branch 255:6.1129)Branch 280:5.8324)Branch 343:6.3551)Branch
364:6.347)Branch 379:6.8134)Branch 386:6.6424)Branch 399:6.4011)Branch 413:6.3015)Branch
420:2.6014,((NC_012855_Ralstonia:5.0623,((NC_008826_Methylibium:5.5918,(NC_009230_Burkholderia:6.2921,(NC_006466_Cupriavidus:1.2851,NC_007971_Cupri
avidus:1.3607)Branch 42:5.6031)Branch 299:6.2922)Branch 344:6.382,(NC_007949_Polaromonas:4.4814,NC_007950_Polaromonas:5.2654)Branch
359:5.1899)Branch 397:5.8229)Branch
404:5.5985,((((NC_008243_Chelativorans:4.6261,(NC_014035_Rhodobacter:4.066,(NC_007488_Rhodobacter:2.3214,NC_011962_Rhodobacter:3.0638)Branch
113:5.4979)Branch 317:5.4002)Branch
400:5.1248,(NC_009671_Ochrobactrum:4.1879,((NC_013545_Sinorhizobium:5.1553,(NC_011986_Agrobacterium:3.7645,(NC_009622_Sinorhizobium:4.5967,NC_009
669_Ochrobactrum:4.7841)Branch 256:4.8955)Branch 279:5.0717)Branch
363:5.3925,((NC_009670_Ochrobactrum:4.5372,(NC_010865_Sinorhizobium:5.2061,(NC_011991_Agrobacterium:3.6104,((NC_008383_Rhizobium:4.7192,(NC_01198
4_Agrobacterium:5.2582,(((NC_011990_Agrobacterium:3.6122,(NC_002147_Agrobacterium:2.6829,NC_003065_Agrobacterium:3.2332)Branch 41:4.8706)Branch
123:5.3509,(NC_011982_Agrobacterium:4.3684,(NC_002377_Agrobacterium:3.1073,NC_010929_Agrobacterium:4.6387)Branch 107:4.5583)Branch
136:5.2298)Branch 184:5.6688,(NC_007762_Rhizobium:4.3529,(NC_002575_Agrobacterium:2.8087,NC_010841_Agrobacterium:2.5764)Branch 55:5.3124)Branch
201:5.0925)Branch 219:5.7995)Branch 245:5.6461)Branch 254:5.3488,(NC_005873_Oligotropha:4.0644,NC_007961_Nitrobacter:4.4209)Branch 267:4.2756)Branch
271:4.8369)Branch 273:4.4558)Branch 288:5.388)Branch
362:5.1654,(NC_008382_Rhizobium:4.8248,(NC_011758_Methylobacterium:6.3176,NC_012987_Methylobacterium:3.6323)Branch 315:5.3873)Branch
370:5.2266)Branch 385:5.284)Branch 396:5.3791)Branch 402:4.8933)Branch
419:5.0822,((NC_008042_Ruegeria:3.7877,(NC_004574_Ruegeria:3.9377,(NC_009955_Dinoroseobacter:3.902,NC_009957_Dinoroseobacter:2.7313)Branch
106:5.841)Branch 278:5.4319)Branch
336:5.226,(NC_007959_Nitrobacter:5.0165,(NC_009430_Rhodobacter:5.2721,(NC_011960_Rhodobacter:2.0596,(NC_009007_Rhodobacter:1.6016,NC_009040_Rhod
obacter:2.1401)Branch 90:2.2748)Branch 109:5.2378)Branch 357:5.6148)Branch 401:5.4524)Branch 423:5.3896)Branch
427:5.2869,(((NC_008757_Polaromonas:4.4302,(NC_002682_Mesorhizobium:4.8192,NC_014007_Sphingobium:4.2913)Branch 347:4.7505)Branch
372:4.8846,(NC_010333_Caulobacter:4.1277,NC_010335_Caulobacter:4.7041)Branch 391:4.5574)Branch
412:4.9012,((NC_005793_Achromobacter:3.8366,((NC_006824_Aromatoleum:4.2545,NC_013193_Candidatus:4.4058)Branch
259:4.6343,(NC_009704_Yersinia:2.6978,(NC_013176_Pseudomonas:2.8428,((NC_013666_Burkholderia:2.7219,((NC_004956_Pseudomonas:2.9992,(NC_001735_Ent
erobacter:2.203,NC_005088_Delftia:2.4875)Branch 151:2.562)Branch 168:3.1837,((NC_008766_Acidovorax:1.9339,NC_010935_Comamonas:2.6487)Branch
149:2.6463,(NC_006830_Achromobacter:3.349,(NC_005912_Ralstonia:1.2959,NC_007337_Ralstonia:1.3499)Branch 72:3.0874)Branch 165:3.3741)Branch
173:3.3032)Branch 198:3.3781)Branch 207:3.3647,(NC_008357_Pseudomonas:3.6811,NC_012919_Photobacterium:2.7996)Branch 226:3.336)Branch
235:3.5901)Branch 247:3.5207)Branch 281:3.2236)Branch 287:4.2227)Branch
324:4.8589,((NC_008758_Polaromonas:4.4752,NC_008760_Polaromonas:4.1271)Branch
406:4.5764,((NC_006823_Aromatoleum:3.8793,((NC_010070_Burkholderia:0.99987,NC_010802_Burkholderia:1.0001)Branch
71:5.193,(NC_008765_Acidovorax:2.6189,(NC_011003_Burkholderia:2.8372,(NC_008545_Burkholderia:3.987,(NC_012720_Burkholderia:3.0935,NC_012725_Burkhol
deria:2.8226)Branch 175:3.3266)Branch 208:4.3506)Branch 262:4.2136)Branch 334:3.9487)Branch 416:4.7562)Branch
446:4.3833,((((NC_009931_Acaryochloris:3.2921,(NC_009929_Acaryochloris:4.17,(NC_009926_Acaryochloris:4.1417,NC_009928_Acaryochloris:3.4741)Branch
221:4.3708)Branch 264:4.4516)Branch 283:4.4997,((NC_009927_Acaryochloris:5.1014,NC_009930_Acaryochloris:3.0839)Branch
260:4.4632,(NC_009932_Acaryochloris:2.9878,NC_009933_Acaryochloris:2.4894)Branch 284:2.8665)Branch 294:4.1279)Branch
361:4.9353,((NC_006672_Gluconobacter:3.5491,(NC_013210_Acetobacter:3.6465,NC_013211_Acetobacter:3.9034)Branch 325:4.0988)Branch
369:4.3031,(NC_009468_Acidiphilium:4.3159,(NC_009467_Acidiphilium:4.1383,NC_009469_Acidiphilium:3.7357)Branch 392:4.11)Branch 422:4.3514)Branch
440:4.4062)Branch 454:4.7037,((NC_012718_Burkholderia:3.064,NC_012723_Burkholderia:3.7184)Branch
351:4.1011,((((NC_009780_Cronobacter:2.0705,NC_013283_Cronobacter:1.8025)Branch
139:4.8572,(NC_011281_Klebsiella:4.5435,((NC_003905_Proteus:3.8721,NC_004349_Shewanella:3.7437)Branch
403:3.9873,((NC_011383_Klebsiella:3.2036,((NC_009132_Escherichia:2.5799,NC_014368_Klebsiella:2.8053)Branch
143:3.5794,(NC_009980_Salmonella:2.4006,((NC_011617_Klebsiella:2.6541,(NC_007682_Escherichia:2.1022,NC_014231_Escherichia:1.8978)Branch
144:2.2353)Branch 162:2.7681,(NC_011385_Klebsiella:3.0475,NC_014208_Klebsiella:2.4297)Branch 169:2.9208)Branch 180:2.9716)Branch 193:2.8627)Branch
210:3.5047)Branch 234:4.8845,(NC_011743_Escherichia:3.3667,(NC_009651_Klebsiella:4.4767,NC_010886_Klebsiella:4.1836)Branch 356:4.4749)Branch
373:4.1536)Branch 407:4.6794)Branch 429:4.3695)Branch 437:4.5093)Branch
445:4.7738,(((NC_005246_Erwinia:1.8671,(NC_004464_Citrobacter:2.6353,NC_011641_Klebsiella:1.8368)Branch 101:2.5107)Branch
116:5.4765,(NC_002523_Serratia:3.206,((NC_004632_Pseudomonas:2.2489,NC_004633_Pseudomonas:2.2232)Branch
220:3.1231,(NC_011759_Yersinia:2.9257,(NC_012752_Candidatus:2.0685,NC_013973_Erwinia:2.7273)Branch 269:2.6723)Branch 296:2.8308)Branch
303:3.005)Branch 304:3.7167)Branch 389:5.005,((NC_002142_Escherichia:4.8058,(NC_009790_Escherichia:3.498,NC_014233_Escherichia:2.9827)Branch
332:3.5503)Branch 376:4.4713,(NC_011416_Escherichia:2.6984,NC_013717_Citrobacter:3.1325)Branch 395:3.0173)Branch 415:3.9396)Branch 452:4.5139)Branch
459:4.6953,((NC_009228_Burkholderia:3.5152,NC_009229_Burkholderia:3.6959)Branch
421:3.8913,((NC_008386_Roseobacter:3.3963,NC_010580_Beijerinckia:4.2849)Branch
449:4.0321,((NC_010815_Geobacter:2.7069,(NC_005838_Thermus:1.6346,NC_006462_Thermus:2.2384)Branch 188:3.9511)Branch
435:3.59,(NC_010404_Acinetobacter:3.5788,(((NC_010474_Synechococcus:3.6427,(NC_007713_Sodalis:1.2546,(NC_007182_Sodalis:0,NC_007183_Sodalis:0)Branch
69:1.5738)Branch 147:4.1009)Branch 448:4.0662,((NC_003267_Nostoc:2.6444,NC_010630_Nostoc:2.046)Branch
358:2.7911,(((NC_010631_Nostoc:3.9755,(NC_010632_Nostoc:3.3923,(NC_003276_Nostoc:3.2934,NC_007410_Anabaena:3.7777)Branch 110:4.2095)Branch
138:4.9978)Branch 228:5.9668,(NC_003240_Nostoc:3.1265,NC_007412_Anabaena:4.0149)Branch 330:3.9751)Branch
384:5.302,(NC_011737_Cyanothece:2.2595,NC_011738_Cyanothece:3.5715)Branch 393:3.0983)Branch 434:4.2956)Branch 453:3.7072)Branch
494:3.9463,(((NC_007490_Rhodobacter:2.4039,NC_009431_Rhodobacter:3.7605)Branch
439:3.2911,(NC_007489_Rhodobacter:2.8851,((NC_007801_Jannaschia:2.1241,(NC_004929_Ruegeria:2.0174,NC_008387_Roseobacter:1.7243)Branch
320:2.1786)Branch 368:2.5157,(NC_009956_Dinoroseobacter:4.4236,(NC_009958_Dinoroseobacter:3.244,NC_013860_Azospirillum:4.6933)Branch
418:4.0948)Branch 436:4.4146)Branch 462:3.5194)Branch 474:3.2741)Branch
492:3.3699,(((((NC_003922_Xanthomonas:2.9048,NC_007507_Xanthomonas:2.0952)Branch
340:3.0077,(NC_007274_Pseudomonas:3.5075,NC_013212_Acetobacter:2.8956)Branch 417:3.2955)Branch
433:3.418,(NC_008607_Pelobacter:2.5827,NC_011667_Thauera:3.4173)Branch 450:3.1988)Branch
```

```
503:3.3547,((NC_009036_Shewanella:3.1846,NC_010377_Yersinia:2.4722)Branch
482:2.9491,(((NC_009777_Vibrio:3.1596,((NC_009701_Vibrio:1.5703,NC_009703_Vibrio:1.2581)Branch
163:3.7803,((NC_011311_Aliivibrio:2.1331,(NC_010614_Vibrio:2.5634,(NC_005128_Vibrio:2.1683,NC_009702_Vibrio:2.8317)Branch 189:3.2977)Branch
258:3.0775)Branch
277:2.9184,(NC_005871_Photobacterium:2.211,((NC_009998_Shewanella:2.157,(NC_011664_Shewanella:2.1718,(NC_009661_Shewanella:2.0658,NC_009999_Shew
anella:1.8072)Branch 128:2.3475)Branch 142:2.8663)Branch 187:2.9806,(NC_011665_Shewanella:2.3505,NC_011668_Shewanella:2.4453)Branch 229:2.4393)Branch
233:3.3943)Branch 289:2.9309)Branch 308:3.6946)Branch 430:3.8915)Branch
464:3.6585,((NC_005918_Pseudomonas:2.1201,(NC_005205_Pseudomonas:2.3142,NC_007275_Pseudomonas:2.882)Branch 257:2.7979)Branch
282:3.632,(NC_011148_Salmonella:1.8801,((NC_010697_Erwinia:2.3651,(NC_002525_Escherichia:2.9944,NC_011351_Escherichia:2.2017)Branch 275:2.8417)Branch
295:3.3864,(((NC_010422_Salmonella:2.3839,NC_011204_Salmonella:1.8587)Branch
104:4.4615,(NC_013503_Escherichia:2.0871,(NC_010378_Escherichia:2.0988,NC_010860_Salmonella:1.9012)Branch 174:2.1406)Branch 183:2.9235)Branch
251:4.5106,(NC_010555_Proteus:2.4348,(NC_010657_Shigella:1.6712,NC_013718_Citrobacter:1.7929)Branch 232:2.549)Branch 311:2.7208)Branch
348:4.0759)Branch 424:3.9419)Branch 463:2.969)Branch 473:3.4012)Branch
498:3.575,(((NC_008341_Nitrosomonas:3.0489,NC_008342_Nitrosomonas:2.1473)Branch
375:2.9283,(NC_009444_Pseudomonas:3.1198,(NC_008573_Shewanella:5.0965,NC_008739_Marinobacter:4.1231)Branch 388:4.8938)Branch 428:4.0844)Branch
441:3.7351,((NC_011409_Haemophilus:2.5463,NC_014304_Erwinia:3.2847)Branch
426:3.1641,(NC_013285_Cronobacter:3.8997,NC_014108_Enterobacter:2.4248)Branch 460:3.2694)Branch 486:3.2853)Branch 501:3.5429)Branch 507:3.5732)Branch
509:3.2711)Branch
510:3.3453,((NC_008738_Marinobacter:2.6936,((NC_007100_Pseudomonas:3.1831,(NC_006143_Aeromonas:3.7487,NC_013509_Edwardsiella:2.3341)Branch
333:3.5614)Branch 414:3.6946,(NC_010509_Methylobacterium:1.6344,(NC_002490_Xylella:3.1327,NC_014144_Thiomonas:2.5242)Branch 405:3.0512)Branch
457:2.4854)Branch 489:3.1618)Branch 508:2.9794,((NC_013731_Spirosoma:2.2447,NC_013732_Spirosoma:1.8784)Branch
328:2.9238,((((NC_006365_Legionella:2.2159,(NC_009726_Coxiella:0.53838,NC_010258_Coxiella:0.46162)Branch 216:1.8456)Branch
394:2.3887,(NC_010373_Methylobacterium:2.6035,(NC_011880_Cyanothece:2.1461,NC_011885_Cyanothece:2.4364)Branch 382:2.7299)Branch 505:2.706)Branch
513:2.587,(((NC_006139_Desulfotalea:2.2912,NC_011061_Prosthecochloris:1.7088)Branch
442:2.1879,((NC_006366_Legionella:1.211,NC_009966_Fluoribacter:1.789)Branch
339:2.0615,(NC_007901_Rhodoferax:2.1904,(NC_009705_Yersinia:2.0502,NC_012970_Methylovorus:2.4219)Branch 444:2.2803)Branch 465:2.3284)Branch
484:2.2346)Branch 490:2.2851,((NC_005250_Listonella:2.1373,NC_012797_Desulfovibrio:1.8627)Branch
472:2.1504,(NC_009973_Herpetosiphon:2.0346,(NC_005230_Synechocystis:2.0978,NC_010480_Synechococcus:2.0253)Branch 408:2.2567)Branch
480:2.2598)Branch 512:2.2387)Branch 518:2.2736)Branch
520:2.4558,((NC_009035_Shewanella:2.4463,(((NC_002806_Microscilla:1.8384,(NC_008761_Polaromonas:2.4748,NC_011185_Vibrio:1.3982)Branch
468:2.056)Branch
488:1.9832,(NC_010606_Acinetobacter:1.9675,(NC_013160_Cyanothece:1.8486,((NC_008014_Lawsonia:1.5853,(NC_012983_Hirschia:1.4589,NC_014157_Salinibact
er:1.1868)Branch 354:1.6467)Branch 443:1.7427,(((NC_010112_Vibrio:1.3523,(NC_009829_Serratia:0.84561,NC_012109_Desulfobacterium:0.88645)Branch
323:1.1953)Branch 381:1.4007,(NC_006842_Vibrio:1.7936,((NC_006134_Campylobacter:0.54202,NC_008790_Campylobacter:0.45798)Branch
244:1.2759,(NC_005012_Campylobacter:0.70711,NC_008770_Campylobacter:0.70711)Branch 297:1.0669)Branch 365:1.4052)Branch 467:1.6079)Branch
469:1.5729,(NC_011893_Methylobacterium:1.5636,((NC_007110_Rickettsia:1.3766,((NC_009226_Burkholderia:0.90315,(NC_005915_Caedibacter:0.68668,(NC_0101
13_Vibrio:0.66821,(NC_009974_Herpetosiphon:0.6304,(NC_013502_Rhodothermus:0.55435,(NC_012439_Persephonella:0,NC_013518_Sebaldella:0)Branch
241:0.44565)Branch 309:0.57671)Branch 322:0.64245)Branch 326:0.67576)Branch 327:0.80904)Branch
355:1.0284,(NC_010124_Gluconacetobacter:1.0487,NC_013958_Nitrosococcus:1.1874)Branch 377:1.1942)Branch 383:1.1677)Branch
409:1.3366,(NC_003273_Nostoc:1.4115,NC_013784_Zymomonas:1.417)Branch 470:1.4315)Branch 471:1.4089)Branch 477:1.501)Branch 479:1.5715)Branch
483:1.6848)Branch 487:1.796)Branch 491:1.901)Branch 497:1.9774,(((NC_005229_Synechocystis:2.0355,NC_005232_Synechocystis:1.8375)Branch
346:2.5683,(NC_014249_Nostoc:1.6673,(NC_004073_Synechococcus:0,NC_007595_Synechococcus:0)Branch 131:2.8048)Branch 478:2.3293)Branch
495:2.5033,(NC_009959_Dinoroseobacter:2.8081,((NC_010844_Leptospira:0,NC_010846_Leptospira:0)Branch
186:1.6894,(NC_005863_Desulfovibrio:0.53641,NC_008741_Desulfovibrio:0.46359)Branch 202:2.2471)Branch 481:2.057)Branch 496:2.4827)Branch
517:2.517)Branch 521:2.2582)Branch
523:2.3608,((NC_011856_Borrelia:2.1971,(((NC_006129_Borrelia:2.0238,(NC_011784_Borrelia:1.9653,(NC_012234_Borrelia:2.7118,(((NC_011859_Borrelia:1.1845,N
C_011877_Borrelia:1.0516)Branch
87:1.5301,(NC_012177_Borrelia:1.1589,(NC_011786_Borrelia:0.94611,(NC_008564_Borrelia:0.47551,NC_012164_Borrelia:0.52449)Branch 49:0.91992)Branch
62:1.4267)Branch 89:1.6392)Branch 100:1.9582,(NC_013128_Borrelia:1.5351,((NC_011876_Borrelia:1.0413,NC_012241_Borrelia:0.95873)Branch
74:1.4204,((NC_012199_Borrelia:1.2701,(NC_012505_Borrelia:1.1356,(NC_012175_Borrelia:0.93335,(NC_012193_Borrelia:0.59079,(NC_012194_Borrelia:0.47666,NC
_012244_Borrelia:0.52334)Branch 46:0.61632)Branch 50:0.99169)Branch 61:1.1627)Branch 73:1.3296)Branch
82:1.4257,((NC_013129_Borrelia:0,NC_013130_Borrelia:0)Branch 43:1.126,(NC_001857_Borrelia:1.3192,NC_012504_Borrelia:1.1303)Branch 80:1.3234)Branch
86:1.2929)Branch 88:1.4792)Branch 91:1.9541)Branch 115:1.8451)Branch 119:2.083)Branch 125:2.6449)Branch 133:2.8997)Branch
171:4.2735,((NC_011246_Borrelia:1.1724,NC_011247_Borrelia:1.6561)Branch
177:2.9445,(NC_012247_Borrelia:2.0001,(NC_000956_Borrelia:1.5762,(NC_011720_Borrelia:0.79966,(NC_012106_Borrelia:0.70816,NC_012253_Borrelia:0.70606)Br
anch 105:0.90745)Branch 117:1.5821)Branch 148:2.4125)Branch 218:3.3572)Branch 338:3.2347)Branch
350:4.2964,(((NC_011860_Borrelia:1.6136,NC_012166_Borrelia:1.0322)Branch
310:1.7487,(NC_008567_Borrelia:1.5935,(NC_011842_Borrelia:1.5123,NC_011869_Borrelia:1.65)Branch 272:2.1481)Branch 345:2.0911)Branch
380:2.0475,((NC_012185_Borrelia:2.9507,(NC_012167_Borrelia:0.63894,NC_012182_Borrelia:0.77528)Branch 200:1.8451)Branch
321:2.8473,((NC_012202_Borrelia:2.1312,(NC_001855_Borrelia:0.83349,NC_012184_Borrelia:0.89856)Branch 172:1.5424)Branch
236:2.9853,(NC_008565_Borrelia:1.8965,(NC_011857_Borrelia:1.5224,NC_011867_Borrelia:1.1234)Branch 285:1.6211)Branch 331:2.0073)Branch 374:2.6067)Branch
387:2.8299)Branch 410:2.5156)Branch 431:3.5913)Branch 476:3.0708,(((NC_008759_Polaromonas:2.0162,NC_013852_Allochromatium:1.9838)Branch
456:2.1988,(NC_010679_Burkholderia:2.5937,(NC_010683_Ralstonia:1.4133,NC_012851_Ralstonia:1.0362)Branch 181:3.4011)Branch 485:3.0671)Branch
499:2.6654,(NC_013370_Escherichia:2.7414,NC_013930_Thioalkalivibrio:2.3577)Branch 500:2.5941)Branch 515:2.6689)Branch 525:5.4557)Branch
524:2.7996)Branch 522:3.1171)Branch 519:3.2667)Branch 516:3.516)Branch 514:3.6431)Branch 511:3.8884)Branch 506:3.7582)Branch 504:3.9081)Branch
502:4.1987)Branch 493:4.2865)Branch 475:4.5835)Branch 466:4.434)Branch 461:4.6974)Branch 458:4.6522)Branch 455:4.8943)Branch 451:5.1264)Branch
447:5.3377)Branch 438:5.6757)Branch 432:5.9906)Branch 425:3.6294)Root;
```