Analysis of the effect of conserved regions on bacterial plasmid host range

Master Thesis

CS5000: Master Thesis

Connor Schilder

Delft University of Technology

TU Delft

Analysis of the effect of conserved regions on bacterial plasmid host range

Master Thesis

by

Connor Schilder

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Monday April 26, 2024 at 14:00 AM.

Student number:	4375726	
Project duration:	September 1, 2023 – April 26, 2024	
Thesis committee:	Dr. T. Abeel,	TU Delft, supervisor
	Dr. C. Lofi,	TU Delft
	Ir. M. Teixeira,	Broad Institute

Cover: Canadarm 2 Robotic Arm Grapples SpaceX Dragon by NASA under CC BY-NC 2.0 (Modified)
Style: TU Delft Report Style, with modifications by Connor Schilder

Abstract

Background: *Genetic information is shared between different bacteria through mobile genetic elements, among which plasmids. Some plasmids are able to transfer and spread genetic information between different species. Understanding which genes allow plasmids to replicate in different species is useful in containing the antibiotic resistance epidemic, as well as aid in the creation of vectors in the field of genetic engineering. Which characteristics determine plasmid host range is not fully known. Current approaches either use nucleotide composition or a combination of the genes, also called a genetic backbone, comprising replication and conjugation machinery to predict plasmid host range, but these methods are not perfect in their predictions. As research has shown that plasmids have a certain genetic backbone, we test the hypothesis that conserved regions around replication, conjugation genes and post-segregational killing systems affect host range.*
Results: *We found that the genes in our conserved regions were almost exclusively encountered in that specific order. When studying the functional annotations of genes in a conserved region, they seem to be functionally related and serve a more grand biological function. Although the replication and conjugation machinery, which are often used as predictors for host range, were found to be associated with plasmid host range, the conserved regions were generally similarly or more specific than the combination of these predictors, which suggests that some conserved regions affect plasmid host range. Several genes related to evading bacterial immune systems and partitioning systems were found to be less conserved, with more divergence in their sequence, for which we hypothesize that these genes usually mutate to ameliorate the cost of plasmid carriage.*

Contents

0.1. Introduction

Bacteria have been one of the major causes of mortality for most of human history [99], but their mortality decreased with the discovery of penicillin by Alexander Fleming in 1928, kicking off the era of antibiotics [26]. In the following years, millions of lives were saved through the use of antibiotics [26, 99]. However, in the past years, pathogenic bacteria have become resistant to different types of antibiotics, giving rise to multi-drug resistant organisms [26]. The effects of this can already be seen in the medical field, with antibiotic resistance contributing to and resulting in 4.95 and 1.27 million deaths respectively in 2019. Furthermore, it is estimated that antimicrobial resistance will cause 10 million deaths annually by 2050, if no action is taken [99, 42, 93].

Resistance mechanisms to antibiotics can be found in a wide variety of bacteria, whether these are meant to protect the antibiotics producers [137] or to stave off the effects of the antibiotics in their environment. Abundant (mis)use of antibiotic compounds by humans in various fields, among which the clinical, industrial and agricultural sector, has increased the selective pressure for these resistance mechanisms and thus resulted in more resistant bacterial strains [136, 99]. Antibiotic resistance can either be intrinsic or acquired [26]. With intrinsic resistance, the genes encoding resistance are passed during mitosis and augmentation of their transcription leads to a resistant phenotype. Bacteria can also acquire resistance via mutations, or via the acquisition of genetic material from other bacteria through horizontal gene transfer (HGT) [10]. HGT occurs through mobile genetic elements (MGEs), among which plasmids and phages, which are able to acquire genes from the genome of their host and traverse between different organisms. Mutation is thought to be the molecular basis of resistance [144, 83], but HGT is considered the main reason for the spreading of resistance genes [132]. An overview of the different methods by which genes transfer can be found in Figure 1, but of these methods, this thesis will mainly focus on plasmids and thus conjugation of self-transmissible and mobilizable plasmids.

Plasmids are pieces of extrachromosomal, most often circular DNA that encode functions that are not essential, but useful to the cell [15]. They are also commonly used as vectors for genetic engineering [90]. Although plasmids occur across all three domains of life [17], there is a focus on the domain of Bacteria, as these have been deemed a public health risk: it has been found that MGEs, specifically plasmids, are a source of virulence factors (VFs), causing disease, and antibiotic resistance genes (ARGs), making it hard to treat pathogenic bacteria containing them. These genes are often flanked with insertion sequences (ISs) or transposons, allowing these genes to be transferred from a plasmid to other plasmids or chromosomes [84, 139, 82, 18, 49, 67, 23, 133, 96, 15].

As mentioned, some plasmids can be exchanged between different bacteria and even to different species, leading to rapid dissemination of genetic information, among which the ARGs [15]. Other plasmids lack this ability to be transferred between species and are linked to a single host. It is important to know what causes the distinction between these groups, as it would allow us to target these systems and stop the spread of ARGs.

Even though plasmids can offer beneficial traits to their host, their maintenance is costly, which is especially disadvantageous to the host in the absence of selective pressure [124]. As such, it is expected that beneficial genes on plasmids are either integrated into the host genome or lost over time [17]. Research has however discovered plasmids that persist over a long time in the absence of selection for their genes [17].

There are several factors that are thought to contribute to long-term plasmid persistence in a bacterial community. On one hand, mutation of genes on plasmids and chromosomes can result in co-dependency, which establishes the plasmid in the host genome ensuring vertical inheritance. On the other hand, high plasmid transfer rates allow plasmids to spread to different cells in a community faster than the plasmid is lost, which ensures the long-term survival of the plasmid. These factors create opportunities for plasmids and their genes to spread to and replicate in a larger variety of hosts. The range of hosts a plasmid can be transferred to, replicate in and persist in after prolonged periods are called the transfer, replication and persistence host range, respectively [145]. Within this thesis, we use the terms broad- and narrow host range to refer to the breadth of a plasmid's replication host range, with broad host range (BHR) plasmids being defined as plasmids that can traverse to different species. If we understand what plasmid features determine host range, we may use this knowledge to limit plasmid dissemination and stop the spread of ARGs and construct better plasmid vectors for industrial purposes [86].

Plasmids are one of many different mobile genetic elements MGEs, an umbrella term for genetic material that accommodates the movement of genes inside a genome (intragenomic) or between different genomes (intergenomic). Aside from plasmids, intergenomic MGEs include bacteriophages, often referred to as pha-ges, and ICEs.

Delineation between these elements is blurry: some phages exist in the cell like plasmids and encode their own replication and partitioning machinery while also able to infect bacteria and produce capsids similar to bacteriophages [97]. Excised ICEs tend to form circular, double-stranded DNA fragments like plasmids, with some ICEs even being known to replicate autonomously like plasmids [55]. The elements are also not easy to distinguish through size, with plasmids varying between 746 bp to 2.5 Mbp, ICEs varying between ~20 kbp to more than 500 kbp and phage DNA varying between 3.3 kbp to more than 500 kbp [48, 45, 55]. Very large plasmids, also referred to as megaplasmids, are also hard to distinguish from secondary chromosomes.

Plasmids and other intergenomic MGEs are thought to acquire the genes beneficial to their host through intragenomic MGEs. Intragenomic MGEs include insertion sequences, integrons and transposons. These elements are known to excise themselves from the host genome and insert themselves in another

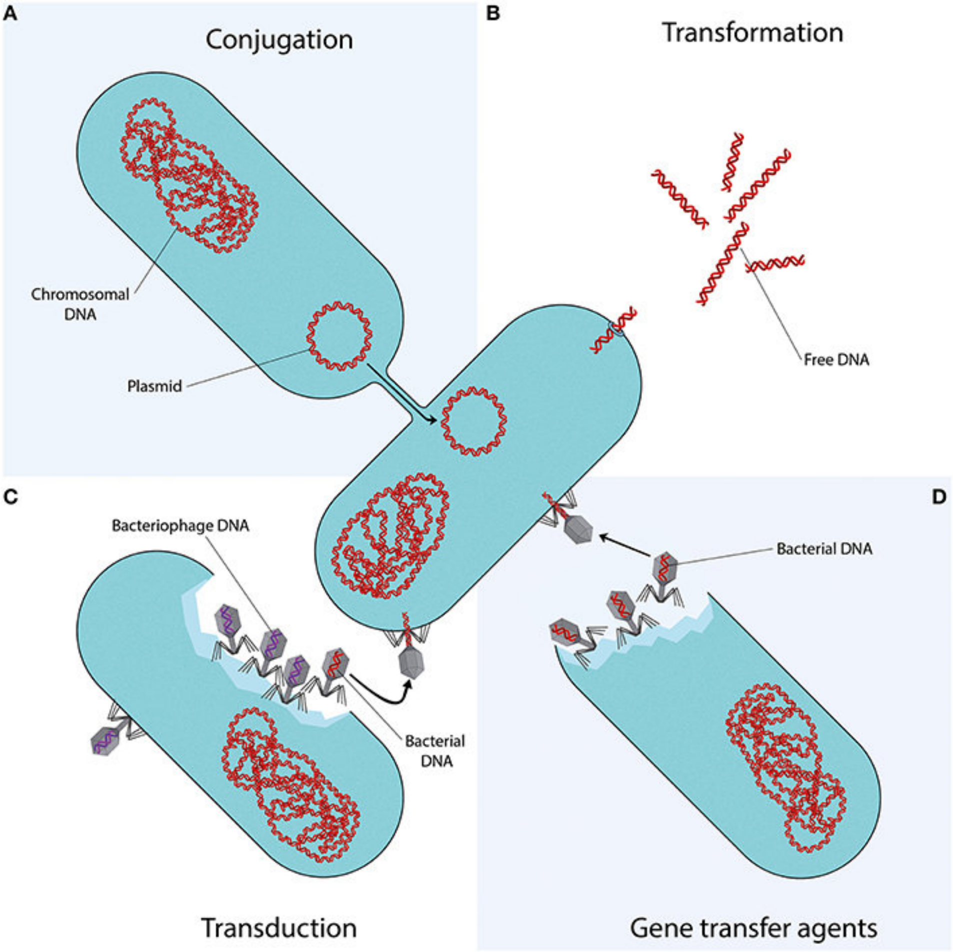

Figure 1: Methods of horizontal gene transfer (image taken from [143]). a) Mobilizable plasmids contain the mechanisms necessary to drive conjugation, in which a donor cell creates a sex pilus through which DNA is transferred; b) Bacteria are able to integrate genetic elements through the intake of genetic material from their environment, e.g. via their porin proteins or outer membrane vesicles. This process is called transformation; c) Bacteriophages, viruses that infect bacteria, can insert their DNA in the host. This DNA can be integrated in the host cell's genome amongst others, but bacteriophage infection can also provide opportunities to spread parts of the host cell's DNA through replacing the bacteriophage DNA with bacterial DNA [13]; d) Gene transfer agents are phage-like particles released by the "host" cell when the cell lyses, and are thought to aid in the adaptation and survival of surrounding cells.

site-specific locus. Aside from the elements that accommodate their movement, they often enable the movement of genes that are beneficial to the cell. However, they are limited in their reach, being confined to the genome of the cell they reside in. Intercellular movement of these elements relies on ICEs, plasmids and phages, in that they can integrate themselves into the DNA of these intergenomic MGEs, which in turn spread to other organisms through conjugation or transduction [30, 103, 81].

Subsets of plasmids have been shown to contain common elements or even a common genetic backbone [127], which gave rise to different plasmid classification schemes. Plasmids can for instance be divided into:

- Immobilizable, if they do not contain any genes encoding plasmid mobility;
- Mobilizable, if they encode an origin of transfer and a relaxase;
- Conjugative/self-transmissible if they encode all genes necessary for their transportation, among which the relaxase and mating pair formation (MPF) genes [119].

Mobilizable and conjugative plasmids can be further divided based on conserved sequences of relaxase and mating pair formation genes, resulting in the so-called MOB-classification. Although relaxase and MPF genes have diverged in ancient types, certain relaxases are found more commonly with certain MPF types than others [38]. Research on these types has found relations between relaxase and MPF types and the diversity of hosts they are found in [38, 119], but these mostly relate to *Pseudomonadota*. Although rare, some plasmids encode multiple relaxases, resulting in them being classified in multiple groups. MOB-classification pertains to the elements allowing for plasmid transfer and affects plasmid host range: immobilizable plasmids lack transfer mechanisms and are narrow host range as a result. Furthermore, some MOB-types have been linked to specific hosts [38, 119].

Incompatibility types have also been used to categorize plasmids into distinct groups. This typing scheme was based on whether two plasmids can co-exist in a cell [24]. Replication genes were historically thought to be the determinant of incompatibility groups, which led to the incompatibility groups also being called replicon types. Although it has been debunked that plasmids of the same replicon type cannot coexist [150], it is still a useful partial determinant for plasmid host range. Some replicon types require host machinery in order to start their replication, whereas others are fully independent, which increases plasmid host range [120, 52]. Furthermore, plasmids have been shown to have broader host range when they fall into multiple replicon types [52].

The replicon, MOB and MPF typing schemes were based on the presence of certain genes without considering co-occurence. Taking this into account led to the development of the plasmid multi-locus sequence typing (pMLST) classification scheme [16]. As plasmids are prone to rapid alterations in their genetic material, they tend to have few shared genes [95]. For different incompatibility typings, the genes that are shared, can be used to further subtyping, leading to the pMLST typing. This *in silico* approach is one currently the most ubiquitous method of classifying plasmids and predicting host range [16].

Plasmids can also be classified based on their function. These functional groups are called F-plasmids, R-plasmids, Col-plasmids, degradative plasmids and virulence plasmids. F-plasmids, or fertility plasmids, enable conjugation [68]. They contain a 60 kbp plasmid backbone and are associated with transposons and integrons.
R-plasmids are known to convey resistance to one or more antibiotics, as well as the resistance transfer factor that enable the transfer of resistance genes [123]. Col-plasmids harbor genes that create bacteriocins, special proteins that kill other bacteria [27]. Virulence plasmids contain virulence genes, which turn the host species, i.e. a bacterium, into a pathogen [56]. Finally, degradative plasmids encode proteins that allow the metabolism of otherwise non-degradable organic compounds [118]. Plasmids can be found categorized into more than one of these groups: it has for instance been found that virulence factors and antibiotic resistance genes are often found on F-plasmids enabling their dissemination [65]. Functional classification of plasmids is less linked to host range, although F-plasmids contain the conjugation machinery that allows dissemination to different hosts.

Plasmid host range is fundamentally determined via conjugation assays [129]. Although this method is accurate, it relies on the costly process of growing the donor and recipient strain *in vitro* and determining whether plasmid DNA is detected in the recipient strain through the use of sequencing. As such, this method has several limitations, as not all bacteria can be cultivated *in vitro* and the methodology to determine host range takes up a lot of time and resources [6]. For this reason, sequence-based approaches to predict host range are preferred in scenarios where time and resources are constrained. *In silico* prediction of plasmid host range currently bases itself around plasmid types and how plasmids with similar genetic backbones, i.e. replicon, relaxase and mpf types, are associated with similar hosts.

Host range is also not static, and it is likely affected by the genetic content of the plasmid, by their hosts and by environmental effects [52, 41, 145], with several factors acting as barriers for the plasmid to establish itself in a host and others aiding the expansion of plasmid host range. The establishment of a plasmid in a new host requires the transfer of the plasmid from a donor to this new host, for which a plasmid either needs to use the conjugation machinery encoded on other plasmids or ICEs in the host or encode their own [119, 38]. Whether transmission is successful is also dependent on the donor-recipient combination [145]: once transferred, the plasmid needs to invade the host's defense system [116, 7]. If the plasmid is not able to replicate itself in its new host, it will be lost as well. If the plasmid cost outweighs the benefit it brings, copy number is too low and plasmid copies are not distributed well during cell division, the plasmid is expected to be lost as well [14].

Once established, the plasmid and host can co-evolve in order to alleviate the costs associated with plasmid maintenance, either through the acquisition of genes resulting from recombination, large-scale deletions or mutation of single genes [44]. Restriction-modification (RM-)systems and toxin-antitoxin (TA-)systems, colloquially post-segregational killing systems, ensure that if plasmid-free daughter cells arise, their survival is prevented [131]. Although these genes have been associated with host range, the do not fully explain it. Certain incompatibility groups are thought to be narrow host range, but plasmids with these have been found in different taxa. *In silico* methods for plasmid host range prediction relying solely on incompatibility, relaxase and MPF types do not achieve sufficient accuracy. Current *in silico* methods provide an accuracy of 80% on the genus level [54]. Therefore, we do not understand which plasmid characteristics influence host range, nor if host range can be explained fully by plasmid content alone.

Due to the plastic nature of plasmids, plasmid phylogeny has recently been proposed to be studied through network analysis [1, 107]. These methods group similar plasmids through creating a plasmid similarity network and using a clustering algorithm, and associate plasmid host range with specific clusters. This network analysis research was not able to find a link between plasmid host range and MOB or replicon types, leaving the suggestion to research combinations of these elements, other phenotypic traits and plasmid evolvability to find host range [107]. Research has suggested that large recombination events can increase host range, as found through studying IncP-plasmids [145]. On the other hand, it has been suggested that single compensatory mutations on certain genes causes the plasmid to conform to their host [44]. Several instances of large conserved regions with similar gene order between distantly related plasmids have been noticed and although authors suggested that these regions serve a purpose, their function remains unclear [107].

As gene-based methods currently base themselves around a genetic backbone of replicon, relaxase and MPF type combinations, we hypothesize that conserved regions based around these genes may aid in more accurately finding plasmid host range. With a synteny-aware plasmid network construction method [127], we are able to uncover if conserved regions based around replication machinery, relaxases, MPF systems and addiction mechanisms can function as better predictors for plasmid host range, and if so, how they do so. Under the assumption that other phenotypic traits that influence plasmid host range would be conserved in combination with the replication machinery, relaxase or MPF system, we set out to find syntenic blocks containing genes influencing this host range. Ergo, this project aims to answer the question of whether syntenic blocks of genes, i.e. regions of neighboring genes in plasmid DNA sequences, have an influence on the host range of plasmids.

0.2. Related work

Traditionally, plasmid host range was thought to be determined by a small set of genes involved in the replication and transfer of the plasmid [52], and is as such the basis for *in silico* plasmid host range prediction. However, other factors have also been found to affect host range, among which the bacterial immune response and plasmid genes that increase host reliance on the plasmid. Plasmid host range is dynamic and evolving, and there is no clearly defined idea of what makes a plasmid BHR. Several studies have researched plasmid host range and what affects it, which will be discussed in this section, starting with the initiation of transfer from a donor to factors aiding in ensuring the plasmid is persistent in their new host, as well as methods that have been developed to predict plasmid host range and drawbacks they have.

0.2.1. Host range-associated genes

Mechanisms of plasmid transmission

As mentioned in Section 1, host range is thought to be partially determined by a plasmid's conjugation machinery. Plasmids can contain one or multiple relaxase enzymes, which are proteins that induce cleavage at a specific site on the plasmid, the origin of transfer (*oriT*). For conjugation, the relaxase-single stranded DNA (ssDNA) complex, also called relaxosome, needs to bind to a type IV coupling protein (T4CP) that recognizes substrates for secretion. This T4CP then links these to the type IV secretion system (T4SS), part of the mating pair formation (MPF) system, for transport to the recipient cell [119, 38]. MPF and relaxase genes are in part the basis for the plasmid classification tools pMLST and MOB-typer, where the genetic backbone of found plasmids are related to the hosts plasmids with this backbone have been found in [16, 109].

Several studies have suggested that the type of conjugation machinery present on a plasmid is an important determinant of plasmid host range, with relaxase and MPF families being associated with specific bacterial taxa [151, 119, 38, 43]. Relaxases are thought to co-evolve with T4CPs, and together they form the basis for plasmid MOB classification [119]. Relaxases and T4CPs are divided into six (super-)families, denoted as MOB_F, MOB_H, MOB_C, MOB_Q, MOB_P and MOB_V [38]. MOB_F relaxases are mainly present in the phyla *Actinomycetota*, *Cyano-bacteria* and *Pseudomonadota*. MOB_H relaxases are associated with *Gammaproteobacteria* and *Betaproteobacteria*. The MOB_C relaxase family is found on MGEs from *Gamma-proteobacteria, Baccilota* and *Mycoplasmatota*. MOB_Q, a relaxase family that overlaps with the MOB_P family, is mainly found on plasmids in the orders *Hyphomicrobiales, Nitrospirota, Bacillota, Pseudomonadota and Actinomycetota*. Subfamily MOB_{Q2} is associated with *Agrobacterium* TI plasmids. MOB_P is the largest relaxase superfamily with unclear boundaries, although most have been found in *Pseudomonadota*, but also in *Firmicutes* and *Bacteroidetes* [22]. Finally, the MOB_V family is generally found in *Bacillota* and *Bacteroidota*, but also found in *Pseudomonadota, Cyanobacteria* and *Spiro-chaetes*[38, 119].

Like relaxases, MPF systems are also divided into different types based on the divergence in their amino acid sequences, namely MPF_F, MPF_G, MPF_I and MPF_T. Although MPF systems are necessary for the creation of the sex pilus for conjugation, it was noted that these groups have mainly been found and classified in the order of *Pseudomonadota* [119]. In a paper that compared host range prediction tools of complete plasmids, tools relying on these genes for classification do not achieve accuracy higher than 50% on the genus level [54]. Furthermore, classification of MPF types is limited to conjugative plasmids, with classification of relaxase types furthermore being limited to mobilizable and conjugative plasmids, and both cannot be applied to immobilizable plasmids. These immobilizable plasmids should not be swept under the rug, as it has been found that they can be mobilized through using the conjugation machinery of other plasmids [104].

Although relaxase and MPF types were thought to determine (in particular transfer) host range, it was found that they are not the sole factor in determining a plasmid's replication host range. *OriT's*, T4SSs and plasmid hosts also affect the replication host range [151, 43, 5]. For instance, constructed minireplicons of plasmids, in which non-essential genes were removed from their composition, were found to be transferred less often in *E. coli* than in *P. putida*. In the same study, plasmids could be transferred from *P. putida* to *E. coli* and back, from *E. coli* to *C. pinatubonensis* but not from *P. putida* to *C. pinatubonensis* [145]. To conclude, relaxase and MPF genes are thought to be important factors in broadening host range and are as such used in host range prediction, but knowledge on them is scarce outside of *Enterobacterales*. Furthermore, it has been found that immobilizable plasmids that miss these genes can be BHR as well [38, 104].

Barriers of entry for plasmids

Some works have shown that replication and conjugation machinery alone are not sufficient in predicting whether a plasmid can persist in a given host. Both bacterial chromosomes and other plasmids encode several mechanisms that prevent a novel plasmid from establishing itself. An abundance of ssDNA in a bacterium, for instance after the conjugation of a plasmid, activates the bacterial SOS response, which activates genes that are involved in DNA damage repair and can lead to recombination and mutagenesis [8]. Among the many different effects of this SOS response, several genes, such as RM-systems, degrade an incoming plasmid which prevents the establishment of the plasmid in its new host [116]. Recombination of gene cassettes is also promoted, inducing mutations in the genome (including the plasmid) [8]. The SOS response also promotes the transcription of error-prone DNA polymerases, resulting in increased mutation rates [9]. It is thought that the resulting mutations and recombinations create opportunities for plasmid cost amelioration through the inactivation of costly genes and deletion of genomic signatures incompatible with the host. On the other hand, several plasmids carry the gene *psiB* that prevents the activation of the SOS response and antirestriction proteins that evade the effects of the bacterial restriction-modification system, in turn ensuring the plasmid survives in an unchanged form [116]. These factors affect plasmid host range either by altering the genes contained on the plasmids resulting in plasmid cost amelioration, or through inhibiting this alteration and ensuring establishment of the plasmid in the recipient cell.

Like interactions between plasmids and their host alter host range, co-residing plasmids similarly affect plasmid establishment: as mentioned, if the plasmid shares the same replication and/or partition control with another plasmid in the cell, the plasmids will compete for their presence in the host. This incompatibility may be unidirectional or reciprocal based on features of the plasmids [128]. Plasmids have also been shown to encode CRISPR-Cas systems, with most of them targeting plasmid-like elements, specifically those with similar properties as the plasmid that carries the CRISPR-Cas systems, suggesting that these systems are encoded on plasmids for interplasmid competition [101, 25, 100]. Plasmids can also encode proteins that block further conjugation, preventing new plasmids from being established in the host [34]. On the other hand, plasmids have been found to co-operate, with some plasmids stabilizing the presence of other plasmids [148, 142].

Plasmid host range should thus not be thought of as a static plasmid feature, but it is a dynamic and evolving characteristic. A plasmid may not replicate easily upon introduction in its new host, but it can mutate and recombine to ameliorate its cost and ensure survival. Other factors present in the host bacterium may both stabilize and destabilize the establishment of the plasmid. Although no tools explicitly use information of the defense mechanisms of hosts and other plasmids in order to predict plasmid host range, tools using similarity of plasmids such as COPLA [108] may be affected by their presence.

Replicon types

Replicon types, a classification based on different types of replication protein, are also an important determinant in plasmid host range, as a plasmid should be able to replicate itself in order to establish itself in a new host. Plasmid replication happens at the origin of replication, which needs to be recognized by replication initiation proteins. The genes for these replication initiation proteins can either be carried by the plasmid itself or can be found on the host genome [117]. Replication machinery falls into several so-called replicon groups, called as such as it was thought two plasmids with the same replication machinery could not co-exist in a cell [24]. Detection of these incompatibility groups relies on the conserved sequence of replication initiation proteins [117, 145]. There are many different incompatibility groupsand research has also found many which are still unclassified [117]. Several of these incompatibility groups have more divergent backbone genes, such as IncP-1, IncP-9, IncQ, which are thought to be more often BHR as a result. Others are less divergent, such as IncA/C, IncX, IncI and IncP-7 [117, 145, 57], and their host range is more narrow [64]. Replicon types are used in many different host range prediction tools, among which MOB-typer, pMLST, HOTSPOT and PlasmidHostFinder [110, 16, 54, 6]. These types have been classified independently in different genera and some of these overlap. For instance, the IncP-1 group in *Pseudomonas* is identical to the IncP group in *Enterobacteriaceae*, as for IncP-3 and IncA/C, IncP-4 and IncQ, and IncP-6 and IncG/U, suggesting that plasmids in these incompatibility groups are BHR on the class level.

Considering the large amount of different incompatibility groups, listing all of them and their known influence on plasmid host range would be a task outside of the scope of this thesis. Furthermore, not all of the incompatibility types are fully documented in their host range. Nevertheless, we will discuss several of the more important incompatibility types that are thought to be BHR below.

IncP/IncP-1 plasmids
IncP plasmids are well-studied plasmids known for being broad host range (BHR), being found in the different classes *Gammaproteobacteria, Betaproteobacteria* and *Alphaproteobacteria*. Their sizes vary, with some being about 35 kbp and others larger than 500 kbp [75]. They have been subgrouped based on differences in the transfer region of these plasmids. In a study on the host ranges of these plasmids, it was found that although host range does overlap between subgroups, they are not the same: IncP-1γ plasmid backbone had genomic signatures that were found in *Gammaproteobacteria* and could establish themselves well in *Gammaproteobacteria* and *Betaproteobacteria*, whereas they could not in *Alphaproteobacteria*. IncP-1β plasmid backbones were associated with *Betaproteobacterial* genomic signatures and could transfer and reside in all bacteria of these classes [91]. The acquired genetic signatures are thought to be the result from recombination with genes from different hosts that aid in host adaptation [91]. Although the backbones of these plasmids may partially explain their host range, the levels of replication initiation protein was suggested to affect plasmid host range as well [145].

IncW plasmids
IncW plasmids are small, encode a wide spectrum of ARGs and are more conserved in their backbone than IncP plasmids. They are associated with the different *Pseudomonadota*, i.e. *Alpha-, Beta-* and *Deltaproteobacteria*, and *Bacteroidetes* [75, 145]. Despite them being linked to *Pseudomonadota*, it has been shown that specific bacteria of the order, specifically *E. coli*, were not able to harbour these plasmids [33].

IncP-3/IncA/C plasmids
IncA/C plasmids are large, low-copy plasmids. Their host range is thought to be broad as they do not contain genomic signatures that are specific to a certain species [57]. However, plasmids of this incompatibility group are mainly found in *Gammaproteobacteria*.

IncQ plasmids
IncQ plasmids are approximately 5 to 15 kbp, mobilizable through a diverse set of mobilization machinery and found in a wide diversity of hosts, including both gram-positive and gram-negative bacteria, as well as cyanobacteria [76, 106].

Length, GC content, environment and co-evolution

Plasmids vary heavily in size, GC content and genetic content, which influences the host range of a plasmid. Several plasmids are even known to be larger than some bacterial chromosomes [45]. Plasmids generally tend to conform to their host, with hosts with smaller chromosomes containing smaller plasmids [45]. Similarly, it has been found that linear plasmids are more often found in hosts with linear chromosomes [135]. GC content also plays an important role in host range, as most narrow-host plasmids are more found in hosts that have similar GC content as the plasmids [89, 4], although it was also concluded that it was not the sole determinant for host range.

Several authors have also proposed that the environment a host finds itself in also affects a carried plasmid's host range, with different temperatures and the collection of hosts in an environment affecting conjugation [145, 66, 80, 146, 63]. The presence of antibiotics and sterilizing agents in non-lethal doses, as well as common non-antibiotic pharmaceuticals such as anti-epilepsy drugs, has been found to increase mutation and alter conjugation rates [78, 74, 141, 140, 147, 149]. Exposure to antibiotics led to mutations in the *copA* gene controlling plasmid replication, and led to a rise in plasmid copy number, increasing both resistance and conjugation rates despite the absence of susceptible hosts [29].

Both chromosomal mutations as well as mutations on plasmids can affect host range [46, 47, 87, 44, 77, 121, 122]. Mutations on the plasmid sequence affect the plasmid host range in the long run, as these mutations are passed on to a plasmid's “offspring". The phenotype of plasmids as a result of mutation suggests a trade-off between parasitic and mutualistic lifestyles, with plasmids either decreasing their mobility and relying on vertical inheritance or increasing their conjugation rate and spreading through horizontal transmission [121, 77, 102, 12]. For instance, a study on the evolution of plasmid stability found that changes in the replication initiation protein of an IncP-1 plasmid stabilized plasmids in *Shewanella oneidensis* but meant the plasmid could no longer replicate in *Pseudomonas* [121]. It was also found that acquisition of a TA-system and a cointegrate resolution system by an unstable Inc-P plasmid in different hosts allowed plasmid persistence in these hosts [77]. Similarly, deletions of conjugation machinery has been shown to stabilize plasmids in a new host [102]. As mentioned for IncP-plasmids, it is thought that recombination with genes from *Gamma-* or *Betaproteobacteria* leads to host adaptation to these classes [91]. More generally speaking, it is thought that plasmids that incorporate antibiotic resistance genes have a more diverse host range and act more like bacterial parasites [73].

All in all, the host range of a plasmid is not determined by a single feature of the plasmid or the host. A combination of plasmid-encoded genes, host-associated genes, GC content, length, evolutionary adaptability and environmental effects affect plasmid host range. Fundamental determination of plasmid host range is done through the use of conjugation assays [129] . This accurate method relies on growing a donor and a recipient strain *in vitro* and determining whether plasmid DNA is detected in the recipient strain through the use of sequencing, making the task both time- and resource expensive [6] and not applicable to all organisms as some cannot be cultivated. As a ways to lessen the cost of host range determination, *in silico* approaches, where combinations of these features are used for host range prediction, are preferred in scenarios where time and resources are constrained.

0.2.2. In silico approaches

In silico prediction of plasmid host range currently bases itself around plasmid types and how plasmids with similar genetic backbones are associated with similar hosts. MOB-suite uses its own plasmid similarity clusters, created through single-linkage clustering, as a reference to find the host range of a query plasmid. This tool uses the query's replicon type and relaxase type as well as the cluster the query plasmid is most similar to, to predict host range based on the diversity of hosts plasmids in the clusters were found in [110]. Another method is COPLA, which predicts which Plasmid Taxonomic Unit (PTU), a plasmid group as found through hierarchical stochastic block modeling, the plasmid belongs to. PTUs were created through clustering a plasmid similarity network, but this clustering was stated to be done manually [107]. PTU's are then associated with the set of hosts these plasmids are found in in literature, which is then output [108]. PlasmidFinder uses the frequency of oligonucleotides in plasmids to predict host range by comparing oligonucleotide frequencies between plasmids and hosts [16], using the findings that plasmid oligonucleotide frequency is similar to its host. PlasmidHostFinder uses a combination of oligonucleotide frequency, found replicons, relaxases and MPF systems and more fed in a random forest model to make plasmid host range predictions [6]. Finally, HOTSPOT uses a transformer-based classifier to predict host range based on replicon, relaxase and MPF type as well as computationally predicted proteins [54].

Although these plasmid features aid in predicting plasmid host range, the authors of HOTSPOT found that host range prediction of complete plasmids maxed at ~80% in predicting host range on the genus level [54], with HOTSPOT being the most accurate, followed by PlasmidHostFinder (~ 50%) and MOB-typer (~ 30%). Although plasmids are thought to contain a certain genetic backbone based on their replication and conjugation machinery[108, 1], studying whether this is conserved between different plasmids has thus far not been done Furthermore, certain conserved regions were found in different plasmids although their functional origin is still unclear [107]. We suggest that surrounding genetic material may affect the function of conjugation and replication machinery and that conserved regions based around these genes as well as addiction modules may give us more information on whether or not this is the case. A method to find recently transferred conserved regions containing ARGs was developed [127]. This method utilizes a seed-and-extend algorithm where seeds are antibiotic resistance genes. We used this method to find syntenic sequences among plasmids with replicons, relaxases, addiction mechanisms and MPF systems used as seeds and test whether using syntenic sequences as opposed to only found genes may increase plasmid host range prediction accuracy.

0.3. Materials and Methods

A graphical abstract of the methodology we followed is shown in Figure 2. How we performed our analysis is furthermore explained in the following subsections.

0.3.1. Dataset acquisition and curation

The latest version of the Plasmid Sequence Database (PLSDB) was downloaded on November 10, 2023 [37, 113]. This publicly available database was created by collecting records from the NCBI nucleotide database, specifically from RefSeq and INSDC, which includes records from DDBJ, EMBL-EBI and GenBank [72, 61]. The database contains only contains plasmid sequences classified as complete and derived from complete assemblies. Alongside the sequences, PLSDB also provides the taxonomy the

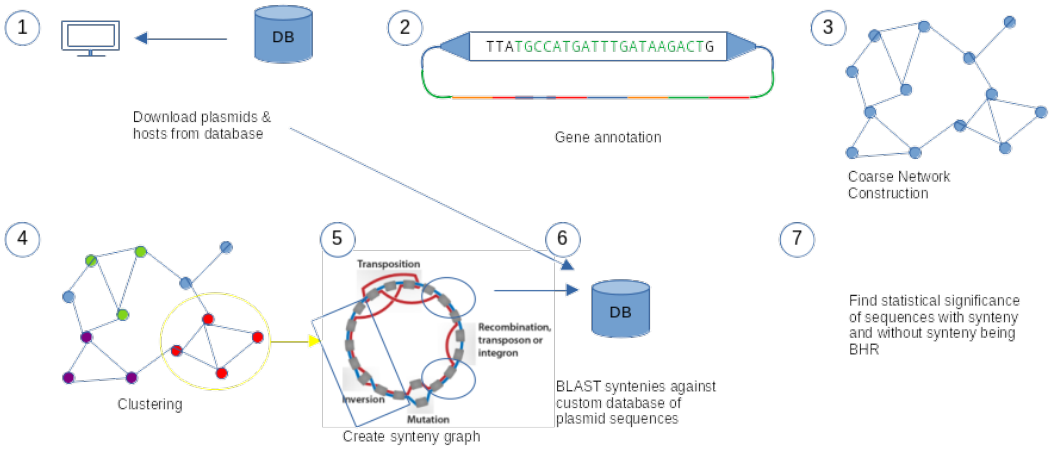

Figure 2: Methodology for finding conserved regions of broad host range plasmids. 1) Obtaining plasmid sequences from PLSDB; 2) Performing gene annotation on the obtained sequences; 3) Creating a plasmid similarity network; 4) Clustering the plasmid similarity network; 5) Running SHIP to find conserved regions based around both origins of replication and transportation, relaxases, toxin-antitoxin systems and restriction-modification systems; 6) Obtaining the sequence of the conserved regions, BLASTn against all the plasmid sequences; 7) Finding statistical significance between plasmids containing a conserved region and host range.

plasmid is derived from to the species level, the sequence lengths, GC% and the detected replicon type. Furthermore, the authors included a list of pairs of plasmids with low mash distance (i.e. 950/1000 shared hashes).

In order to ascertain that the plasmid dataset PLSDB contained only complete plasmids, we filtered out sequences outside of the 746 bp to 2.75 Mbp length range, as these were the sizes of the smallest and largest plasmids found in reviewing literature [21, 45]. To make the ground-truth host range more reliable, plasmids from species with less than five plasmids in the dataset were also removed.

0.3.2. Annotation

Genes in the plasmid sequences were re-annotated with Bakta v1.9.1. Bakta requires an annotation database and provides the option to choose between a full and a light version of the annotation database [115]. We used a locally downloaded version of the full database and the options “complete" to ensure all sequences were deemed complete replicons, “keep-contig-headers" and “skip-plot". We performed replicon, relaxase and MPF typing using MOB-typer (from MOB-suite v3.1.7) with default parameters [110]. Alongside the previously mentioned filters for length and plasmid hosts, plasmids for which Bakta found less than five genes were removed, as these plasmids did not contain enough genes to find conserved regions with SHIP. The resulting sequences were further annotated using NCBI’s AMRFinderPlus (v3.11.26) to find ARGs using the sequences and the General Feature Format (GFF) and protein FASTA files output by Bakta [32]. The proteins from the genes discovered by Bakta were clustered using CD-HIT (v4.8.1) with a threshold of 90% protein similarity, a word length of 5 and the length of the description in the .clstr file set to 0 [71, 70, 36]. Plasmids were considered to be broad host range (BHR) when there was another sequence in the dataset with which it shares more than 950/1000 shared hashes, as found in the supplied file mentioned above, and the sequences were derived from different hosts on the species level.

0.3.3. Building a plasmid similarity network

Similarity between plasmids was calculated using Fast-ANI (v1.33) with the fragment length set to 200 bp to allow comparison of the smallest plasmids in our dataset [53]. FastANI compared all plasmids to one another, causing every two different plasmids to have two different similarity values depending on which plasmid was used as query and as reference. To make the network undirected, the MCL-edge toolkit (v22.282) was used. With mcxload, the plasmid distances in tsv-format were loaded in such a way only the edge with the smallest weight was kept. This was followed by using mcxdump with the 'dump-upper' flag such that each similarity between two plasmids was only output once. Weights in the network were then 0-1 normalized to improve clustering. The plasmid similarities were then used to construct a plasmid similarity network, both in networkx (v3.2.1) and igraph (v0.11.3). We analyzed the degree distribution of nodes through the use of the powerlaw module (v1.4.6), through fitting the network degree distribution with a power law distribution, log normal distribution and binomial distribution using the loglikelihood ratio.

0.3.4. Finding plasmid characteristics associated with host range

We used the Python packages SciPy (v1.11.4) and Pandas (v2.1.4) to find correlations between plasmid host range and features of plasmids contained in the dataset [85, 134, 51]. Characteristics analyzed were plasmid host taxonomy, sequence length, GC content and coding density. We also analyzed replicon, relaxase and MPF types. As plasmids can fall into multiple replicon and relaxase types and MOB-typer had detected one type multiple times for one plasmid, we kept only unique types for plasmid records. When analyzing singular replicon and relaxase types and how these were associated with BHR, we counted each replicon/relaxase type once for each plasmid. To find the significance of relations between host range and categorical variables, the Chi-squared test and Cramer's v were used, whereas we used the two-sided student's t-test and Cohen's d in order to find relations between host range, sequence length and GC content. The annotations from Bakta were used to find the number of genes a plasmid encoded, which was then used to find the coding density of plasmids by dividing the number of genes on a plasmid by its sequence length in kbp. To find specific (combinations of) categorical variables, a post-hoc corrected Chi-squared test was used.

0.3.5. Clustering the plasmid similarity network

We clustered the plasmid similarity network with two clustering methods: Markov and Leiden clustering. Markov clustering was performed with inflation parameters $\{1.1, 1.2, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$; We also clustered the network through the Leiden community detection algorithm [130]. Experiments were performed with different maximum cluster sizes. Final results were gathered after clustering with Leiden clustering using the modularity vertex partition, 10 iterations and a maximum community size of 500. The clustering performance of these algorithms was evaluated based on the v-measure with respect to replicon types, relaxase types and MPF types. The v-measure is the harmonic mean of homogeneity and completeness, which in turn are measurements based on the conditional entropy of the classes given the cluster assignments, defined as:

$$H(C|K) = -\sum_{c=1}^{|C|}\sum_{k=1}^{|K|} \frac{n_{c,k}}{n} \cdot log(\frac{n_{c,k}}{n_k}),$$

and the entropy of the classes, defined by:

$$H(C) = -\sum_{c=1}^{|C|} \frac{n_c}{n} \cdot log(\frac{n_c}{n})$$

Homogeneity is then given by:

$$h = 1 - \frac{H(C|K)}{H(C)},$$

whereas completeness is given by:

$$c = 1 - \frac{H(K|C)}{H(K)},$$

with finally the v-measure being given by:

$$v = 2 \cdot \frac{h \cdot c}{h + c}$$

where $n_{c,k}$ stands for the number of samples with label c in cluster k, n_k is the number of samples in cluster k, n is the total number of samples and n_c is the total number of labels[111].

0.3.6. Finding syntenic regions associated with host range

We searched for syntenic regions associated with BHR plasmids within each cluster in the plasmid network individually. We clustered the protein sequences for each cluster separately using CD-hit (v4.8.1). These regions were found using SHIP (v0.2.6) [127]. Genes with functional annotations containing at least one of the terms "relaxase", "replicon", "replication", "origin", "toxin", "restriction", and "modification" were selected as seeds for SHIP. We kept all regions with a minimum average plasmid distance of 0.1, found in at least 2 plasmids, and containing 5 to 9 genes. To find all plasmids containing each syntenic region found by SHIP in the complete plasmid dataset, each region was searched among all plasmid sequences with BLASTn (v2.15.0), with a maximum of 10,000 target sequences. Hits with less than 90% identity and less than 90% coverage were filtered out. We tested if each of the regions found by SHIP was associated with BHR using Bonferroni-corrected Chi-squared tests. More specifically, we compared the number of BHR and single host plasmids in the complete dataset containing said region against the expected number of plasmids under the null hypothesis that the region was uniformly distributed across BHR and single host plasmids. AMR genes contained in the conserved regions found using SHIP were identified using CARD databases v3.2.9 and Resistance Gene Identifier v6.0.3 [3]. The three conserved regions with a p-value lower than 0.05 and the highest Chi-squared statistic, as well as the one conserved region that was found across different phyla, were analyzed on their gene content. For every individual gene that was found in the conserved region, we performed the same Chi-squared test as with the conserved region, with the null hypothesis that the gene was distributed uniformly across BHR and single host plasmids. Furthermore, we analyzed the correlation between host range and the combination of detected replicon, relaxase and MPF type of the plasmid, to see whether the conserved region was a better indicator than these genes.

0.4. Results

0.4.1. Analysis of plasmid characteristics

Before looking for syntenic regions associated with broad plasmid host range, we assessed if several plasmid characteristics were correlated with host range by inspecting the dataset of plasmid sequences used in this work. We analyzed plasmid characteristics that are currently used in state-of-the-art host range predeiction. These characteristics include length, GC%, coding density and replicon, relaxase and MPF type. Aside from these characteristics, we also inspected the number of ARGs on the plasmids from PLSDB using AMRFinderPlus and found 357 unique ARGs 109,548 times across the 50,667 plasmids.

Larger plasmids tend to be single host

After filtering out sequences with less than five genes, sequences for which the host supplied less than five sequences and outside of the sequence length range as mentioned in section 0.3, Materials and Methods, the dataset contained 50,667 plasmids, of which 7,715 were BHR and 42,952 were single host. Sequence lengths for complete, BHR and single host datasets are shown in 3. Literature research had found that a relationship exists between plasmid length and their host range, with very large plasmids (> 500 kbp) usually being immobilizable and narrow host range [119]. For smaller plasmids, this distinction was less obvious. A causal relationship between plasmid length and host range was not found in neither literature nor our work.

We studied the relation between plasmid size and its host range through a two-tailed student's t-test and determined the relation between sequence length and host range classification to be significant with a p-value of $p < 10^{-5}$ but with a Cohen's d of 0.27 directed towards single host plasmids being larger. This suggests that although the difference is deemed significant, the effect size is deemed small and not enough to indicate whether or not a plasmid is BHR or not. We hypothesize that this is the effect of large plasmids behaving more similar to bacterial chromosomes, as was suggested in Hall et al [45].

GC content deemed significantly correlated with host range

We investigated the spread of GC content in BHR and single host plasmids, a boxplot of which can be viewed in Figure 4. The GC of single host plasmids varied between 14.0% to 74.9%, with an average of 46.0%, whereas BHR plasmids varied between 24.1% and 69.6%, with an average of 48.2%. Literature review had found that a plasmid's GC content affects its establishment in a new host [28], and that a plasmid's GC content conforms to their host due to mutation effects and is generally slightly lower [9, 89]. With a large proportion of plasmids within the dataset being hosted by *Enterobacterales*, whose GC

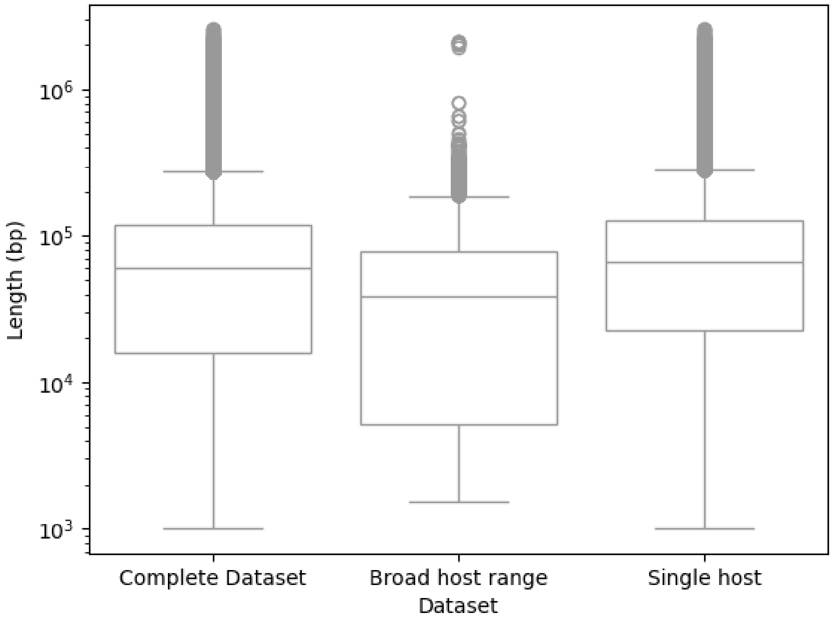

Figure 3: Sequence lengths in basepairs in the complete, BHR and single host datasets. The single host plasmids varied in size from 1.0 kbp to 2.6 Mbp, on average being 66.9 kbp, whereas the BHR plasmids varied in size from 1.6 kbp to 2.1 Mbp, on average being 39.1 kbp.

content is around 50% [79], this explains why the average GC content of plasmids is between 45 and 48%.

Through a two-tailed student's t-test, we found that GC content correlated significantly with plasmids being BHR or not (p-value of $< 10^{-5}$). The Cohen's d was determined to be 0.23 with BHR plasmids generally having higher GC content. Ergo, although plasmid GC content was thought to be important in determining a plasmid's host range, the low effect size suggests that it would not be enough to determine plasmid host range on its own. Because of the narrower range in GC content for single host plasmids, we hypothesize that BHR plasmids are proportionally more often around ~50% GC content to improve their chance at being maintained in as many hosts as possible. Single host plasmids being more accustomed to the GC content for their specific host, which corroborates the literature found [9, 89, 28].

Coding density correlated with host range

Smillie et al. showed that the coding density of large plasmids is higher than the coding density of smaller plasmids [119]. Furthermore, larger plasmids tend to be non-mobilizable and not BHR [119]. We investigated whether or not this plasmid characteristic would suffice in determining whether a plasmid is BHR or not. The distribution of coding density for different plasmids is shown in 5.

Corroborating literature, we found through a two-tailed student's t-test and Cohen's d that the coding density of BHR plasmids is significantly less than that of single host plasmids ($p < 10^{-5}$, Cohen's d = 0.48). Although the Cohen's d shows that this plasmid characteristic is not enough to fully determine that a plasmid is BHR or not, it does reveal that BHR plasmids on average contain fewer genes than single host plasmids.

Several species host more BHR plasmids

We assessed in what hosts the plasmids in our dataset were found in. The 14 hosts that supplied the most plasmids and the number of plasmids is given in Figure 6. *Enterobacterales* make up a 56.6% of our complete dataset, 84.1% of our BHR dataset and 51.6% of our single host dataset. The clinically relevant ESKAPEE pathogens make up 49.2%, 65.4% and 46.3% for the complete, BHR and single host datasets respectively. In total, there are 877 species from 265 genera in the complete dataset, 277 species from 73

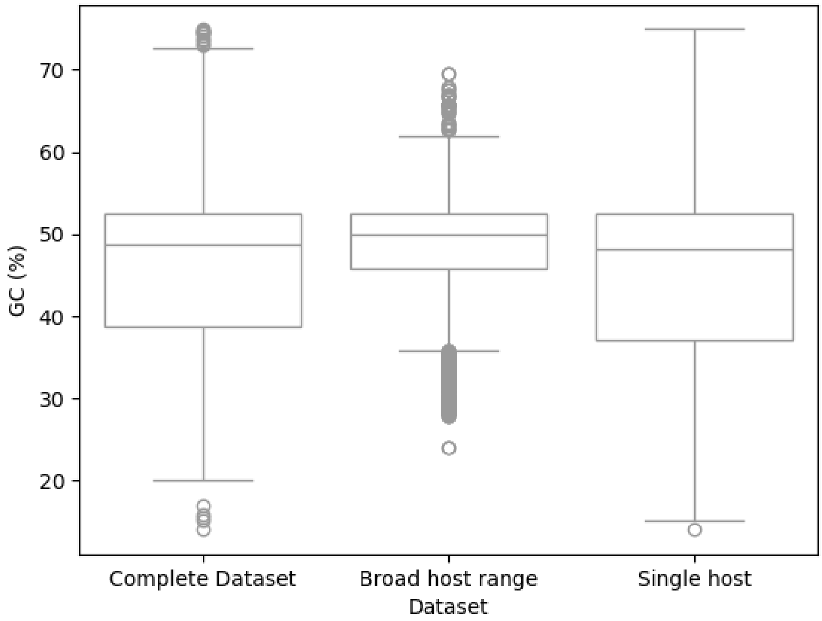

Figure 4: Guanine/Cytosine (GC) content of the sequences in the dataset.

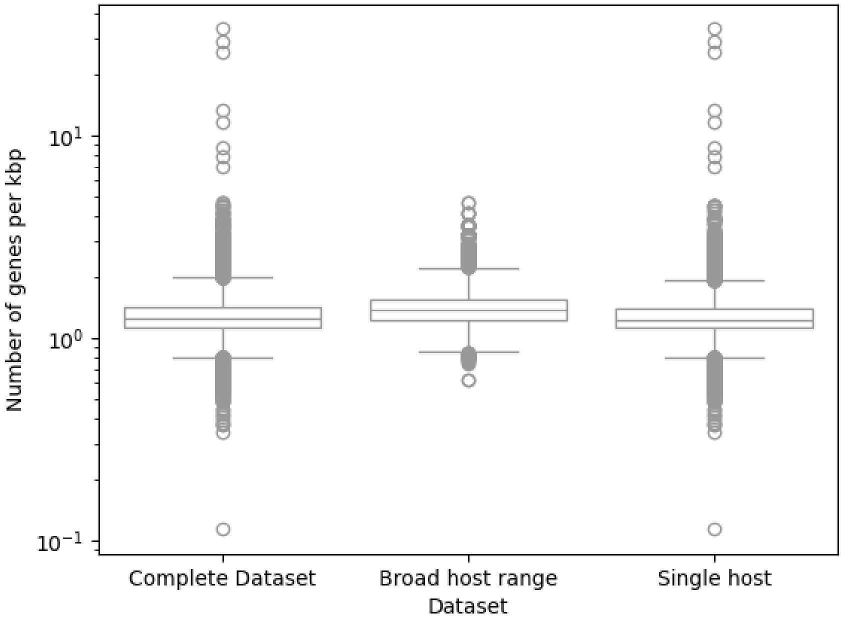

Figure 5: The coding density of the sequences in the complete, BHR and single host datasets, measured in number of genes per kbp.

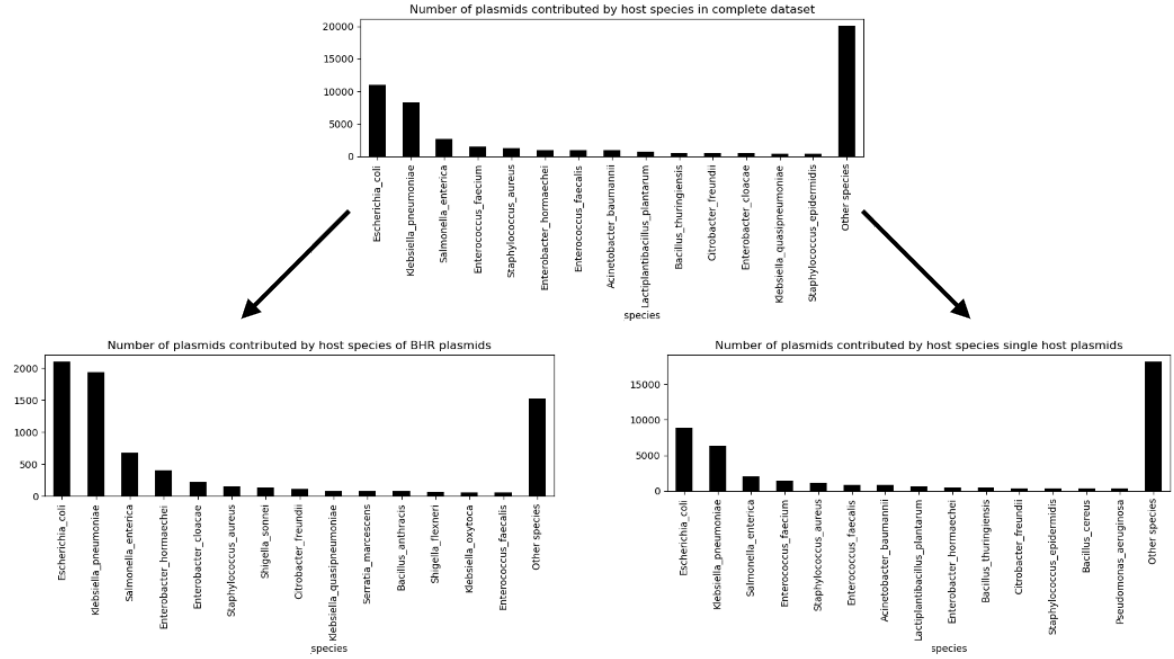

Figure 6: Number of plasmids per host species in the complete dataset, broad host range (BHR) plasmids and single host plasmids.

genera in the BHR dataset and 876 species from 265 genera in the single host dataset.

Through a Chi-Squared test, we found that several species were more associated with BHR plasmids ($p < 10^{-5}$) through a post-hoc corrected Chi-squared test, namely *Acinetobacter sp., Bacillus anthracis, Bacter-oides thetaiotaomicron, Citrobacter freundii, Citrobacter kose-ri, Enterobacter cloacae, Enterobacter hormaechei, Escherichia coli, Escherichia fergusonii, Escherichia sp., Klebsiella aerogenes, Klebsiella michiganensis, Klebsiella oxytoca, Klebsiella pneumoniae, Salmonella enterica, Salmonella sp., Serratia entomo-phila, Serratia marcescens, Shigella flexneri, Shigella sonnei* and *Tolypothrix sp*. Most of these species belong to the order of *Enterobacterales*. It may be the case that plasmids found in *Enterobacterales* are more often BHR, but it is more likely a consequence of the overrepresentation of this order in our dataset.

Although the broad host range category was defined as plasmids being found in different species, we investigated the promiscuity of BHR plasmids between other taxonomic ranks as well, resulting in Figure 7. This Figure demonstrated that most BHR plasmids navigated to species of different genera, i.e. confined in the same family, and plasmids spreading between different orders, classes and phyla were far less abundant.

All in all, plasmids in the dataset were more likely confined to species of the same family, and the more distantly related bacteria were, the fewer plasmids they shared. As we defined BHR plasmids to be able to spread to distinct species, this suggests that a lot of the BHR plasmids we found are exchanged between different members of the order *Enterobacterales*. It also suggests that plasmid exchange is a common occurrence for bacteria in the same family.

Specific replicon, relaxase and MPF types correlate with host range

As has been mentioned in section 0.2, Related work, state-of-the-art gene-based *in silico* host range prediction relies on finding a query's replicon, relaxase and MPF types before outputting in which species these type combinations have been found [110]. We assessed the correlation between these types and plasmids being BHR.

Individual replicon types were found to correlate significantly with plasmids being BHR through a Chi-Squared test with a rather large effect size ($p < 10^{-5}$, Cramer's v = 0.48). Through a Bonferroni-corrected Chi-squared test we investigated if certain of these genes were disproportionally associated with BHR plasmids. We found 34 replicon types that were associated with BHR, of which most

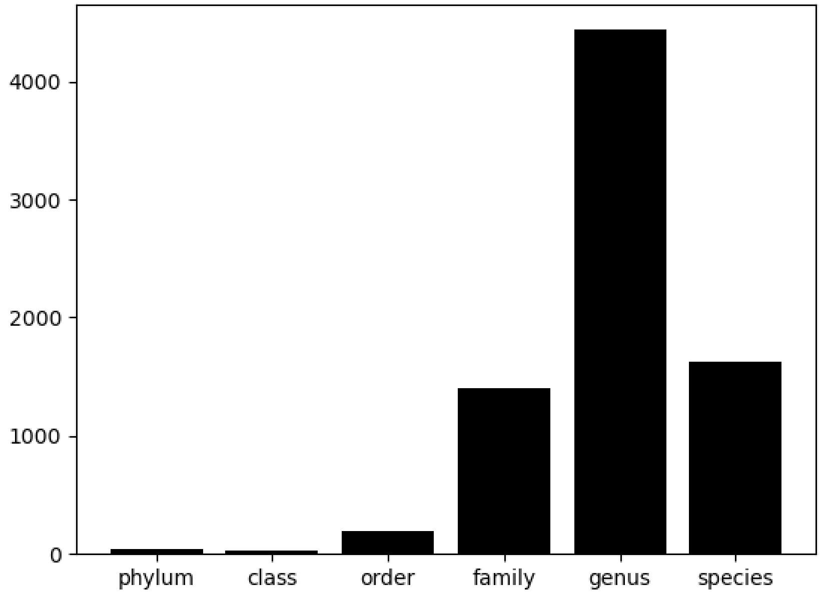

Figure 7: Number of BHR plasmids per definition of BHR in the dataset. The bars exclude plasmids from higher taxa, meaning that, for instance, plasmids present in different genera are not included in plasmids present solely different species.

have are corroborated by literature [126, 98]. On the other hand, we found that IncI, IncX3, IncX4 and IncX6-plasmids were BHR, although they were generally thought to be narrow-host range in literature [126, 125, 58, 69]. The IncX replicons are generally thought to be conjugated between different *Enterobacterales* [126]. The IncI plasmids too were found mainly in *Enterobacteriaceae*, but they were recently also discovered in environmental bacteria [35].

As research has found that multiple replicons on a plasmid broadens their host range [52], we also checked whether unique combinations of replicon types correlated with plasmid host range through a Chi-squared test and Cramer's v. Replicon types were found to correlate significantly with a large effect size ($p < 10^{-5}$, Cramer's v = 0.57). Once again, most of the found replicon types that were also present in replicon combinations were thought to be BHR, but some exceptions that literature thought to be narrow were IncFIA, IncFIB and IncFIC [105]. A full list of replicons and replicon combinations, their observed frequencies and their expected frequencies for being single host and BHR is given in the appendix 0.5. Even though some replicon types and replicon type combinations occurred significantly more often on BHR plasmids than single host plasmids, they were not unique to BHR plasmids, suggesting that replicon type is only a partial determinant of plasmid host range.

Following this, we assessed whether or not relaxase types and combinations were associated with BHR plasmids. We found through a Chi-squared test and Cramer's v that individual relaxase types correlated significantly with plasmids being BHR, but with a medium small effect size ($p < 10^{-5}$, Cramer's v = 0.23). The same was found for relaxase type combinations, although with a larger effect size (Cramer's v = 0.25). This effect size suggests that relaxase types are only a partial determinant for plasmid host range. Through a Bonferroni-corrected Chi-squared test we found that the MOB_C, MOB_P, and MOB_H type relaxases were associated with BHR plasmids ($p < 10^{-5}$). Investigating plasmids with multiple relaxase types revealed that plasmids with multiple MOB_P and MOB_H relaxase types as well as MOB_V relaxases also correlated with plasmids being BHR. That these relaxase types are thought to occur more often in BHR plasmids, is largely corroborated by literature. MOB_H relaxases are found in the different classes *Beta-* and *Gammaproteobacteria*, whereas MOB_C relaxases are known to span different phyla. MOB_P relaxases are very diverse and contains many relaxases that are still currently evolving. Members of the MOB_P relaxase superfamily have been found across many different bacterial phyla [38]. MOB_V plasmids are also found across different phyla, mainly *Bacillota* and *Bacteroidetes*, but also in *Pseudomonadota*, *Cyanobacteria* and *Spirochaetes*.

Finally, MPF types were investigated and it was found that although a significant correlation between these and BHR was found ($p < 10^{-5}$), the effect size found was rather small (Cramer's v = 0.14). MPF types that found to be associated with plasmids being BHR were those of type T and type I, of which the first is associated with the phylum-crossing IncP plasmids whereas the latter is more associated with plasmids from *Enterobacterales* [119]. The table of replicon types is also shown in the appendix 0.5.

We also looked at the use of replicon, relaxase and MPF type combinations as is currently used in state-of-the-art gene-based *in silico* classification methods, and found through a Chi-squared test that these correlated significantly with a large effect size ($p < 10^{-5}$ Cramer's v = 0.60). However, literature review has measured that these genes are not good enough to predict plasmid hosts, maxing at 50% accuracy for complete plasmids on the genus level [54].

0.4.2. Plasmid similarity network shows hubs and singletons

With FastANI and networkx, we constructed a plasmid similarity network. Isolated nodes were removed, leading to a loss of 192 nodes. The resulting plasmid similarity network counted 48,709,138 undirected edges connecting the 50,475 nodes, see also Figure 8.

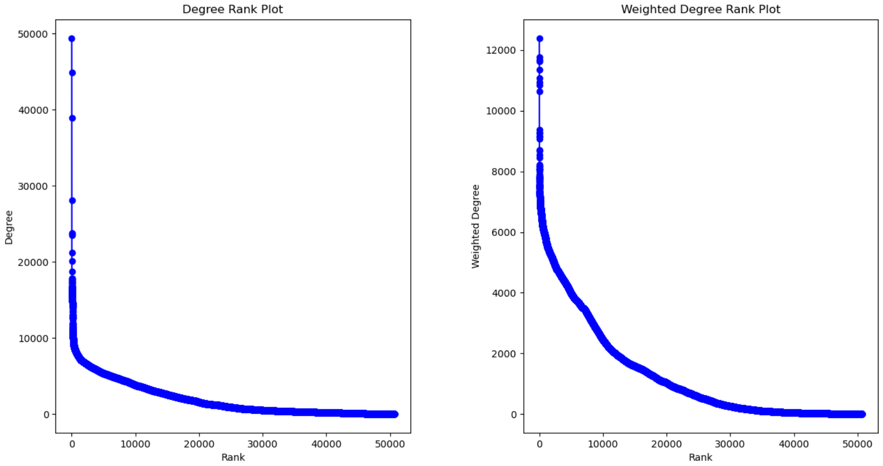

Figure 8: Left: Degree rank plot of the plasmid similarity network. Several sequences have similarities to many sequences, whereas most have a low degree. Right: Weighted degree rank plot of the plasmid similarity network, calculated by adding the 0-1 normalized weights of all edges between the inspected sequence and all other sequences.

Nodes in the network had an average degree of 1,931, ranging from having one neighbor to 49,416 neighbors. We fit the degrees of the nodes to a power-law distribution, a log-linear distribution and an exponential distribution and it was found that the network followed a power-law distribution ($p < 10^{-5}$ when comparing to exponential, $p = 0.014$ when comparing to a log-normal distribution). This indicates that whereas most plasmids are only related to a few other plasmids in the network, several plasmids are shown to share part of their sequences with most other plasmids. Although it is likely that this is because default settings of FastANI only output ANI values when 20 percent of the smallest plasmid of a pair is shared between two plasmids. Increasing this threshold would decrease the number of plasmids that are deemed similar.

0.4.3. Obtaining similar plasmid sequences through clustering plasmid similarity network

We wanted to find similar plasmids that contain conserved regions in order to investigate whether these regions are the underlying cause for plasmid hos range. To find similar plasmids, we clustered the plasmid sequences in the plasmid similarity network. Previous literature has suggested the use of Markov clustering to cluster the plasmid similarity network [127]. Markov clustering is based around (weighted) random walks to simulate flow in the network in order to find connected components. However, the network was too dense to cluster using this method. For this reason, we performed clustering using the Leiden algorithm, which gave more evenly-sized clusters. Leiden clustering is an

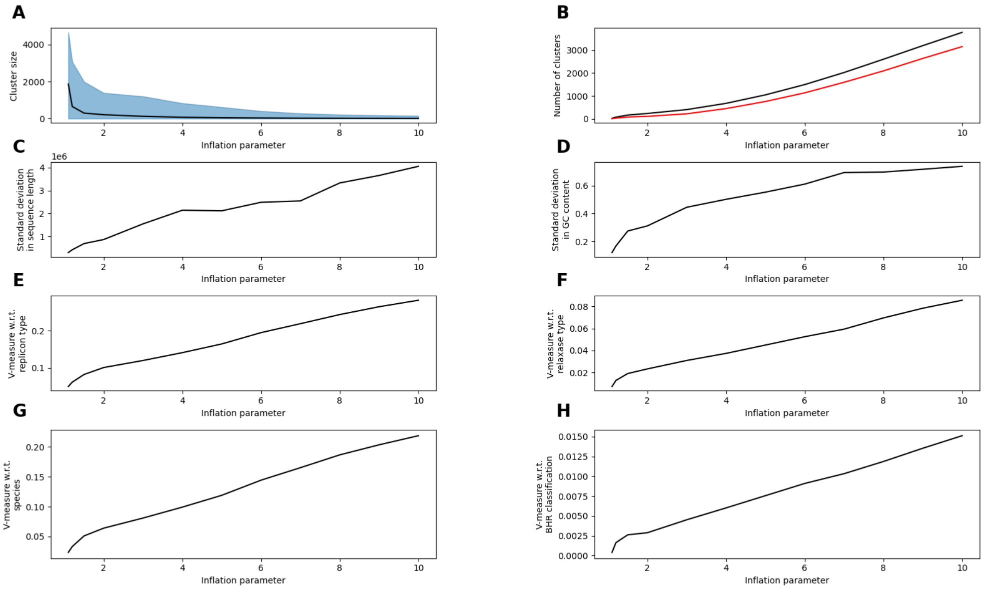

Figure 9: Metrics of clustering quality of clusters derived using Markov Clustering with different inflation parameters. A) Average cluster sizes as a black line, with the blue filling representing standard deviation in cluster sizes. B) Total number of clusters (in black) and the number of clusters containing less than 5 sequences (in red). C) standard deviation for sequence length, calculated by adding the variance in sequence length for each cluster resulting from different inflation parameters and taking the square root. D) Standard deviation for GC content, calculated by adding the variance in GC content for each cluster resulting from different inflation parameters and taking the square root. E) V-measure scores w.r.t replicon type. F) V-measure scores w.r.t. relaxase type. G) V-measure score w.r.t. host species. H) V-measure scores w.r.t. BHR classification.

improvement on the Louvain algorithm and maximizes modularity in order to detect communities [130].

Previous literature has shown that community detection in plasmid similarity networks results in communities that are homogeneous in relaxase type, sequence length and GC content [1]. Concordance has been shown concerning replicon type and host organism as well, albeit to a lesser extent than relaxase type and GC content; mainly because certain communities tend to contain plasmids that can replicate multiple different hosts [107]. We assessed the clusters resulting from both Markov clustering and Leiden clustering using different parameters through analyzing these features, in order to assess cluster quality.

Markov Clustering resulted in inconsistent cluster sizes

Markov clustering resulted in several very large clusters and a lot of very small clusters, with low concordance in replicon and relaxase type as well as GC content and sequence length. The cluster sizes, the number of clusters with less than 5 plasmids and the total number of clusters, the standard deviation in sequence length and GC content, and the v-measure scores between replicon, relaxase, species and BHR classification is given in Figure 9.

In Figure 9A, one can see that increasing the inflation parameter leads to more, smaller clusters. However, there are also a few very large clusters. This can also be seen in Figure 9B, where the number of clusters containing less than 5 sequences grows rather large when compared to the total amount of clusters. The remaining clusters would account for the rest of the dataset, in turn being very large. This was an unwanted effect, as large clusters increases the computation time for finding conserved regions significantly, whereas too small clusters would not allow us to derive conserved regions that would be diverse enough to find evidence of conserved regions. Increasing the inflation parameter has an effect on the standard deviation in sequence length, but the result is variable. Increasing the inflation parameter did however decrease standard deviation in GC content, meaning sequences with similar GC content tended to cluster together. Nevertheless, the smaller cluster size leads to an increase in v-measure scores with regards to replicon, relaxase, species and BHR classification (Figure 9E-H), although these scores remain low.

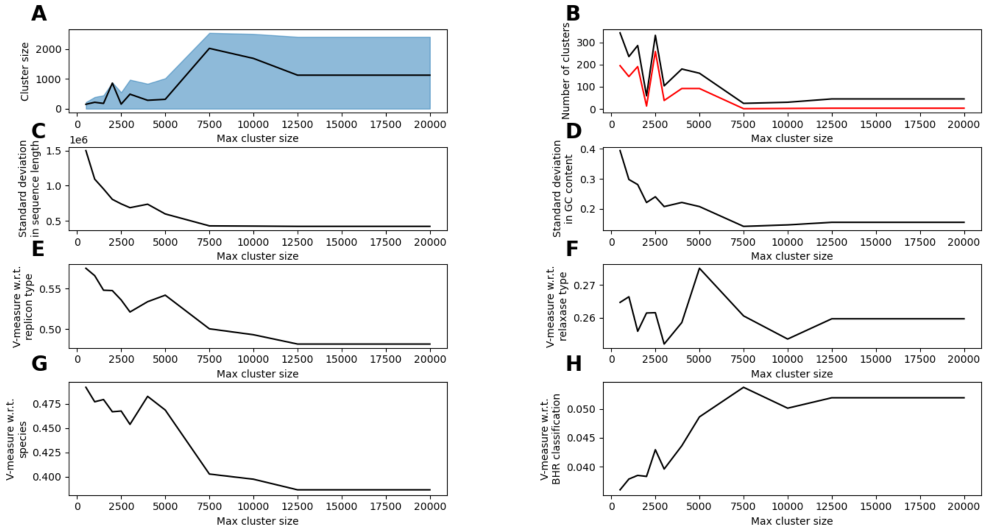

Figure 10: Metrics to measure clustering quality of clusters derived using Leiden Clustering with different maximum cluster sizes. A) Average cluster sizes as a black line, with the blue filling representing standard deviation in cluster sizes. B) Total number of clusters (in black) and the number of clusters containing less than 5 sequences (in red). C) standard deviation for sequence length, calculated by adding the variance in sequence length for each cluster resulting from different inflation parameters and taking the square root. D) Standard deviation for GC content, calculated by adding the variance in GC content for each cluster resulting from different inflation parameters and taking the square root. E) V-measure scores w.r.t replicon type. F) V-measure scores w.r.t. relaxase type. G) V-measure score w.r.t. host species. H) V-measure scores w.r.t. BHR classification.

All in all, increasing the inflation parameter with Mar-kov clustering led to few large clusters with many clusters containing less than five sequences, while not increasing the v-measure concerning replicon, relaxase and species to the extent that was expected. Whereas for larger clusters, we expected that the computation time for finding conserved regions would be too much for this project, for smaller clusters we expected that the conserved regions we would extract were not diverse enough to find evidence of conserved regions. For this reason, we decided to use the Leiden algorithm.

Leiden algorithm resulted in more homogeneous clusters of similar size

The network was then clustered through the Leiden algorithm with different max cluster sizes. Statistics on the resulting clusters is shown in Figure 10. Decreasing the max cluster size created more, smaller clusters. Although the number of clusters with less than 5 plasmids did increase with this decreasing max cluster size, sequences in these clusters would total less than 1000 plasmids. As these clusters would be too small to extract conserved regions which were reliably found across different plasmids, missing out on these 1000 plasmids was deemed acceptable. Decreasing the max cluster size when clustering with the Leiden algorithm increased the v-measure scores for replicon and species. Differing the cluster sizes varied the v-measure score concerning relaxase types which peaked with a max cluster size of 5000. As we thought that smaller clusters would result in more similar plasmids and thus more similar GC content, we unexpectedly found that decreasing the max cluster size had an adverse effect on the standard deviation of sequence length and GC content contained in a cluster.

When comparing Leiden Clustering to Markov Clustering, Leiden clustering with the smallest cluster size resulted in many clusters of said max cluster size, as opposed to one large cluster with many smaller clusters courtesy of Markov Clustering. The standard deviation in sequence length and GC content both rose when decreasing cluster size, but Leiden clustering seemed to cluster sequences of similar length (length standard deviation through Leiden clustering at max 500 cluster size: ~1.5e6, length standard deviation through MCL with inflation parameter 10: ~4e6) and GC content (GC standard deviation through Leiden at max 500: ~0.4, GC standar deviation through MCL with inflation parameter 10: ~0.7) better than the previously applied Markov clustering. The v-measure scores increased when creating finer clusters through both algorithms, but the cluster size was more consistent through Leiden clustering. As a trade-off between the computation time of SHIP and detecting conserved regions in diverse enough plasmids to obtain valuable results, the clusters resulting from Leiden clustering with a maximum cluster size of 500 sequences was chosen to create synteny graphs to obtain the conserved regions.

0.4.4. Analysis of conserved regions found underlying functions related to host range expansion

We extracted 21,975 conserved regions using SHIP. Table 1 shows us the seeds and the number of conserved regions found. We observe that TA-systems are often used as seeds in our conserved regions, indicating that they are prevalent in our dataset. Restriction-modification systems are also often found as seeds in our conserved regions. We screened the conserved regions on the presence of antibiotic resistance genes, but found none. This suggests that our criteria for finding conserved regions were too stringent and thus did not allow us to find ARGs. We also analyzed which taxonomic barriers the conserved regions were able to transcend, as can be seen in 11. One conserved region was found to cross different phyla. We can see that generally, the number of conserved regions increases as the criteria for BHR becomes less stringent.

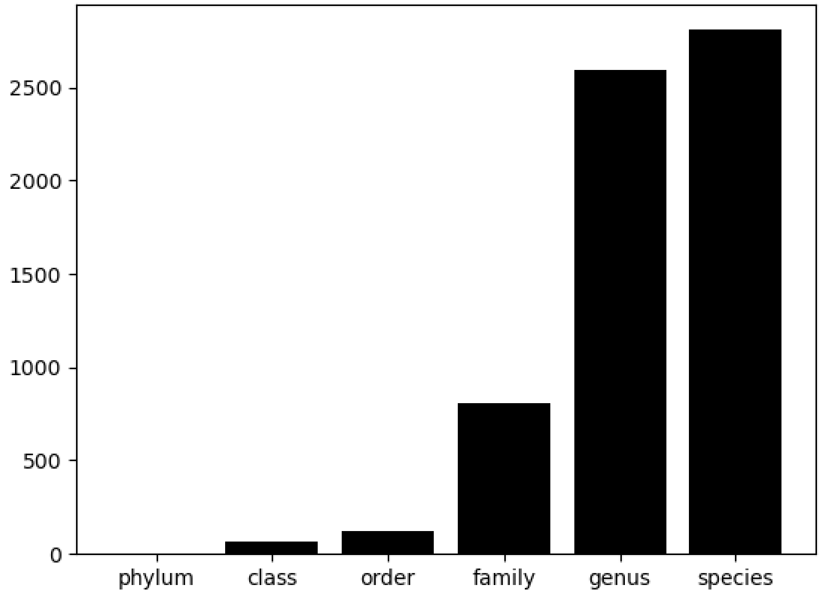

Figure 11: Number of conserved regions and the highest taxonomical rank they were able to spread to. One conserved region was found in plasmids that belong to different phyla.

Table 1: Table of the number of gathered conserved regions and the seeds that were used to gather these conserved regions. An exception to this were AMR genes which were not used as seeds for the conserved regions but screened using CARD tools and databases [3, 2].

Seed gene	Number of conserved regions
Replicon	1942
Relaxase	1827
Conjugation	1473
Toxin/antitoxin	13009
Restriction/Modification	4310
AMR genes*	0

One of the conserved regions, which we will name CR1, was found in 720 plasmids, of which 576 were deemed BHR. A schematic representation of this region is shown in Figure 12. CR1 contained three genes associated with the creation of the MPF system, namely a pilus assembly protein, an unnamed conjugal transfer protein and PilX2, all proteins necessary for the creation of the sex pilus through which plasmids are transported during conjugation. These regions are known for the transport of large molecules, for instance proteins and protein-DNA complexes among which plasmids, and as such contribute to plasmid host range expansion and bacterial virulence [138]. Therefore, it is likely that this conserved region plays a major role in the broad host range of the plasmids in which it is found. The other found proteins were a transglycosylase, important for the biosynthesis of the bacterial peptidoglycan cell wall [19]; followed by two RfaH transcription anti-termination proteins, known to promote expressions of operons that encode proteins related to the cell surface [112]; a CotH spore coat protein, known to create spores of *Bacillus subtilis* and act as a promotor of the invasion of epithelial cells in *Rhizopus delemar* [39, 40]; and finally a hypothetical protein whose function is unknown.

The conserved region was found in *E. coli, K. pneumoniae, E. hormaechei, C. freundii, K. quasipneumoniae* and *E. cloacae*, all members of the order *Enterobacterales*. We assessed how different the plasmids were through checking the different ANI values between plasmids. The lowest found ANI between two plasmids was 82.53%, with the median distance being 99.93% [53]. This suggests that although most plasmids with the conserved region were deemed highly similar, the conserved region was also found in plasmids that are rather dissimilar.

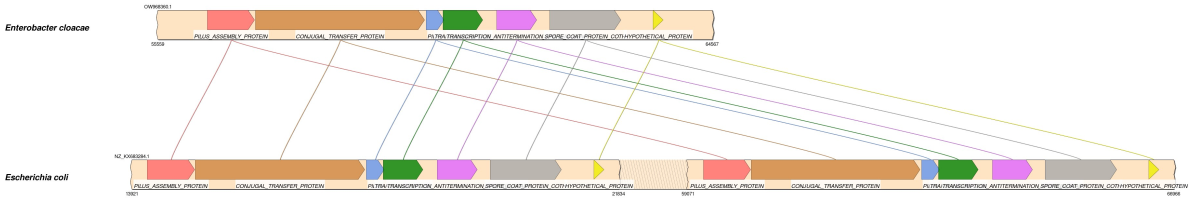

Figure 12: Schematic representation of the conserved region CR1 on two different plasmids, with the named functional annotation given per gene. The conserved region was found on NZ_KX683284.1 twice, suggesting that this sequence is a plasmid dimer.

Table 2: The complete region CR1 and individual genes that were found in conserved region CR1, how often the genes were found in the plasmid dataset in total, the number of these plasmids that were found to be BHR and the expected number of BHR plasmids under the assumption that genes are independently distributed between BHR plasmids and single host plasmids. In the bottom rows, we gathered the replicon, MPF and relaxase and the MPF and relaxase types respectively for plasmids containing the conserved region, picked the combination that occurred most often in these plasmids and the total number of plasmids with these regions, the number of BHR plasmids and the expected number of plasmids.

Gene in conserved region	All	BHR	Exp. BHR
Complete region	720	576	110
Pilus assembly protein	718	576	109
Conjugal transfer protein	718	576	109
PilX2	720	576	110
Transglycosylase	723	576	110
Rfah transcription anti-termination protein	738	579	112
CotH spore coat protein	723	576	110
Hypothetical protein	1076	654	164
IncX3, MOB P, MPF T	744	595	113
MOB P, MPF T	4269	1363	650

A table on the individual genes and how many plasmids they are found in in our complete dataset is shown in table 2. Most of the genes found were present in ~720 plasmids like the complete region was. The hypothetical protein was an exception to these individual genes, occurring in 1076 plasmids of which 654 were deemed BHR. The individual genes seemed to contribute to plasmids being BHR, but the individual genes occurred as often as the conserved region. This suggests that the genes are co-located for an underlying biological reason.

Current gene-based approaches use the combination of replicon, relaxase and MPF types to predict plasmid host range. Of the replicon types found on the plasmids the syntenic region was derived from, 79.5% were type IncX3 replicons, which is known for its association with *Enterobacterales* [59]. Furthermore, 96.2% of the found relaxases on the plasmids were of type MOB_P and 98.2% of the found MPF types were of type MPF_T. These replicons, relaxases and MPF types are related to plasmids found in different species of the order *Enterobacterales*. Plasmids with these replicon, relaxase and MPF types were found to be BHR about as often as plasmids containing the conserved region. The conserved region is, however, based on a conjugal transfer protein, suggesting that the conserved region relates to relaxase and MPF type. For this reason, we also checked solely the MPF and relaxase types that were most often found, also seen in 2. When comparing this combination to the conserved region, we find that the conserved region is located on less plasmids of which proportionally more are BHR. All in all, this conserved region may not be the sole determinant of plasmid host range, but the region correlates more with the plasmids being BHR than the genes that are currently used in gene-based *in silico* host range prediction and thus suggests that this conserved region would be better in determining plasmid host range than the state-of-the-art approaches.

A second conserved region, which we will call CR2, was analyzed similarly to the previously mentioned conserved region and displayed in Figure 13 and table 3. The genes contained on the conserved region were a DNA topoisomerase and a nucleoid-associated DNA binding protein, responsible for DNA supercoiling [31]; a flagellin N-methylase, which seems to modify flagella and is thought to increase virulence by increasing cell adhesion through the flagella [50]; a hypothetical protein; antirestriction protein ArdR which is often found on promiscuous plasmids to affect restriction activity of restriction/modification-systems selectively [88] and a SprT-like domain-containing proteins, which does not have a known function. Combined, the genes suggest that the region alters transcription and increases virulence in its host, but the lack of annotation of half of the found proteins hinders more speculation. The functional annotations of the genes suggest that it is unlikely that this region plays a major role in determining host range, as the genes are not involved in plasmid transfer or replication mechanisms. The lowest ANI between different plasmids was 92.46%, with plasmids having a median ANI of 99.93%. This once again suggests that most of the plasmids have a common ancestor causing them to be more similar.

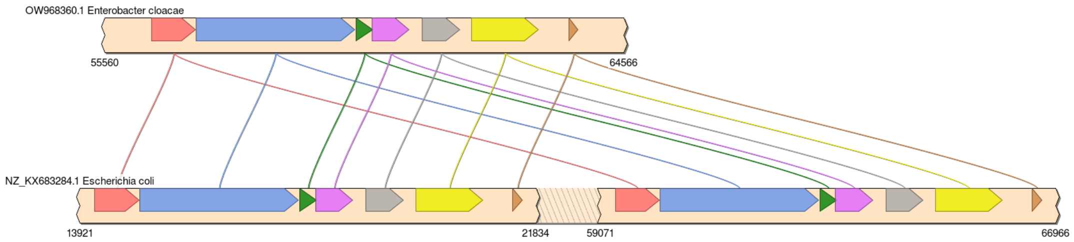

Figure 13: Schematic representation of the conserved region CR2 on two different plasmids, with the named functional annotation given per gene.

Like the previous conserved region, most of the individual genes were found about as often in plasmids as the conserved region, but the antirestriction protein was less prevalent in the conserved region. As we used the same similarity and identity threshold for BLASTn, this suggests that the entire conserved region is generally kept very conserved, whereas the antirestriction protein sequence was less conserved and often missing. We checked whether the antirestriction protein was more often found with less stringent similarity thresholds, but this was not the case. This protein is also more prevalent in non-BHR plasmids, which we hypothesize is due to the plasmid adapting to the host.

Of the plasmids with the conserved region, 91% were found in *E. coli*, *K. pneumoniae*, *E. hormaechei*, *E. cloacae*, *C. freundii* and *K. quasipneumoniae*, all *Enterobacterales*. The syntenic block was found 75% of the time on plasmids containing IncX3 replicons, which are common in *Enterobacterales* [59, 94], and 92.5% of the plasmids contained MOB_P relaxases and 94.7% were of MPF type T. Like the previous region, the combination of replicon, relaxase and MPF types also correlated significantly with the plasmid being BHR. This, together with the individual genes being located in the same number of plasmids, suggests that plasmid host range is not determined by this specific syntenic block.

Another conserved region, which we will dub CR3, was investigated as this was also associated with BHR plasmids. The syntenic region contained conjugal transfer protein traG, which is required for the conjugative transfer of plasmids and mediates interactions between DNA-processing and Mpf systems [114]; a VirB-family protein, which is likely a protein that aids in forming the type IV secretion system [20]; an inner membrane protein; a putative partition-related protein, which are known to give each daughter cell a plasmid copy during cell division; and antirestriction protein ArdA, which is often found on promiscuous plasmids to affect restriction activity of restriction/modification-systems selectively [88]. Like the previous region, the antirestriction protein was less often found, alongside the partition protein. Why the conserved region is found, but the individual genes are missing, is because the complete sequence does match the identity and similarity thresholds whereas the individual genes would be found if the similarity threshold was lowered. We inspected the BLAST results for the antirestriction protein, and it was found on 72 plasmids (26 BHR) within the 80-100% similarity range. Similarly, for the partitioning protein, it was found on 749 plasmids (576 BHR) when also considering similarities in the 80-100% similarity range. This lower similarity suggests that the antirestriction and partitioning proteins are altered in some hosts as a result of host adaptation. The plasmid was once again mostly found in highly similar plasmids with the median ANI being 99.93%, but the lowest being 92.70%, suggesting that the conserved region is found on a plasmid that had spread and ameliorated to different species than that it is found on multiple different plasmids.

Similar to the previously mentioned conserved region, the conserved region was mainly found on plasmids of *E. coli, K. pneumoniae, E. hormaechei, E. cloacae, C. freundii* and *K. quasipneumoniae*. The syntenic block was found 78.9% on plasmids with the IncX3 replicon type, 95.2% on plasmids with the MOBF relaxase types and 98.1% on plasmids with MPF type T, similar to the other two mentioned conserved regions.

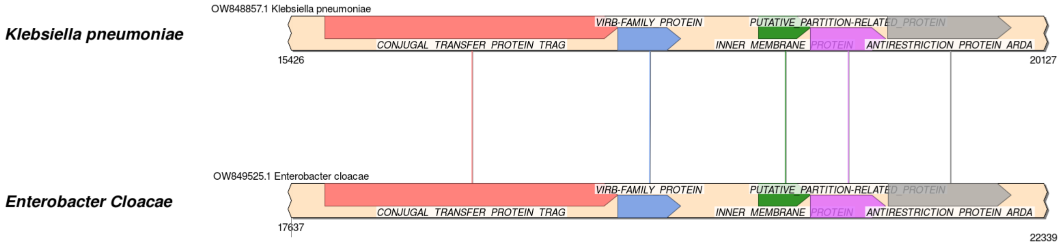

Figure 14: Schematic representation of the conserved region CR3 on two different plasmids, with the named functional annotation given per gene.

Finally, we checked the conserved region that was found to cross different phyla, CR4, which is depicted in Figure 15 and, together with the individual genes and replicon/relaxase/MPF type combination, in table 5. The conserved region was found in bacteria of the phyla *Bacteroidota* and *Pseudomonadota*. The conserved region contained a Rep-3 domain-containing protein, described to be similar to known replication proteins through AlphaFold [60]; a DUF4468 domain-containing protein, whose function is unknown [11]; mobilization proteins MbpC, MbpB and MbpA, which have roles in the mobilization process [92]. The conserved region also hosts a Txe-YoeB family toxin-anti-toxin system [62]. MbpC

Table 3: The individual genes that were found in conserved region CR2, how often the genes were found in the plasmid dataset in total, the number of these plasmids that were found to be BHR and the expected number of BHR plasmids under the assumption that genes are independently distributed between BHR plasmids and single host plasmids.

Gene in conserved region	All	BHR	Exp. BHR
Conserved region	724	575	110
DNA topoisomerase	729	576	111
DNA binding protein	713	576	110
YkgJ fam. cysteine cl. protein	718	572	110
Antirestriction protein ArdR	38	2	5.8
SprT-domain protein	730	576	111
Zinc-ribbon domain protein	730	576	111
IncX3, MOB P, MPF T	744	595	113

Table 4: The complete sequence and the individual genes that were found in conserved region CR3, how often the genes were found in the plasmid dataset in total, the number of these plasmids that were found to be BHR and the expected number of BHR plasmids under the assumption that genes are independently distributed between BHR plasmids and single host plasmids.

Gene in conserved region	All	BHR	Exp. BHR
Conserved region	724	575	110
TraG	729	576	111
VirB-family protein	728	576	111
Inner membrane protein	716	570	110
Putative partition protein	37	14	6
Antirestriction protein ArdA	52	16	8
IncX3, MOB P, MPF T	744	595	113

were found to be less conserved but often found when including 80-90% sequence similarity, whereas DUF4468-domain containing protein as well as the toxin-antitoxin systems were often not detected using BLAST. This suggests that mainly the mobilization proteins and the Rep3-domain containing protein were highly conserved. Plasmids from this region were most divergent, with the median ANI being 98.06% and the lowest being 83.30%.

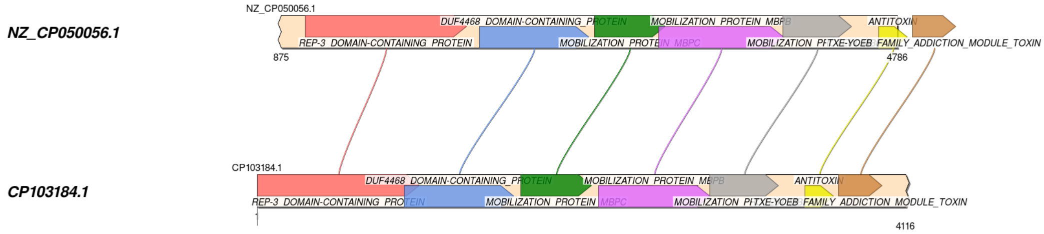

Figure 15: Schematic representation of the conserved region CR4 on two different plasmids, with the named functional annotation given per gene.

Table 5: The complete sequence and the individual genes that were found in conserved region CR4, how often the genes were found in the plasmid dataset in total, the number of these plasmids that were found to be BHR and the expected number of BHR plasmids under the assumption that genes are independently distributed between BHR plasmids and single host plasmids.

Gene in conserved region	All	BHR	Exp. BHR
Conserved region	62	39	9.44
Rep3	48	36	7.31
DUF4468	3	0	0.45
MbpC	8	0	1.2
MbpB	57	38	8.6
MbpA	55	39	8.4
Txe-YoeB toxin	4	0	0.6
Txe-YoeB antitoxin	4	0	0.6
No rep, No MOB, No MPF	7069	150	1076

The sequence was found mostly in the genus *Bacteroides*, but it was once found in *P. aeruginosa* and in two different *Phocaeicola* species. For of the plasmids with the conserved region, generally no MOB, relaxase or MPF type were detected, with one replicon type found belonging to rep_cluster_1019 and some MOBP-relaxases being found. As no replicon, relaxase and MPF type have been detected for most plasmids, these plasmids are likely understudied and very different and only grouped together due to the limited knowledge we have of them. Ergo, we found a region that explains why these plasmids are BHR in a set of plasmids for which we did not know why this was the case. All in all, this conserved region is found in the most diverse group of plasmids which are likely to be understudied, but the genes contained on the conserved region are related to plasmid mobilization and replication, suggesting that it most likely plays a role in making a plasmid BHR.

To summarize, we found that the method to extract conserved regions around replication and conjugation machinery was sound. Inspection of the individual genes in the conserved regions suggests that the genes perform a collective function that caused the order of the genes to be conserved. Some of the extracted conserved regions contained antirestriction proteins, and in one case a partitioning protein, that were found less conserved than the entire conserved region. We hypothesize that these genes are prone to mutation to ameliorate the costs of carrying the plasmid. The conserved regions that are based around replication and conjugation machinery are more specifically related to BHR plasmids than the known replication, relaxase and MPF types that are found on the plasmid, suggesting that conserved regions indeed affect host range through conserving a collective function that affects both how plasmids transfer horizontally as well as ameliorate their cost of surviving in a new host. With the results gathered, we hypothesize that genes surrounding the replication and conjugation machinery affect their effectiveness in different hosts. We conclude that conserved regions are more correlated in determining plasmid host range than solely the genes that were used as their seeds.

However, the results gathered are generally limited to *Enterobacterales* with the same replicon, relaxase and MPF types. Furthermore, the plasmids showed that they were usually similar in nucleotide composition, although larger deviation between plasmid similarities was shown. A different definition of broad host range plasmids in higher taxa, i.e. classes or orders, may allow us to find plasmids that are less specific to *Enterobacterales*. Furthermore, investigation of other clusters may provide us with more plasmids found in different organisms than *Enterobacterales*, in turn giving us different host species and different replicon, relaxase and MPF types giving insight in understudied host species.

0.5. Conclusion and Discussion

In this thesis, we investigated whether a relation between host range and conserved regions around replication and conjugation machinery and addiction mechanisms extracted from plasmids exists. We did this through a seed-and-extend method focused on the mentioned genes. The found conserved regions contained genes that are thought to be functionally related to one another and we hypothesize that these serve a collective function in the cell, which is why their order is conserved.

We analyzed the conserved regions that were the most associated with plasmids classified as BHR on the spe-cies level and one conserved region found in different phyla. Two of these were found to contain plasmids linked to conjugation, of which one also contained both genes for evading the bacterial immune system as well as genes for partitioning, suggesting that the conserved region affects host range. On the other hand, CR3, which was also based around an antirestriction protein, did not contain genes aside from the antirestriction protein that were thought to significantly affect plasmid host range. Finally, the last conserved region was found in different phyla and contained several genes related to replication and conjugation machinery. All in all this suggests that some conserved regions were linked to host range, but to find only conserved regions related to host range we should make the criteria for finding conserved regions more stringent.

Among the conserved regions analyzed, we found several that contained machinery related to bacterial conjugation and were associated with BHR plasmids. In that sense, our approach in finding conserved regions associated with host range was fruitful. However, the found and analyzed conserved regions are usually derived from *Enterobacterales*. A more stringent definition of broad host range plasmids where plasmids are only classified as such when they are found across different classes, orders or phyla may allow us to find different conserved regions in plasmids that are able to conjugate between more distant taxa. Furthermore, we did not find any AMR genes in any of the regions associated with BHR. However, AMR genes were found in other regions of the plasmids. This may be because the AMR genes were not detected in our conserved regions because we focused on regions spanning at most 9 genes from our seeds, with AMR genes not falling into this range, or because regions containing both one of our seeds as well as an ARG did not fulfill our criteria to calling the region conserved through SHIP. It has, however, been found that plasmids with ARGs are more broad in host range [145]. Due to this, in combination with several genes having been found to be less conserved in our regions, we hypothesize that the presence of ARGs and plasmid evolvability affect plasmid host range as well, which cannot be captured solely by our conserved regions. Furthermore, some genes not being as conserved as the complete region suggests that even when solely considering conserved regions, not all genes necessarily need to be present to predict host range.

The conserved regions we have detected were generally as correlated with plasmid host range as the combination of replication and conjugation machinery. Although we have not tested whether using multiple conserved regions and inspecting whether or not the conserved region would allow us to increase host range prediction performance, we suggest to research whether the use of multiple conserved regions may allow us to make more accurate plasmid host range predictions than gene-based approaches that are currently used.

For this prediction, we suggest the creation of a data-base of the conserved regions. This database would be used to screen a novel plasmid assembly and find whether it contains conserved regions. These would then be associated with what hosts the regions were found in, in order to output the predicted plasmid host range.

A drawback of our research, although this is perpetual in both gene and host range research, is that it relies on accurate host range classification and understanding which functions genes fulfill. Fundamentally, plasmid host ranges are found through the expensive process of conjugation assays. Computer-based approaches are also able to find host ranges based on sequence comparison; plasmids that are very similar or even the same, found in different species, can be classified as BHR. However, with most species not being sequenced, information on plasmids and their host ranges remains incomplete [54]. Furthermore, many genes on plasmids lack functional annotations [54], information that would aid in interpreting what genes do and why they are located together.

References

[1] Mislav Acman et al. "Large-scale network analysis captures biological features of bacterial plasmids". In: *Nature Communications* 11.1 (May 2020), p. 2452. ISSN: 2041-1723. DOI: `10.1038/s41467-020-16282-w`. URL: `https://doi.org/10.1038/s41467-020-16282-w`.

[2] Brian P Alcock, William Huynh, Romeo Chalil, et al. "CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database". en. In: *Nucleic Acids Res* 51.D1 (Jan. 2023), pp. D690–D699.

[3] Brian P Alcock, Amogelang R Raphenya, Tammy T Y Lau, et al. "CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database". In: *Nucleic Acids Research* 48.D1 (Oct. 2019), pp. D517–D525. ISSN: 0305-1048. DOI: `10.1093/nar/gkz935`. eprint: `https://academic.oup.com/nar/article-pdf/48/D1/D517/31697411/gkz935.pdf`. URL: `https://doi.org/10.1093/nar/gkz935`.

[4] Apostolos Almpanis et al. "Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages". en. In: *Microb Genom* 4.4 (Apr. 2018).

[5] Manuel Ares-Arroyo, Charles Coluzzi, and Eduardo P C Rocha. "Origins of transfer establish networks of functional dependencies for plasmid transfer by conjugation". In: *Nucleic Acids Research* 51.7 (Nov. 2022), pp. 3001–3016. ISSN: 0305-1048. DOI: `10.1093/nar/gkac1079`. eprint: `https://academic.oup.com/nar/article-pdf/51/7/3001/49875792/gkac1079.pdf`. URL: `https://doi.org/10.1093/nar/gkac1079`.

[6] Derya Aytan-Aktug et al. "PlasmidHostFinder: Prediction of Plasmid Hosts Using Random Forest". In: *mSystems* 7.2 (2022), e01180–21. DOI: `10.1128/msystems.01180-21`. eprint: `https://journals.asm.org/doi/pdf/10.1128/msystems.01180-21`. URL: `https://journals.asm.org/doi/abs/10.1128/msystems.01180-21`.

[7] Zeynep Baharoglu, David Bikard, and Didier Mazel. "Conjugative DNA Transfer Induces the Bacterial SOS Response and Promotes Antibiotic Resistance Development through Integron Activation". In: *PLOS Genetics* 6.10 (Oct. 2010), pp. 1–10. DOI: `10.1371/journal.pgen.1001165`. URL: `https://doi.org/10.1371/journal.pgen.1001165`.

[8] Zeynep Baharoglu, David Bikard, and Didier Mazel. "Conjugative DNA Transfer Induces the Bacterial SOS Response and Promotes Antibiotic Resistance Development through Integron Activation". In: *PLOS Genetics* 6.10 (Oct. 2010), pp. 1–10. DOI: `10.1371/journal.pgen.1001165`. URL: `https://doi.org/10.1371/journal.pgen.1001165`.

[9] Zeynep Baharoglu and Didier Mazel. "SOS, the formidable strategy of bacteria against aggressions". In: *FEMS Microbiology Reviews* 38.6 (Nov. 2014), pp. 1126–1145. ISSN: 0168-6445. DOI: `10.1111/1574-6976.12077`. eprint: `https://academic.oup.com/femsre/article-pdf/38/6/1126/18142660/38-6-1126.pdf`. URL: `https://doi.org/10.1111/1574-6976.12077`.

[10] Pauline M Bennett. "Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria". In: *British journal of pharmacology* 153.S1 (2008), S347–S357.

[11] Helen M. Berman, John Westbrook, Zukang Feng, et al. "The Protein Data Bank". In: *Nucleic Acids Research* 28.1 (Jan. 2000), pp. 235–242. ISSN: 0305-1048. DOI: `10.1093/nar/28.1.235`. eprint: `https://academic.oup.com/nar/article-pdf/28/1/235/9895144/280235.pdf`. URL: `https://doi.org/10.1093/nar/28.1.235`.

[12] Jonathan H Bethke et al. "Vertical and horizontal gene transfer tradeoffs direct plasmid fitness". In: *Molecular Systems Biology* 19.2 (2023), e11300. DOI: `https://doi.org/10.15252/msb.202211300`. eprint: `https://www.embopress.org/doi/pdf/10.15252/msb.202211300`. URL: `https://www.embopress.org/doi/abs/10.15252/msb.202211300`.

[13] Tatiana Borodovich et al. "Phage-mediated horizontal gene transfer and its implications for the human gut microbiome". en. In: *Gastroenterol Rep (Oxf)* 10 (Apr. 2022), goac012.

[14] Michael A. Brockhurst and Ellie Harrison. "Ecological and evolutionary solutions to the plasmid paradox". In: *Trends in Microbiology* 30.6 (2022), pp. 534–543. ISSN: 0966-842X. DOI: `https://doi.org/10.1016/j.tim.2021.11.001`. URL: `https://www.sciencedirect.com/science/article/pii/S0966842X21002675`.

[15] Michal Bukowski et al. "Prevalence of antibiotic and heavy metal resistance determinants and virulence-related genetic elements in plasmids of Staphylococcus aureus". In: *Frontiers in Microbiology* 10 (2019), p. 805.

[16] Alessandra Carattoli, Ea Zankari, Aurora García-Fernández, et al. "In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing". en. In: *Antimicrob Agents Chemother* 58.7 (Apr. 2014), pp. 3895–3903.

[17] Amanda C. Carroll and Alex Wong. "Plasmid persistence: costs, benefits, and the plasmid paradox". In: *Canadian Journal of Microbiology* 64.5 (2018). PMID: 29562144, pp. 293–304. DOI: `10.1139/cjm-2017-0609`. eprint: `https://doi.org/10.1139/cjm-2017-0609`. URL: `https://doi.org/10.1139/cjm-2017-0609`.

[18] Sherwood R Casjens, Eddie B Gilcrease, Marija Vujadinovic, et al. "Plasmid diversity and phylogenetic consistency in the Lyme disease agent Borrelia burgdorferi". In: *BMC genomics* 18.1 (2017), pp. 1–18.

[19] Xiaorui Chen, Chi-Huey Wong, and Che Ma. "Targeting the Bacterial Transglycosylase: Antibiotic Development from a Structural Perspective". In: *ACS Infect. Dis.* 5.9 (Sept. 2019), pp. 1493–1504.

[20] Peter J. Christie, Neal Whitaker, and Christian González-Rivera. "Mechanism and structure of the bacterial type IV secretion systems". In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1843.8 (2014). Protein trafficking and secretion in bacteria, pp. 1578–1591. ISSN: 0167-4889. DOI: `https://doi.org/10.1016/j.bbamcr.2013.12.019`. URL: `https://www.sciencedirect.com/science/article/pii/S0167488913004461`.

[21] Anna Ciok et al. "Identification of miniature plasmids in psychrophilic Arctic bacteria of the genus Variovorax". In: *FEMS Microbiology Ecology* 92.4 (Feb. 2016), fiw043. ISSN: 0168-6496. DOI: `10.1093/femsec/fiw043`. eprint: `https://academic.oup.com/femsec/article-pdf/92/4/fiw043/16736700/fiw043.pdf`. URL: `https://doi.org/10.1093/femsec/fiw043`.

[22] Charles Coluzzi et al. "Evolution of Plasmid Mobility: Origin and Fate of Conjugative and Nonconjugative Plasmids". In: *Molecular Biology and Evolution* 39.6 (May 2022), msac115. ISSN: 1537-1719. DOI: `10.1093/molbev/msac115`. eprint: `https://academic.oup.com/mbe/article-pdf/39/6/msac115/44179683/msac115.pdf`. URL: `https://doi.org/10.1093/molbev/msac115`.

[23] Sean Conlan, Pamela J Thomas, Clayton Deming, et al. "Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae". en. In: *Sci Transl Med* 6.254 (Sept. 2014), 254ra126.

[24] M. Couturier et al. "Identification and classification of bacterial plasmids". In: *Microbiological reviews* 52.3 (1988), pp. 375–395.

[25] Valerie M Crowley et al. "A Type IV-A CRISPR-Cas System in Pseudomonas aeruginosa Mediates RNA-Guided Plasmid Interference In Vivo". en. In: *CRISPR J* 2.6 (Nov. 2019), pp. 434–440.

[26] Julian Davies and Dorothy Davies. "Origins and evolution of antibiotic resistance". en. In: *Microbiol Mol Biol Rev* 74.3 (Sept. 2010), pp. 417–433.

[27] Ian Dewan and Hildegard Uecker. "A mathematician's guide to plasmids: an introduction to plasmid biology for modellers". en. In: *Microbiology (Reading)* 169.7 (July 2023).

[28] Anne-Kathrin Dietel et al. "Selective advantages favour high genomic AT-contents in intracellular elements". In: *PLoS genetics* 15.4 (2019), e1007778.

[29] Tatiana Dimitriu, Andrew C Matthews, and Angus Buckling. "Increased copy number couples the evolution of plasmid horizontal transmission and plasmid-encoded antibiotic resistance". en. In: *Proc Natl Acad Sci U S A* 118.31 (Aug. 2021).

[30] Sara Domingues, Gabriela J da Silva, and Kaare M Nielsen. "Integrons: Vehicles and pathways for horizontal dissemination in bacteria". en. In: *Mob Genet Elements* 2.5 (Sept. 2012), pp. 211–223.

[31] Charles J Dorman. "Function of nucleoid-associated proteins in chromosome structuring and transcriptional regulation". en. In: *J Mol Microbiol Biotechnol* 24.5-6 (Feb. 2015), pp. 316–331.

[32] Michael Feldgarden et al. "AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence". In: *Scientific reports* 11.1 (2021), pp. 1–9.

[33] Raul Fernandez-Lopez et al. "Comparative Genomics of the Conjugation Region of F-like Plasmids: Five Shades of F". In: *Frontiers in Molecular Biosciences* 3 (2016). ISSN: 2296-889X. DOI: `10.3389/fmolb.2016.00071`. URL: `https://www.frontiersin.org/articles/10.3389/fmolb.2016.00071`.

[34] Judith Finger and Viji Krishnapillai. "Host range, entry exclusion, and incompatibility of Pseudomonas aeruginosa FP plasmids". In: *Plasmid* 3.3 (1980), pp. 332–342. ISSN: 0147-619X. DOI: `https://doi.org/10.1016/0147-619X(80)90046-3`. URL: `https://www.sciencedirect.com/science/article/pii/0147619X80900463`.

[35] Steven L Foley et al. "Incompatibility group I1 (IncI1) plasmids: their genetics, biology, and public health relevance". In: *Microbiology and Molecular Biology Reviews* 85.2 (2021), pp. 10–1128.

[36] Limin Fu et al. "CD-HIT: accelerated for clustering the next-generation sequencing data". In: *Bioinformatics* 28.23 (2012), pp. 3150–3152.

[37] Valentina Galata et al. "PLSDB: a resource of complete bacterial plasmids". In: *Nucleic Acids Research* 47.D1 (Oct. 2018), pp. D195–D202. ISSN: 0305-1048. DOI: `10.1093/nar/gky1050`. eprint: `https://academic.oup.com/nar/article-pdf/47/D1/D195/33043681/gky1050.pdf`. URL: `https://doi.org/10.1093/nar/gky1050`.

[38] María Pilar Garcillán-Barcia, María Victoria Francia, and Fernando de la Cruz. "The diversity of conjugative relaxases and its application in plasmid classification". en. In: *FEMS Microbiol Rev* 33.3 (May 2009), pp. 657–687.

[39] Teclegiorgis Gebremariam et al. "CotH3 mediates fungal invasion of host cells during mucormycosis". In: *The Journal of clinical investigation* 124.1 (Jan. 2014), pp. 237–250. ISSN: 0021-9738. DOI: `10.1172/jci71349`. URL: `https://europepmc.org/articles/PMC3871245`.

[40] Teclegiorgis Gebremariam et al. "CotH3 mediates fungal invasion of host cells during mucormycosis". In: *The Journal of clinical investigation* 124.1 (Jan. 2014), pp. 237–250. ISSN: 0021-9738. DOI: `10.1172/jci71349`. URL: `https://europepmc.org/articles/PMC3871245`.

[41] Marta Gibert et al. "Temperature Dependent Control of the R27 Conjugative Plasmid Genes". In: *Frontiers in Molecular Biosciences* 7 (2020). ISSN: 2296-889X. DOI: `10.3389/fmolb.2020.00124`. URL: `https://www.frontiersin.org/articles/10.3389/fmolb.2020.00124`.

[42] Valeria Gigante, Hatim Sati, and Peter Beyer. "Recent advances and challenges in antibacterial drug development". In: *ADMET and DMPK* 10.2 (2022), pp. 147–151.

[43] Julien Guglielmini, Fernando de la Cruz, and Eduardo P C Rocha. "Evolution of conjugation and type IV secretion systems". en. In: *Mol Biol Evol* 30.2 (Sept. 2012), pp. 315–331.

[44] James P. J. Hall et al. "Plasmid fitness costs are caused by specific genetic conflicts". In: *bioRxiv* (2021). DOI: `10.1101/2021.04.10.439128`. eprint: `https://www.biorxiv.org/content/early/2021/04/12/2021.04.10.439128.full.pdf`. URL: `https://www.biorxiv.org/content/early/2021/04/12/2021.04.10.439128`.

[45] James P. J. Hall et al. "What makes a megaplasmid?" In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 377.1842 (2022), p. 20200472. DOI: `10.1098/rstb.2020.0472`. eprint: `https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.2020.0472`. URL: `https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2020.0472`.

[46] Ellie Harrison et al. "Parallel Compensatory Evolution Stabilizes Plasmids across the Parasitism-Mutualism Continuum". In: *Current Biology* 25.15 (Aug. 2015), pp. 2034–2039.

[47] Ellie Harrison et al. “Rapid compensatory evolution promotes the survival of conjugative plasmids”. In: *Mobile Genetic Elements* 6.3 (2016). PMID: 27510852, e1179074. DOI: `10.1080/2159256X.2016.1179074`. eprint: `https://doi.org/10.1080/2159256X.2016.1179074`. URL: `https://doi.org/10.1080/2159256X.2016.1179074`.

[48] Graham F Hatfull and Roger W Hendrix. “Bacteriophages and their genomes”. en. In: *Curr Opin Virol* 1.4 (Oct. 2011), pp. 298–303.

[49] Susu He, Alison Burgess Hickman, Alessandro M. Varani, et al. “Insertion Sequence IS<i>26</i> Reorganizes Plasmids in Clinically Isolated Multidrug-Resistant Bacteria by Replicative Transposition”. In: *mBio* 6.3 (2015), 10.1128/mbio.00762–15. DOI: `10.1128/mbio.00762-15`. eprint: `https://journals.asm.org/doi/pdf/10.1128/mbio.00762-15`. URL: `https://journals.asm.org/doi/abs/10.1128/mbio.00762-15`.

[50] Julia A Horstmann, Michele Lunelli, Hélène Cazzola, et al. “Methylation of Salmonella Typhimurium flagella promotes bacterial adhesion and host cell invasion”. In: *Nature Communications* 11.1 (Apr. 2020), p. 2013.

[51] J. D. Hunter. “Matplotlib: A 2D graphics environment”. In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: `10.1109/MCSE.2007.55`.

[52] Aayushi Jain and Preeti Srivastava. “Broad host range plasmids”. In: *FEMS Microbiology Letters* 348.2 (Nov. 2013), pp. 87–96. ISSN: 0378-1097. DOI: `10.1111/1574-6968.12241`. URL: `https://doi.org/10.1111/1574-6968.12241`.

[53] Chirag Jain et al. “High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries”. In: *Nature communications* 9.1 (2018), p. 5114.

[54] Yongxin Ji et al. “HOTSPOT: hierarchical host prediction for assembled plasmid contigs with transformer”. In: *Bioinformatics* 39.5 (Apr. 2023), btad283. ISSN: 1367-4811. DOI: `10.1093/bioinformatics/btad283`. eprint: `https://academic.oup.com/bioinformatics/article-pdf/39/5/btad283/50204890/btad283.pdf`. URL: `https://doi.org/10.1093/bioinformatics/btad283`.

[55] Christopher M Johnson and Alan D Grossman. “Integrative and Conjugative Elements (ICEs): What They Do and How They Work”. en. In: *Annu Rev Genet* 49 (Oct. 2015), pp. 577–601.

[56] James R. Johnson. “CHAPTER 2 - Evolution of Pathogenic Escherichia coli”. In: *Escherichia Coli*. Ed. by Michael S. Donnenberg. San Diego: Academic Press, 2002, pp. 55–77. ISBN: 978-0-12-220751-8. DOI: `https://doi.org/10.1016/B978-012220751-8/50003-3`. URL: `https://www.sciencedirect.com/science/article/pii/B9780122207518500033`.

[57] Timothy J Johnson and Kevin S Lang. “IncA/C plasmids: An emerging threat to human and animal health?” en. In: *Mob Genet Elements* 2.1 (Jan. 2012), pp. 55–58.

[58] Timothy J Johnson et al. “Expansion of the IncX plasmid family for improved identification and typing of novel plasmids in drug-resistant Enterobacteriaceae”. In: *Plasmid* 68.1 (2012), pp. 43–50.

[59] Timothy J Johnson et al. “Expansion of the IncX plasmid family for improved identification and typing of novel plasmids in drug-resistant Enterobacteriaceae”. en. In: *Plasmid* 68.1 (Mar. 2012), pp. 43–50.

[60] John Jumper, Richard Evans, Alexander Pritzel, et al. “Highly accurate protein structure prediction with AlphaFold”. In: *Nature* 596.7873 (Aug. 2021), pp. 583–589.

[61] Ilene Karsch-Mizrachi et al. “The international nucleotide sequence database collaboration”. en. In: *Nucleic Acids Res.* 46.D1 (Jan. 2018), pp. D48–D51.

[62] Barbara Kedzierska, Lu-Yun Lian, and Finbarr Hayes. “Toxin–antitoxin regulation: bimodal interaction of YefM–YoeB with paired DNA palindromes exerts transcriptional autorepression”. In: *Nucleic Acids Research* 35.1 (Dec. 2006), pp. 325–339. ISSN: 0305-1048. DOI: `10.1093/nar/gkl1028`. eprint: `https://academic.oup.com/nar/article-pdf/35/1/325/7131157/gkl1028.pdf`. URL: `https://doi.org/10.1093/nar/gkl1028`.

[63] Julia Kloos et al. “Piggybacking on Niche Adaptation Improves the Maintenance of Multidrug-Resistance Plasmids”. In: *Molecular Biology and Evolution* 38.8 (Mar. 2021), pp. 3188–3201. ISSN: 1537-1719. DOI: `10.1093/molbev/msab091`. eprint: `https://academic.oup.com/mbe/article-pdf/38/8/3188/39457259/msab091.pdf`. URL: `https://doi.org/10.1093/molbev/msab091`.

[64] Uli Klümper et al. "Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community". In: *The ISME Journal* 9.4 (Apr. 2015), pp. 934–945.

[65] Günther Koraimann. "Spread and Persistence of Virulence and Antibiotic Resistance Genes: A Ride on the F Plasmid Conjugation Module". In: *EcoSal Plus* 8.1 (2018), 10.1128/ecosalplus.ESP-0003–2018. DOI: `10.1128/ecosalplus.esp-0003-2018`. eprint: `https://journals.asm.org/doi/pdf/10.1128/ecosalplus.esp-0003-2018`. URL: `https://journals.asm.org/doi/abs/10.1128/ecosalplus.esp-0003-2018`.

[66] Anastasia Kottara et al. "Multi-host environments select for host-generalist conjugative plasmids". In: *BMC Evolutionary Biology* 16.1 (Apr. 2016), p. 70.

[67] Val F. Lanza, María de Toro, M. Pilar Garcillán-Barcia, et al. "Plasmid Flux in Escherichia coli ST131 Sublineages, Analyzed by Plasmid Constellation Network (PLACNET), a New Method for Plasmid Reconstruction from Whole Genome Sequences". In: *PLOS Genetics* 10.12 (Dec. 2014), pp. 1–21. DOI: `10.1371/journal.pgen.1004766`. URL: `https://doi.org/10.1371/journal.pgen.1004766`.

[68] T.D. Lawley et al. "F factor conjugation is a true type IV secretion system". In: *FEMS Microbiology Letters* 224.1 (July 2003), pp. 1–15. ISSN: 0378-1097. DOI: `10.1016/S0378-1097(03)00430-0`. eprint: `https://academic.oup.com/femsle/article-pdf/224/1/1/19106979/224-1-1.pdf`. URL: `https://doi.org/10.1016/S0378-1097(03)00430-0`.

[69] Bing Li et al. "Dissemination of KPC-2-encoding IncX6 plasmids among multiple Enterobacteriaceae species in a single Chinese hospital". In: *Frontiers in microbiology* 9 (2018), p. 478.

[70] Weizhong Li and Adam Godzik. "Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences". In: *Bioinformatics* 22.13 (May 2006), pp. 1658–1659. ISSN: 1367-4803. DOI: `10.1093/bioinformatics/btl158`. eprint: `https://academic.oup.com/bioinformatics/article-pdf/22/13/1658/48838763/bioinformatics\-22\-13\-1658.pdf`. URL: `https://doi.org/10.1093/bioinformatics/btl158`.

[71] Weizhong Li, Lukasz Jaroszewski, and Adam Godzik. "Tolerating some redundancy significantly speeds up clustering of large protein databases ". In: *Bioinformatics* 18.1 (Jan. 2002), pp. 77–82. ISSN: 1367-4803. DOI: `10.1093/bioinformatics/18.1.77`. eprint: `https://academic.oup.com/bioinformatics/article-pdf/18/1/77/48850435/bioinformatics\-18\-1\-77.pdf`. URL: `https://doi.org/10.1093/bioinformatics/18.1.77`.

[72] Wenjun Li et al. "RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation". In: *Nucleic Acids Research* 49.D1 (Dec. 2020), pp. D1020–D1028. ISSN: 0305-1048. DOI: `10.1093/nar/gkaa1105`. eprint: `https://academic.oup.com/nar/article-pdf/49/D1/D1020/35364279/gkaa1105.pdf`. URL: `https://doi.org/10.1093/nar/gkaa1105`.

[73] Xiaobin Li et al. "Diversification of broad host range plasmids correlates with the presence of antibiotic resistance genes". In: *FEMS Microbiology Ecology* 92.1 (Dec. 2015), fiv151. ISSN: 0168-6496. DOI: `10.1093/femsec/fiv151`. eprint: `https://academic.oup.com/femsec/article-pdf/92/1/fiv151/16735255/fiv151.pdf`. URL: `https://doi.org/10.1093/femsec/fiv151`.

[74] Gang Liu et al. "Antibiotic-Induced, Increased Conjugative Transfer Is Common to Diverse Naturally Occurring ESBL Plasmids in Escherichia coli". In: *Frontiers in Microbiology* 10 (2019). ISSN: 1664-302X. DOI: `10.3389/fmicb.2019.02119`. URL: `https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2019.02119`.

[75] Hao-Yu Lo et al. "IncP-type plasmids carrying genes for antibiotic resistance or for aromatic compound degradation are prevalent in sequenced Aromatoleum and Thauera strains". In: *Environmental Microbiology* 24.12 (2022), pp. 6411–6425. DOI: `https://doi.org/10.1111/1462-2920.16262`. eprint: `https://enviromicro-journals.onlinelibrary.wiley.com/doi/pdf/10.1111/1462-2920.16262`. URL: `https://enviromicro-journals.onlinelibrary.wiley.com/doi/abs/10.1111/1462-2920.16262`.

[76] Wesley Loftie-Eaton and Douglas E. Rawlings. "Diversity, biology and evolution of IncQ-family plasmids". In: *Plasmid* 67.1 (2012), pp. 15–34. ISSN: 0147-619X. DOI: `https://doi.org/10.1016/j.plasmid.2011.10.001`. URL: `https://www.sciencedirect.com/science/article/pii/S0147619X11000928`.

[77] Wesley Loftie-Eaton et al. “Evolutionary Paths That Expand Plasmid Host-Range: Implications for Spread of Antibiotic Resistance”. en. In: *Mol Biol Evol* 33.4 (Dec. 2015), pp. 885–897.

[78] Allison J Lopatkin et al. “Persistence and reversal of plasmid-mediated antibiotic resistance”. In: *Nature Communications* 8.1 (Nov. 2017), p. 1689.

[79] Saurabh Mahajan and Deepa Agashe. “Evolutionary jumps in bacterial GC content”. en. In: *G3 (Bethesda)* 12.8 (July 2022).

[80] D Maher and D E Taylor. “Host range and transfer efficiency of incompatibility group HI plasmids”. en. In: *Can J Microbiol* 39.6 (June 1993), pp. 581–587.

[81] J Mahillon and M Chandler. “Insertion sequences”. en. In: *Microbiol Mol Biol Rev* 62.3 (Sept. 1998), pp. 725–774.

[82] Surbhi Malhotra-Kumar et al. “Colistin resistance gene mcr-1 harboured on a multidrug resistant plasmid”. In: *The Lancet infectious diseases* 16.3 (2016), pp. 283–284.

[83] JL Martinez and F Baquero. “Mutation frequencies and antibiotic resistance”. In: *Antimicrobial agents and chemotherapy* 44.7 (2000), pp. 1771–1777.

[84] Amy J Mathers, Gisele Peirano, and Johann DD Pitout. “The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae”. In: *Clinical microbiology reviews* 28.3 (2015), pp. 565–591.

[85] Wes McKinney. “Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 56–61. DOI: `10.25080/Majora-92bf1922-00a`.

[86] Claudia Michaelis and Elisabeth Grohmann. “Horizontal Gene Transfer of Antibiotic Resistance Genes in Biofilms”. In: *Antibiotics* 12.2 (2023). ISSN: 2079-6382. DOI: `10.3390/antibiotics12020328`. URL: `https://www.mdpi.com/2079-6382/12/2/328`.

[87] A San Millan et al. “Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids”. In: *Nature Communications* 5.1 (Oct. 2014), p. 5208.

[88] Sergei V. Nekrasov et al. “Plasmid-encoded Antirestriction Protein ArdA Can Discriminate between Type I Methyltransferase and Complete Restriction–Modification System”. In: *Journal of Molecular Biology* 365.2 (2007), pp. 284–297. ISSN: 0022-2836. DOI: `https://doi.org/10.1016/j.jmb.2006.09.087`. URL: `https://www.sciencedirect.com/science/article/pii/S0022283606013222`.

[89] Hiromi Nishida. “Comparative analyses of base compositions, DNA sizes, and dinucleotide frequency profiles in archaeal and bacterial chromosomes and plasmids”. en. In: *Int J Evol Biol* 2012 (Mar. 2012), p. 342482.

[90] Luísa Czamanski Nora et al. “The art of vector engineering: towards the construction of next-generation genetic tools”. In: *Microbial biotechnology* 12.1 (2019), pp. 125–147.

[91] Peter Norberg et al. “The IncP-1 plasmid backbone adapts to different host bacterial species and evolves through homologous recombination”. In: *Nature Communications* 2.1 (Apr. 2011), p. 268.

[92] T J Novicki and D W Hecht. “Characterization and DNA sequence of the mobilization region of pLV22a from Bacteroides fragilis”. In: *Journal of Bacteriology* 177.15 (1995), pp. 4466–4473. DOI: `10.1128/jb.177.15.4466-4473.1995`. eprint: `https://journals.asm.org/doi/pdf/10.1128/jb.177.15.4466-4473.1995`. URL: `https://journals.asm.org/doi/abs/10.1128/jb.177.15.4466-4473.1995`.

[93] Jim O’Neill. “Tackling drug-resistant infections globally: final report and recommendations”. In: *amr review* (2016).

[94] Érica Maria de Oliveira et al. “High plasmid variability, and the presence of IncFIB, IncQ, IncA/C, IncHI1B, and IncL/M in clinical isolates of Klebsiella pneumoniae with bla KPC and bla NDM from patients at a public hospital in Brazil”. en. In: *Rev Soc Bras Med Trop* 53 (Oct. 2020), e20200397.

[95] Alex Orlek, Nicole Stoesser, Muna F Anjum, et al. “Plasmid Classification in an Era of Whole-Genome Sequencing: Application in Studies of Antibiotic Resistance Epidemiology”. en. In: *Front Microbiol* 8 (Feb. 2017), p. 182.

[96] Mitchell W. Pesesky, Rayna Tilley, and David A.C. Beck. "Mosaic plasmids are abundant and unevenly distributed across prokaryotic taxa". In: *Plasmid* 102 (2019), pp. 10–18. ISSN: 0147-619X. DOI: `https://doi.org/10.1016/j.plasmid.2019.02.003`. URL: `https://www.sciencedirect.com/science/article/pii/S0147619X18301215`.

[97] Eugen Pfeifer et al. "Bacteria have numerous distinctive groups of phage–plasmids with conserved phage and variable plasmid gene repertoires". In: *Nucleic Acids Research* 49.5 (Feb. 2021), pp. 2655–2673. ISSN: 0305-1048. DOI: `10.1093/nar/gkab064`. eprint: `https://academic.oup.com/nar/article-pdf/49/5/2655/36627348/gkab064.pdf`. URL: `https://doi.org/10.1093/nar/gkab064`.

[98] Minh-Duy Phan and John Wain. "IncHI plasmids, a dynamic link between resistance and pathogenicity". In: *The Journal of Infection in Developing Countries* 2.04 (2008), pp. 272–278.

[99] Stephanie Pillay et al. "Metagenomic-based surveillance systems for antibiotic resistance in non-clinical settings". en. In: *Front Microbiol* 13 (Dec. 2022), p. 1066995.

[100] Rafael Pinilla-Redondo et al. "CRISPR-Cas systems are widespread accessory elements across bacterial and archaeal plasmids". In: *Nucleic Acids Research* 50.8 (Oct. 2021), pp. 4315–4328. ISSN: 0305-1048. DOI: `10.1093/nar/gkab859`. eprint: `https://academic.oup.com/nar/article-pdf/50/8/4315/43547346/gkab859.pdf`. URL: `https://doi.org/10.1093/nar/gkab859`.

[101] Rafael Pinilla-Redondo et al. "Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids". In: *Nucleic Acids Research* 48.4 (Dec. 2019), pp. 2000–2012. ISSN: 0305-1048. DOI: `10.1093/nar/gkz1197`. eprint: `https://academic.oup.com/nar/article-pdf/48/4/2000/32615310/gkz1197.pdf`. URL: `https://doi.org/10.1093/nar/gkz1197`.

[102] Andreas Porse et al. "Survival and Evolution of a Large Multidrug Resistance Plasmid in New Clinical Bacterial Hosts". en. In: *Mol Biol Evol* 33.11 (Aug. 2016), pp. 2860–2873.

[103] Leslie Pray. *Transposons: The Jumping Genes*. 2008. URL: `https://www.nature.com/scitable/topicpage/transposons-the-jumping-genes-518/`.

[104] Joshua P Ramsay et al. "An updated view of plasmid conjugation and mobilization in Staphylococcus". en. In: *Mob Genet Elements* 6.4 (July 2016), e1208317.

[105] Anuradha Ravi et al. "Transmission and persistence of IncF conjugative plasmids in the gut microbiota of full-term infants". In: *FEMS Microbiology Ecology* 94.1 (Nov. 2017), fix158. ISSN: 0168-6496. DOI: `10.1093/femsec/fix158`. eprint: `https://academic.oup.com/femsec/article-pdf/94/1/fix158/23576737/fix158.pdf`. URL: `https://doi.org/10.1093/femsec/fix158`.

[106] Douglas E. Rawlings and Erhard Tietze. "Comparative Biology of IncQ and IncQ-Like Plasmids". In: *Microbiology and Molecular Biology Reviews* 65.4 (2001), pp. 481–496. DOI: `10.1128/mmbr.65.4.481-496.2001`. eprint: `https://journals.asm.org/doi/pdf/10.1128/mmbr.65.4.481-496.2001`. URL: `https://journals.asm.org/doi/abs/10.1128/mmbr.65.4.481-496.2001`.

[107] Santiago Redondo-Salvo, Raúl Fernández-López, et al. "Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids". In: *Nature Communications* 11.1 (July 2020), p. 3602.

[108] Santiago Redondo-Salvo et al. "COPLA, a taxonomic classifier of plasmids". In: *BMC bioinformatics* 22.1 (2021), pp. 1–9.

[109] James Robertson and John H E Nash. "MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies". en. In: *Microb Genom* 4.8 (July 2018).

[110] James Robertson et al. "Universal whole-sequence-based plasmid typing and its utility to prediction of host range and epidemiological surveillance". In: *Microbial Genomics* 6.10, e000435 (2020). ISSN: 2057-5858. DOI: `https://doi.org/10.1099/mgen.0.000435`. URL: `https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000435`.

[111] Andrew Rosenberg and Julia Hirschberg. "V-measure: A conditional entropy-based external cluster evaluation measure". In: *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*. 2007, pp. 410–420.

[112] Thomas J Santangelo and Jeffrey W Roberts. “RfaH, a Bacterial Transcription Antiterminator”. In: *Molecular Cell* 9.4 (2002), pp. 698–700. issn: 1097-2765. doi: https://doi.org/10.1016/S1097-2765(02)00516-6. url: https://www.sciencedirect.com/science/article/pii/S1097276502005166.

[113] Georges P Schmartz et al. “PLSDB: advancing a comprehensive database of bacterial plasmids”. In: *Nucleic Acids Research* 50.D1 (Nov. 2021), pp. D273–D278. issn: 0305-1048. doi: 10.1093/nar/gkab1111. eprint: https://academic.oup.com/nar/article-pdf/50/D1/D273/42058178/gkab1111.pdf. url: https://doi.org/10.1093/nar/gkab1111.

[114] Gunnar Schröder et al. “TraG-like proteins of DNA transfer systems and of the Helicobacter pylori type IV secretion system: inner membrane gate for exported substrates?” en. In: *J Bacteriol* 184.10 (May 2002), pp. 2767–2779.

[115] Oliver Schwengers et al. “Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification”. In: *Microbial Genomics* 7.11, 000685 (2021). issn: 2057-5858. doi: https://doi.org/10.1099/mgen.0.000685. url: https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000685.

[116] Zhen Shen, Christoph M Tang, and Guang-Yu Liu. “Towards a better understanding of antimicrobial resistance dissemination: what can be learnt from studying model conjugative plasmids?” In: *Military Medical Research* 9.1 (Jan. 2022), p. 3.

[117] Masaki Shintani, Zoe K. Sanchez, and Kazuhide Kimbara. “Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy”. In: *Frontiers in Microbiology* 6 (2015). issn: 1664-302X. doi: 10.3389/fmicb.2015.00242. url: https://www.frontiersin.org/articles/10.3389/fmicb.2015.00242.

[118] Masaki Shintani et al. “The Behavior and Significance of Degradative Plasmids Belonging to Inc Groups in <i>Pseudomonas</i> within Natural Environments and Microcosms”. In: *Microbes and Environments* 25.4 (2010), pp. 253–265. doi: 10.1264/jsme2.ME10155.

[119] Chris Smillie et al. “Mobility of plasmids”. en. In: *Microbiol Mol Biol Rev* 74.3 (Sept. 2010), pp. 434–452.

[120] G del Solar et al. “Replication and control of circular bacterial plasmids”. en. In: *Microbiol Mol Biol Rev* 62.2 (June 1998), pp. 434–464.

[121] Masahiro Sota et al. “Shifts in the host range of a promiscuous plasmid through parallel evolution of its replication initiation protein”. en. In: *ISME J* 4.12 (June 2010), pp. 1568–1580.

[122] Thibault Stalder et al. “Emerging patterns of plasmid-host coevolution that stabilize antibiotic resistance”. In: *Scientific Reports* 7.1 (July 2017), p. 4853.

[123] A B Stone. “R factors: plasmids conferring resistance to antibacterial agents”. en. In: *Sci Prog* 62.245 (1975), pp. 89–101.

[124] Murugan Subbiah et al. “Selection pressure required for long-term persistence of blaCMY-2-positive IncA/C plasmids”. en. In: *Appl Environ Microbiol* 77.13 (May 2011), pp. 4486–4493.

[125] Jian Sun et al. “Genetic analysis of the IncX4 plasmids: implications for a unique pattern in the mcr-1 acquisition”. In: *Scientific reports* 7.1 (2017), p. 424.

[126] Haruo Suzuki et al. “Predicting plasmid promiscuity based on genomic signature”. In: *Journal of bacteriology* 192.22 (2010), pp. 6045–6055.

[127] Marco Teixeira et al. “SHIP: identifying antimicrobial resistance gene transfer between plasmids”. en. In: *Bioinformatics* 39.10 (Oct. 2023).

[128] Christopher M. Thomas. “Plasmid Incompatibility”. In: *Molecular Life Sciences: An Encyclopedic Reference*. New York, NY: Springer New York, 2021, pp. 1–3. isbn: 978-1-4614-6436-5. doi: 10.1007/978-1-4614-6436-5-565-2. url: https://doi.org/10.1007/978-1-4614-6436-5-565-2.

[129] Maho Tokuda et al. “Determination of Plasmid pSN1216-29 Host Range and the Similarity in Oligonucleotide Composition Between Plasmid and Host Chromosomes”. In: *Frontiers in Microbiology* 11 (2020). issn: 1664-302X. doi: 10.3389/fmicb.2020.01187. url: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2020.01187.

[130] V A Traag, L Waltman, and N J van Eck. “From Louvain to Leiden: guaranteeing well-connected communities”. In: *Scientific Reports* 9.1 (Mar. 2019), p. 5233.

[131] Jennifer Tsang. “Bacterial plasmid addiction systems and their implications for antibiotic drug development”. en. In: *Postdoc J* 5.5 (May 2017), pp. 3–9.

[132] Angela HAM Van Hoek et al. “Acquired antibiotic resistance genes: an overview”. In: *Frontiers in microbiology* 2 (2011), p. 203.

[133] Poonam G Vinayamohan, Abraham J Pellissery, and Kumar Venkitanarayanan. “Role of horizontal gene transfer in the dissemination of antimicrobial resistance in food animal production”. In: *Current Opinion in Food Science* 47 (2022), p. 100882. ISSN: 2214-7993. DOI: `https://doi.org/10.1016/j.cofs.2022.100882`. URL: `https://www.sciencedirect.com/science/article/pii/S2214799322000844`.

[134] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: `10.1038/s41592-019-0686-2`.

[135] Jean-Nicolas Volff and Josef Altenbuchner. “A new beginning with new ends: linearisation of circular chromosomes during bacterial evolution”. In: *FEMS Microbiology Letters* 186.2 (May 2000), pp. 143–150. ISSN: 0378-1097. DOI: `10.1111/j.1574-6968.2000.tb09095.x`. eprint: `https://academic.oup.com/femsle/article-pdf/186/2/143/19115223/186-2-143.pdf`. URL: `https://doi.org/10.1111/j.1574-6968.2000.tb09095.x`.

[136] Christian JH Von Wintersdorff et al. “Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer”. In: *Frontiers in microbiology* 7 (2016), p. 173.

[137] Nicholas Waglechner, Elizabeth J. Culp, and Gerard D. Wright. “Ancient Antibiotics, Ancient Resistance”. In: *EcoSal Plus* 9.2 (2021), eESP-0027–2020. DOI: `10.1128/ecosalplus.ESP-0027-2020`. eprint: `https://journals.asm.org/doi/pdf/10.1128/ecosalplus.esp-0027-2020`. URL: `https://journals.asm.org/doi/abs/10.1128/ecosalplus.esp-0027-2020`.

[138] Karin Wallden, Angel Rivera-Calzada, and Gabriel Waksman. “Type IV secretion systems: versatility and diversity in function”. en. In: *Cell Microbiol* 12.9 (July 2010), pp. 1203–1212.

[139] Bing Wang and Dongchang Sun. “Detection of NDM-1 carbapenemase-producing Acinetobacter calcoaceticus and Acinetobacter junii in environmental samples from livestock farms”. In: *Journal of Antimicrobial Chemotherapy* 70.2 (2015), pp. 611–613.

[140] Yue Wang et al. “Antiepileptic drug carbamazepine promotes horizontal transfer of plasmid-borne multi-antibiotic resistance genes within and across bacterial genera”. In: *The ISME journal* 13.2 (2019), pp. 509–522.

[141] Yue Wang et al. “Non-antibiotic pharmaceuticals promote the transmission of multidrug resistance plasmids through intra- and intergenera conjugation”. In: *The ISME Journal* 15.9 (Mar. 2021), pp. 2493–2508. ISSN: 1751-7362. DOI: `10.1038/s41396-021-00945-7`. eprint: `https://academic.oup.com/ismej/article-pdf/15/9/2493/55257416/41396\-2021\-article\-945.pdf`. URL: `https://doi.org/10.1038/s41396-021-00945-7`.

[142] Martin Werisch, Uta Berger, and Thomas U. Berendonk. “Conjugative plasmids enable the maintenance of low cost non-transmissible plasmids”. In: *Plasmid* 91 (2017). SI: ISPB Plasmid 2016, pp. 96–104. ISSN: 0147-619X. DOI: `https://doi.org/10.1016/j.plasmid.2017.04.004`. URL: `https://www.sciencedirect.com/science/article/pii/S0147619X16301160`.

[143] Christian von Wintersdorff, John Penders, Julius Niekerk, et al. “Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer”. In: *Frontiers in Microbiology* 7 (Feb. 2016). DOI: `10.3389/fmicb.2016.00173`.

[144] N. Woodford and M.J. Ellington. “The emergence of antibiotic resistance by mutation”. In: *Clinical Microbiology and Infection* 13.1 (2007), pp. 5–18. ISSN: 1198-743X. DOI: `https://doi.org/10.1111/j.1469-0691.2006.01492.x`. URL: `https://www.sciencedirect.com/science/article/pii/S1198743X14615500`.

[145] Hirokazu Yano et al. “Host range diversification within the IncP-1 plasmid group”. In: *Microbiology* 159.Pt 11 (2013), p. 2303.

[146] Chengsong Ye et al. "Dormancy induced by oxidative damage during disinfection facilitates conjugation of ARGs through enhancing efflux and oxidative stress: A lagging response". In: *Water Research* 221 (2022), p. 118798. ISSN: 0043-1354. DOI: `https://doi.org/10.1016/j.watres.2022.118798`. URL: `https://www.sciencedirect.com/science/article/pii/S0043135422007515`.

[147] Zhigang Yu et al. "Nonnutritive sweeteners can promote the dissemination of antibiotic resistance through conjugative gene transfer". In: *The ISME Journal* 15.7 (2021), pp. 2117–2130.

[148] Xue Zhang et al. "Evolution of satellite plasmids can prolong the maintenance of newly acquired accessory genes in bacteria". In: *Nature Communications* 10.1 (Dec. 2019), p. 5809.

[149] Shuyao Zhu et al. "Augmented dissemination of antibiotic resistance elicited by non-antibiotic factors". In: *Ecotoxicology and Environmental Safety* 262 (2023), p. 115124. ISSN: 0147-6513. DOI: `https://doi.org/10.1016/j.ecoenv.2023.115124`. URL: `https://www.sciencedirect.com/science/article/pii/S0147651323006280`.

[150] Chao Zhuo et al. "Epidemic Plasmid Carrying blaCTX-M-15 in Klebsiella penumoniae in China". In: *PLOS ONE* 8.1 (Jan. 2013), pp. 1–8. DOI: `10.1371/journal.pone.0052222`. URL: `https://doi.org/10.1371/journal.pone.0052222`.

[151] Jan Zrimec and Ales Lapanje. "DNA structure at the plasmid origin-of-transfer indicates its potential transfer range". In: *Scientific Reports* 8.1 (Jan. 2018), p. 1820.

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
No replicon type	468	9957	1404.3	9020.7
Col(BS512)	0	1	0.13	0.87
Col(MG828)	24	95	16.03	102.97
Col(MP18)	0	8	1.08	6.92
Col(SD853)	9	75	11.32	72.68
Col(VCM04)	7	72	10.64	68.36
Col(YF27601)	0	1	0.13	0.87
Col(Ye4449)	8	33	5.52	35.48
Col156	97	128	30.31	194.69
Col3M	2	20	2.96	19.04
Col8282	1	2	0.4	2.6
ColE10	9	27	4.85	31.15
ColKP3	1	2	0.4	2.6
ColRNAI rep cluster 1857	412	251	89.31	573.69
ColRNAI rep cluster 1987	501	879	185.89	1194.11
ColRNAI rep cluster 1998	0	2	0.27	1.73
ColpVC	3	24	3.64	23.36
Inc11	5	63	9.16	58.84
Inc13	0	50	6.74	43.26
Inc18	0	142	19.13	122.87
IncA	12	37	6.6	42.4
IncC	208	645	114.9	738.1
IncFIA	306	4865	696.56	4474.44
IncFIB	521	6510	947.11	6083.89
IncFIC	215	1949	291.5	1872.5
IncFII	402	4615	675.81	4341.19
IncHI1A	33	80	15.22	97.78
IncHI1B	144	976	150.87	969.13
IncHI2A	213	764	131.61	845.39
IncI-gamma/K1	198	984	159.22	1022.78
IncI1	2	111	15.22	97.78
IncI1/B/O	54	136	25.59	164.41
IncI2	240	221	62.1	398.9
IncK2/Z	25	262	38.66	248.34
IncL/M	388	207	80.15	514.85
IncN	254	703	128.91	828.09
IncP	86	358	59.81	384.19
IncQ1	97	490	79.07	507.93
IncQ2	29	28	7.68	49.32
IncR	80	1281	183.33	1177.67
IncT	0	8	1.08	6.92
IncU	145	685	111.8	718.2
IncW	0	12	1.62	10.38
IncX1	124	578	94.56	607.44
IncX2	2	11	1.75	11.25
IncX3	624	196	110.46	709.54
IncX4	250	157	54.82	352.18
IncX5	10	35	6.06	38.94
IncX6	12	3	2.02	12.98
IncX7	0	4	0.54	3.46
IncX8	0	4	0.54	3.46
IncX9	0	4	0.54	3.46
IncY	2	309	41.89	269.11
rep cluster 1	3	129	17.78	114.22
rep cluster 100	0	4	0.54	3.46

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1002	0	1	0.13	0.87
rep cluster 1005	0	1	0.13	0.87
rep cluster 1006	0	37	4.98	32.02
rep cluster 101	0	4	0.54	3.46
rep cluster 1010	0	16	2.16	13.84
rep cluster 1011	0	3	0.4	2.6
rep cluster 1012	0	4	0.54	3.46
rep cluster 1015	0	1	0.13	0.87
rep cluster 1016	0	3	0.4	2.6
rep cluster 1017	47	1060	149.12	957.88
rep cluster 1018	6	83	11.99	77.01
rep cluster 1019	2	4	0.81	5.19
rep cluster 102	0	1	0.13	0.87
rep cluster 1021	0	1	0.13	0.87
rep cluster 1022	0	2	0.27	1.73
rep cluster 1023	0	4	0.54	3.46
rep cluster 1027	0	87	11.72	75.28
rep cluster 103	0	1	0.13	0.87
rep cluster 1030	0	4	0.54	3.46
rep cluster 1031	0	1	0.13	0.87
rep cluster 1039	0	1	0.13	0.87
rep cluster 104	0	2	0.27	1.73
rep cluster 1041	2	8	1.35	8.65
rep cluster 1042	0	3	0.4	2.6
rep cluster 1044	0	1	0.13	0.87
rep cluster 1046	2	6	1.08	6.92
rep cluster 1049	0	4	0.54	3.46
rep cluster 105	0	1	0.13	0.87
rep cluster 1050	0	2	0.27	1.73
rep cluster 1053	0	44	5.93	38.07
rep cluster 1057	13	49	8.35	53.65
rep cluster 1058	0	40	5.39	34.61
rep cluster 1059	0	1	0.13	0.87
rep cluster 106	0	3	0.4	2.6
rep cluster 1060	0	2	0.27	1.73
rep cluster 1061	0	16	2.16	13.84
rep cluster 1066	0	2	0.27	1.73
rep cluster 1067	0	10	1.35	8.65
rep cluster 1068	4	15	2.56	16.44
rep cluster 1069	0	31	4.18	26.82
rep cluster 107	0	6	0.81	5.19
rep cluster 1070	0	6	0.81	5.19
rep cluster 1071	4	10	1.89	12.11
rep cluster 1072	0	15	2.02	12.98
rep cluster 1074	0	2	0.27	1.73
rep cluster 1077	0	13	1.75	11.25
rep cluster 1078	1	14	2.02	12.98
rep cluster 1079	2	7	1.21	7.79
rep cluster 1081	0	3	0.4	2.6
rep cluster 1085	0	4	0.54	3.46
rep cluster 1087	0	2	0.27	1.73
rep cluster 1088	213	779	133.63	858.37
rep cluster 1092	0	1	0.13	0.87
rep cluster 1093	0	115	15.49	99.51
rep cluster 1096	0	3	0.4	2.6

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1097	9	3	1.62	10.38
rep cluster 1099	0	25	3.37	21.63
rep cluster 110	20	52	9.7	62.3
rep cluster 1100	0	2	0.27	1.73
rep cluster 1101	0	1	0.13	0.87
rep cluster 1103	0	1	0.13	0.87
rep cluster 1104	0	1	0.13	0.87
rep cluster 1110	3	68	9.56	61.44
rep cluster 1111	0	2	0.27	1.73
rep cluster 1112	0	3	0.4	2.6
rep cluster 1113	0	2	0.27	1.73
rep cluster 1115	7	85	12.39	79.61
rep cluster 1117	0	2	0.27	1.73
rep cluster 1118	96	311	54.82	352.18
rep cluster 112	0	2	0.27	1.73
rep cluster 1121	0	22	2.96	19.04
rep cluster 1123	0	2	0.27	1.73
rep cluster 1124	0	1	0.13	0.87
rep cluster 1125	0	1	0.13	0.87
rep cluster 1127	0	3	0.4	2.6
rep cluster 1128	0	3	0.4	2.6
rep cluster 1129	0	1	0.13	0.87
rep cluster 1131	0	2	0.27	1.73
rep cluster 1134	2	18	2.69	17.31
rep cluster 1136	0	3	0.4	2.6
rep cluster 1138	0	2	0.27	1.73
rep cluster 1139	0	1	0.13	0.87
rep cluster 114	0	1	0.13	0.87
rep cluster 1140	0	3	0.4	2.6
rep cluster 1142	16	107	16.57	106.43
rep cluster 1143	0	3	0.4	2.6
rep cluster 1144	0	1	0.13	0.87
rep cluster 1148	0	1	0.13	0.87
rep cluster 1149	0	5	0.67	4.33
rep cluster 1151	0	9	1.21	7.79
rep cluster 1152	2	5	0.94	6.06
rep cluster 1154	0	9	1.21	7.79
rep cluster 1155	6	23	3.91	25.09
rep cluster 1156	0	2	0.27	1.73
rep cluster 1159	0	5	0.67	4.33
rep cluster 1164	0	17	2.29	14.71
rep cluster 1166	0	2	0.27	1.73
rep cluster 1167	0	1	0.13	0.87
rep cluster 1168	0	8	1.08	6.92
rep cluster 1169	2	41	5.79	37.21
rep cluster 1171	0	19	2.56	16.44
rep cluster 1172	2	183	24.92	160.08
rep cluster 1173	0	2	0.27	1.73
rep cluster 1174	0	1	0.13	0.87
rep cluster 1175	0	2	0.27	1.73
rep cluster 1176	0	1	0.13	0.87
rep cluster 1178	0	2	0.27	1.73
rep cluster 1180	0	1	0.13	0.87
rep cluster 1181	0	4	0.54	3.46
rep cluster 1186	0	1	0.13	0.87

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1187	0	2	0.27	1.73
rep cluster 1188	0	2	0.27	1.73
rep cluster 1189	0	1	0.13	0.87
rep cluster 119	0	5	0.67	4.33
rep cluster 1190	0	1	0.13	0.87
rep cluster 1192	0	1	0.13	0.87
rep cluster 1193	0	4	0.54	3.46
rep cluster 1194	0	2	0.27	1.73
rep cluster 1195	113	57	22.9	147.1
rep cluster 1197	0	126	16.97	109.03
rep cluster 12	0	16	2.16	13.84
rep cluster 120	0	3	0.4	2.6
rep cluster 1200	0	1	0.13	0.87
rep cluster 1202	0	1	0.13	0.87
rep cluster 1203	0	10	1.35	8.65
rep cluster 1205	0	4	0.54	3.46
rep cluster 1206	0	1	0.13	0.87
rep cluster 1207	0	1	0.13	0.87
rep cluster 1208	0	1	0.13	0.87
rep cluster 1209	0	3	0.4	2.6
rep cluster 1212	0	1	0.13	0.87
rep cluster 1213	0	1	0.13	0.87
rep cluster 1214	0	1	0.13	0.87
rep cluster 1215	3	380	51.59	331.41
rep cluster 1216	0	14	1.89	12.11
rep cluster 1218	0	72	9.7	62.3
rep cluster 1219	0	15	2.02	12.98
rep cluster 122	0	1	0.13	0.87
rep cluster 1220	0	46	6.2	39.8
rep cluster 1222	0	1	0.13	0.87
rep cluster 1223	0	4	0.54	3.46
rep cluster 1224	0	2	0.27	1.73
rep cluster 1225	0	1	0.13	0.87
rep cluster 1226	11	60	9.56	61.44
rep cluster 1229	2	46	6.47	41.53
rep cluster 123	0	1	0.13	0.87
rep cluster 1230	8	20	3.77	24.23
rep cluster 1231	0	1	0.13	0.87
rep cluster 1232	0	69	9.29	59.71
rep cluster 1233	0	22	2.96	19.04
rep cluster 1234	0	1	0.13	0.87
rep cluster 1239	0	2	0.27	1.73
rep cluster 124	0	1	0.13	0.87
rep cluster 1241	0	3	0.4	2.6
rep cluster 1244	0	37	4.98	32.02
rep cluster 1246	0	2	0.27	1.73
rep cluster 1247	0	5	0.67	4.33
rep cluster 1249	0	37	4.98	32.02
rep cluster 125	65	240	41.08	263.92
rep cluster 1254	48	389	58.87	378.13
rep cluster 1255	0	32	4.31	27.69
rep cluster 1256	4	25	3.91	25.09
rep cluster 1257	0	3	0.4	2.6
rep cluster 1258	0	1	0.13	0.87
rep cluster 1259	2	23	3.37	21.63

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1260	0	2	0.27	1.73
rep cluster 1261	0	11	1.48	9.52
rep cluster 1262	0	9	1.21	7.79
rep cluster 1264	0	3	0.4	2.6
rep cluster 1265	4	23	3.64	23.36
rep cluster 1269	0	3	0.4	2.6
rep cluster 1274	0	1	0.13	0.87
rep cluster 1277	0	1	0.13	0.87
rep cluster 1278	0	3	0.4	2.6
rep cluster 1280	0	13	1.75	11.25
rep cluster 1281	3	43	6.2	39.8
rep cluster 1282	0	31	4.18	26.82
rep cluster 1283	0	19	2.56	16.44
rep cluster 1285	0	1	0.13	0.87
rep cluster 1286	0	1	0.13	0.87
rep cluster 1287	0	2	0.27	1.73
rep cluster 1288	0	9	1.21	7.79
rep cluster 1289	0	10	1.35	8.65
rep cluster 1290	0	1	0.13	0.87
rep cluster 1291	0	17	2.29	14.71
rep cluster 1294	0	7	0.94	6.06
rep cluster 1295	0	2	0.27	1.73
rep cluster 1297	0	1	0.13	0.87
rep cluster 1299	0	7	0.94	6.06
rep cluster 1300	0	3	0.4	2.6
rep cluster 1301	0	1	0.13	0.87
rep cluster 1302	0	11	1.48	9.52
rep cluster 1304	2	20	2.96	19.04
rep cluster 1306	0	2	0.27	1.73
rep cluster 1307	0	42	5.66	36.34
rep cluster 1309	0	1	0.13	0.87
rep cluster 1310	0	2	0.27	1.73
rep cluster 1311	0	12	1.62	10.38
rep cluster 1313	0	1	0.13	0.87
rep cluster 1316	0	1	0.13	0.87
rep cluster 1317	0	11	1.48	9.52
rep cluster 1319	0	16	2.16	13.84
rep cluster 1320	2	6	1.08	6.92
rep cluster 1322	0	1	0.13	0.87
rep cluster 1323	0	20	2.69	17.31
rep cluster 1325	0	1	0.13	0.87
rep cluster 1328	0	89	11.99	77.01
rep cluster 1331	0	1	0.13	0.87
rep cluster 1332	3	10	1.75	11.25
rep cluster 1333	2	9	1.48	9.52
rep cluster 1336	0	9	1.21	7.79
rep cluster 1337	2	1	0.4	2.6
rep cluster 1340	0	1	0.13	0.87
rep cluster 1341	0	2	0.27	1.73
rep cluster 1344	0	1	0.13	0.87
rep cluster 1345	0	3	0.4	2.6
rep cluster 1346	0	1	0.13	0.87
rep cluster 1347	0	1	0.13	0.87
rep cluster 135	0	1	0.13	0.87
rep cluster 1351	0	1	0.13	0.87

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1355	0	1	0.13	0.87
rep cluster 1357	0	7	0.94	6.06
rep cluster 1359	0	1	0.13	0.87
rep cluster 136	0	3	0.4	2.6
rep cluster 1362	0	5	0.67	4.33
rep cluster 1364	3	77	10.78	69.22
rep cluster 1367	33	27	8.08	51.92
rep cluster 1368	2	10	1.62	10.38
rep cluster 137	0	70	9.43	60.57
rep cluster 1370	0	1	0.13	0.87
rep cluster 1371	0	6	0.81	5.19
rep cluster 1374	0	29	3.91	25.09
rep cluster 1375	0	2	0.27	1.73
rep cluster 1376	2	2	0.54	3.46
rep cluster 1377	0	3	0.4	2.6
rep cluster 1378	2	4	0.81	5.19
rep cluster 1379	0	1	0.13	0.87
rep cluster 138	0	8	1.08	6.92
rep cluster 1380	0	4	0.54	3.46
rep cluster 1381	0	5	0.67	4.33
rep cluster 1392	0	2	0.27	1.73
rep cluster 1393	0	2	0.27	1.73
rep cluster 1394	0	1	0.13	0.87
rep cluster 1396	0	15	2.02	12.98
rep cluster 1398	0	9	1.21	7.79
rep cluster 1399	0	1	0.13	0.87
rep cluster 1403	0	9	1.21	7.79
rep cluster 1407	0	1	0.13	0.87
rep cluster 141	0	5	0.67	4.33
rep cluster 1410	0	41	5.52	35.48
rep cluster 1416	0	1	0.13	0.87
rep cluster 1417	0	10	1.35	8.65
rep cluster 1418	56	588	86.75	557.25
rep cluster 142	0	1	0.13	0.87
rep cluster 1420	0	18	2.42	15.58
rep cluster 1426	0	3	0.4	2.6
rep cluster 1431	0	1	0.13	0.87
rep cluster 1433	0	108	14.55	93.45
rep cluster 1435	0	1	0.13	0.87
rep cluster 1436	0	4	0.54	3.46
rep cluster 1439	0	19	2.56	16.44
rep cluster 144	0	1	0.13	0.87
rep cluster 1440	0	2	0.27	1.73
rep cluster 1441	0	29	3.91	25.09
rep cluster 1442	2	51	7.14	45.86
rep cluster 1443	0	7	0.94	6.06
rep cluster 1444	17	18	4.71	30.29
rep cluster 1445	0	2	0.27	1.73
rep cluster 1448	0	33	4.45	28.55
rep cluster 1449	0	24	3.23	20.77
rep cluster 1450	0	1	0.13	0.87
rep cluster 1451	0	8	1.08	6.92
rep cluster 1454	0	23	3.1	19.9
rep cluster 1455	2	0	0.27	1.73
rep cluster 1456	0	4	0.54	3.46

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 146	0	4	0.54	3.46
rep cluster 1461	0	1	0.13	0.87
rep cluster 1462	0	5	0.67	4.33
rep cluster 1464	0	3	0.4	2.6
rep cluster 1465	0	77	10.37	66.63
rep cluster 1467	0	52	7	45
rep cluster 1469	0	1	0.13	0.87
rep cluster 1474	0	4	0.54	3.46
rep cluster 1478	0	3	0.4	2.6
rep cluster 1480	0	2	0.27	1.73
rep cluster 1481	0	6	0.81	5.19
rep cluster 1485	0	4	0.54	3.46
rep cluster 1486	25	32	7.68	49.32
rep cluster 1489	0	1	0.13	0.87
rep cluster 1493	0	1	0.13	0.87
rep cluster 1495	0	1	0.13	0.87
rep cluster 1497	0	1	0.13	0.87
rep cluster 1498	0	1	0.13	0.87
rep cluster 150	0	23	3.1	19.9
rep cluster 1500	0	7	0.94	6.06
rep cluster 1501	0	1	0.13	0.87
rep cluster 1502	0	14	1.89	12.11
rep cluster 1503	3	10	1.75	11.25
rep cluster 1506	2	37	5.25	33.75
rep cluster 1508	0	1	0.13	0.87
rep cluster 1509	0	6	0.81	5.19
rep cluster 151	0	2	0.27	1.73
rep cluster 1510	0	8	1.08	6.92
rep cluster 1511	0	6	0.81	5.19
rep cluster 1513	0	1	0.13	0.87
rep cluster 1515	0	1	0.13	0.87
rep cluster 1516	0	5	0.67	4.33
rep cluster 1519	0	60	8.08	51.92
rep cluster 152	0	1	0.13	0.87
rep cluster 1520	0	23	3.1	19.9
rep cluster 1522	0	3	0.4	2.6
rep cluster 1523	0	3	0.4	2.6
rep cluster 1524	0	2	0.27	1.73
rep cluster 1527	2	11	1.75	11.25
rep cluster 1529	0	12	1.62	10.38
rep cluster 153	0	3	0.4	2.6
rep cluster 1533	0	2	0.27	1.73
rep cluster 1534	0	1	0.13	0.87
rep cluster 1535	0	2	0.27	1.73
rep cluster 1536	0	10	1.35	8.65
rep cluster 1537	0	3	0.4	2.6
rep cluster 154	0	3	0.4	2.6
rep cluster 1542	0	1	0.13	0.87
rep cluster 1546	0	2	0.27	1.73
rep cluster 1547	0	4	0.54	3.46
rep cluster 1548	0	1	0.13	0.87
rep cluster 155	3	77	10.78	69.22
rep cluster 1551	0	1	0.13	0.87
rep cluster 1552	0	1	0.13	0.87
rep cluster 1556	0	1	0.13	0.87

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 156	6	17	3.1	19.9
rep cluster 1561	0	1	0.13	0.87
rep cluster 1566	0	7	0.94	6.06
rep cluster 1567	0	21	2.83	18.17
rep cluster 1570	0	1	0.13	0.87
rep cluster 1573	0	1	0.13	0.87
rep cluster 1574	0	1	0.13	0.87
rep cluster 1575	0	7	0.94	6.06
rep cluster 1578	0	1	0.13	0.87
rep cluster 158	0	1	0.13	0.87
rep cluster 1581	0	2	0.27	1.73
rep cluster 1582	0	1	0.13	0.87
rep cluster 159	0	1	0.13	0.87
rep cluster 1590	0	4	0.54	3.46
rep cluster 1592	0	1	0.13	0.87
rep cluster 1593	0	2	0.27	1.73
rep cluster 1596	0	11	1.48	9.52
rep cluster 1598	0	1	0.13	0.87
rep cluster 1599	0	3	0.4	2.6
rep cluster 16	0	4	0.54	3.46
rep cluster 160	0	5	0.67	4.33
rep cluster 1601	0	3	0.4	2.6
rep cluster 1603	0	2	0.27	1.73
rep cluster 1604	0	1	0.13	0.87
rep cluster 1605	0	1	0.13	0.87
rep cluster 1607	0	36	4.85	31.15
rep cluster 1608	0	1	0.13	0.87
rep cluster 1609	0	1	0.13	0.87
rep cluster 1610	0	4	0.54	3.46
rep cluster 1613	0	4	0.54	3.46
rep cluster 1614	0	1	0.13	0.87
rep cluster 1617	0	9	1.21	7.79
rep cluster 1618	0	6	0.81	5.19
rep cluster 1621	0	32	4.31	27.69
rep cluster 1622	0	7	0.94	6.06
rep cluster 1625	0	14	1.89	12.11
rep cluster 1626	0	6	0.81	5.19
rep cluster 1628	0	11	1.48	9.52
rep cluster 1629	0	12	1.62	10.38
rep cluster 163	22	113	18.19	116.81
rep cluster 1631	0	19	2.56	16.44
rep cluster 1634	0	1	0.13	0.87
rep cluster 1636	0	1	0.13	0.87
rep cluster 1637	0	1	0.13	0.87
rep cluster 1639	0	1	0.13	0.87
rep cluster 164	6	52	7.81	50.19
rep cluster 1640	0	2	0.27	1.73
rep cluster 1642	0	3	0.4	2.6
rep cluster 1645	0	4	0.54	3.46
rep cluster 1648	0	3	0.4	2.6
rep cluster 1650	0	1	0.13	0.87
rep cluster 1652	0	1	0.13	0.87
rep cluster 1656	0	62	8.35	53.65
rep cluster 1657	2	3	0.67	4.33
rep cluster 166	0	1	0.13	0.87

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1662	0	19	2.56	16.44
rep cluster 1664	0	2	0.27	1.73
rep cluster 1665	4	7	1.48	9.52
rep cluster 167	0	127	17.11	109.89
rep cluster 1671	2	3	0.67	4.33
rep cluster 1673	0	1	0.13	0.87
rep cluster 1674	0	1	0.13	0.87
rep cluster 1675	0	9	1.21	7.79
rep cluster 1676	0	2	0.27	1.73
rep cluster 1678	0	4	0.54	3.46
rep cluster 1679	0	1	0.13	0.87
rep cluster 168	0	1	0.13	0.87
rep cluster 1685	0	31	4.18	26.82
rep cluster 1688	0	1	0.13	0.87
rep cluster 169	0	35	4.71	30.29
rep cluster 1693	0	1	0.13	0.87
rep cluster 1695	0	1	0.13	0.87
rep cluster 1696	0	1	0.13	0.87
rep cluster 1698	7	35	5.66	36.34
rep cluster 170	0	2	0.27	1.73
rep cluster 1701	0	3	0.4	2.6
rep cluster 1702	12	19	4.18	26.82
rep cluster 1704	8	89	13.07	83.93
rep cluster 1705	0	3	0.4	2.6
rep cluster 1706	0	3	0.4	2.6
rep cluster 1707	0	1	0.13	0.87
rep cluster 1710	0	3	0.4	2.6
rep cluster 1712	0	1	0.13	0.87
rep cluster 1713	0	34	4.58	29.42
rep cluster 1714	0	1	0.13	0.87
rep cluster 1717	0	3	0.4	2.6
rep cluster 1719	0	1	0.13	0.87
rep cluster 172	3	29	4.31	27.69
rep cluster 1720	0	3	0.4	2.6
rep cluster 1721	0	159	21.42	137.58
rep cluster 1722	0	2	0.27	1.73
rep cluster 1723	0	3	0.4	2.6
rep cluster 1725	0	1	0.13	0.87
rep cluster 1729	0	1	0.13	0.87
rep cluster 1730	0	3	0.4	2.6
rep cluster 1731	0	1	0.13	0.87
rep cluster 1732	0	3	0.4	2.6
rep cluster 1733	98	478	77.59	498.41
rep cluster 1734	0	1	0.13	0.87
rep cluster 1735	0	2	0.27	1.73
rep cluster 1737	0	1	0.13	0.87
rep cluster 1738	0	11	1.48	9.52
rep cluster 1739	0	6	0.81	5.19
rep cluster 1740	0	4	0.54	3.46
rep cluster 1741	0	1	0.13	0.87
rep cluster 1742	9	61	9.43	60.57
rep cluster 1744	0	4	0.54	3.46
rep cluster 1747	0	1	0.13	0.87
rep cluster 1749	0	3	0.4	2.6
rep cluster 1751	0	13	1.75	11.25

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1756	2	7	1.21	7.79
rep cluster 1760	0	19	2.56	16.44
rep cluster 1763	0	84	11.32	72.68
rep cluster 1767	0	8	1.08	6.92
rep cluster 1769	0	1	0.13	0.87
rep cluster 177	0	2	0.27	1.73
rep cluster 1775	3	3	0.81	5.19
rep cluster 1776	0	3	0.4	2.6
rep cluster 1777	0	2	0.27	1.73
rep cluster 1778	134	149	38.12	244.88
rep cluster 1779	0	2	0.27	1.73
rep cluster 1780	0	6	0.81	5.19
rep cluster 1787	0	5	0.67	4.33
rep cluster 179	0	2	0.27	1.73
rep cluster 1791	0	4	0.54	3.46
rep cluster 1798	0	5	0.67	4.33
rep cluster 18	0	1	0.13	0.87
rep cluster 180	0	132	17.78	114.22
rep cluster 1800	0	2	0.27	1.73
rep cluster 1801	0	1	0.13	0.87
rep cluster 1804	0	22	2.96	19.04
rep cluster 1805	0	1	0.13	0.87
rep cluster 1806	0	7	0.94	6.06
rep cluster 1809	0	1	0.13	0.87
rep cluster 181	0	4	0.54	3.46
rep cluster 1814	0	3	0.4	2.6
rep cluster 1821	0	14	1.89	12.11
rep cluster 1822	0	1	0.13	0.87
rep cluster 1825	0	5	0.67	4.33
rep cluster 1826	0	1	0.13	0.87
rep cluster 1827	0	11	1.48	9.52
rep cluster 183	4	57	8.22	52.78
rep cluster 1830	0	14	1.89	12.11
rep cluster 1832	0	17	2.29	14.71
rep cluster 1833	2	1	0.4	2.6
rep cluster 1836	0	24	3.23	20.77
rep cluster 1837	0	1	0.13	0.87
rep cluster 1838	0	1	0.13	0.87
rep cluster 1840	0	1	0.13	0.87
rep cluster 1848	0	13	1.75	11.25
rep cluster 185	37	363	53.88	346.12
rep cluster 1850	0	14	1.89	12.11
rep cluster 1852	0	3	0.4	2.6
rep cluster 1855	0	7	0.94	6.06
rep cluster 1857	0	4	0.54	3.46
rep cluster 1858	0	1	0.13	0.87
rep cluster 1861	0	2	0.27	1.73
rep cluster 1863	0	5	0.67	4.33
rep cluster 1864	0	4	0.54	3.46
rep cluster 1868	0	2	0.27	1.73
rep cluster 1869	2	44	6.2	39.8
rep cluster 187	0	5	0.67	4.33
rep cluster 1871	0	5	0.67	4.33
rep cluster 1874	0	9	1.21	7.79
rep cluster 1875	0	11	1.48	9.52

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1879	0	14	1.89	12.11
rep cluster 1880	2	25	3.64	23.36
rep cluster 1887	0	1	0.13	0.87
rep cluster 1891	0	2	0.27	1.73
rep cluster 1892	0	5	0.67	4.33
rep cluster 1895	2	23	3.37	21.63
rep cluster 1897	0	1	0.13	0.87
rep cluster 190	2	5	0.94	6.06
rep cluster 1902	0	4	0.54	3.46
rep cluster 1903	0	15	2.02	12.98
rep cluster 1905	0	1	0.13	0.87
rep cluster 1906	0	34	4.58	29.42
rep cluster 1907	2	3	0.67	4.33
rep cluster 1908	0	2	0.27	1.73
rep cluster 1909	0	12	1.62	10.38
rep cluster 1910	0	15	2.02	12.98
rep cluster 1911	0	1	0.13	0.87
rep cluster 1912	0	40	5.39	34.61
rep cluster 1913	0	2	0.27	1.73
rep cluster 1914	0	2	0.27	1.73
rep cluster 1919	0	5	0.67	4.33
rep cluster 1922	0	16	2.16	13.84
rep cluster 1925	5	2	0.94	6.06
rep cluster 1927	0	3	0.4	2.6
rep cluster 1929	0	1	0.13	0.87
rep cluster 193	0	7	0.94	6.06
rep cluster 1930	0	1	0.13	0.87
rep cluster 1931	7	0	0.94	6.06
rep cluster 1934	0	3	0.4	2.6
rep cluster 1938	0	3	0.4	2.6
rep cluster 1943	0	3	0.4	2.6
rep cluster 1944	0	3	0.4	2.6
rep cluster 1945	0	2	0.27	1.73
rep cluster 1947	0	15	2.02	12.98
rep cluster 1948	0	55	7.41	47.59
rep cluster 1949	0	1	0.13	0.87
rep cluster 195	0	8	1.08	6.92
rep cluster 1950	0	5	0.67	4.33
rep cluster 1951	0	2	0.27	1.73
rep cluster 1952	0	3	0.4	2.6
rep cluster 1953	2	3	0.67	4.33
rep cluster 1956	0	4	0.54	3.46
rep cluster 1958	0	5	0.67	4.33
rep cluster 196	0	14	1.89	12.11
rep cluster 1964	0	2	0.27	1.73
rep cluster 1965	0	2	0.27	1.73
rep cluster 1966	3	3	0.81	5.19
rep cluster 1967	0	4	0.54	3.46
rep cluster 1968	0	9	1.21	7.79
rep cluster 197	0	12	1.62	10.38
rep cluster 1970	0	1	0.13	0.87
rep cluster 1971	0	17	2.29	14.71
rep cluster 1974	0	8	1.08	6.92
rep cluster 1975	0	3	0.4	2.6
rep cluster 1977	0	4	0.54	3.46

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 198	0	1	0.13	0.87
rep cluster 1980	0	8	1.08	6.92
rep cluster 1981	0	1	0.13	0.87
rep cluster 1985	0	1	0.13	0.87
rep cluster 1986	0	6	0.81	5.19
rep cluster 1988	0	2	0.27	1.73
rep cluster 199	0	7	0.94	6.06
rep cluster 1991	0	48	6.47	41.53
rep cluster 1992	2	3	0.67	4.33
rep cluster 1993	0	5	0.67	4.33
rep cluster 2002	0	4	0.54	3.46
rep cluster 2005	0	4	0.54	3.46
rep cluster 2006	0	13	1.75	11.25
rep cluster 2007	0	1	0.13	0.87
rep cluster 2011	0	1	0.13	0.87
rep cluster 2013	0	7	0.94	6.06
rep cluster 2014	0	1	0.13	0.87
rep cluster 2015	0	1	0.13	0.87
rep cluster 2020	0	13	1.75	11.25
rep cluster 2021	0	13	1.75	11.25
rep cluster 2022	4	21	3.37	21.63
rep cluster 2023	3	5	1.08	6.92
rep cluster 2025	0	23	3.1	19.9
rep cluster 203	2	0	0.27	1.73
rep cluster 2037	0	22	2.96	19.04
rep cluster 204	0	11	1.48	9.52
rep cluster 2042	2	11	1.75	11.25
rep cluster 2045	0	5	0.67	4.33
rep cluster 2046	0	4	0.54	3.46
rep cluster 2051	0	1	0.13	0.87
rep cluster 2061	0	2	0.27	1.73
rep cluster 2062	0	1	0.13	0.87
rep cluster 207	0	1	0.13	0.87
rep cluster 2077	0	5	0.67	4.33
rep cluster 2078	0	6	0.81	5.19
rep cluster 208	0	2	0.27	1.73
rep cluster 2082	0	2	0.27	1.73
rep cluster 209	0	4	0.54	3.46
rep cluster 210	0	1	0.13	0.87
rep cluster 2100	12	371	51.59	331.41
rep cluster 2101	0	16	2.16	13.84
rep cluster 2104	1	19	2.69	17.31
rep cluster 2109	0	4	0.54	3.46
rep cluster 211	0	5	0.67	4.33
rep cluster 2117	0	5	0.67	4.33
rep cluster 2119	7	51	7.81	50.19
rep cluster 2120	0	18	2.42	15.58
rep cluster 2127	0	9	1.21	7.79
rep cluster 213	0	1	0.13	0.87
rep cluster 2131	61	460	70.18	450.82
rep cluster 2132	0	1	0.13	0.87
rep cluster 2133	0	8	1.08	6.92
rep cluster 2134	0	10	1.35	8.65
rep cluster 214	0	11	1.48	9.52
rep cluster 2140	0	1	0.13	0.87

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 2141	1	16	2.29	14.71
rep cluster 2146	5	21	3.5	22.5
rep cluster 2147	0	1	0.13	0.87
rep cluster 215	0	3	0.4	2.6
rep cluster 2150	16	22	5.12	32.88
rep cluster 216	0	87	11.72	75.28
rep cluster 2164	0	2	0.27	1.73
rep cluster 2165	0	11	1.48	9.52
rep cluster 2167	0	1	0.13	0.87
rep cluster 2169	0	1	0.13	0.87
rep cluster 2172	0	1	0.13	0.87
rep cluster 2175	10	13	3.1	19.9
rep cluster 2176	0	26	3.5	22.5
rep cluster 2183	131	1320	195.46	1255.54
rep cluster 2187	0	6	0.81	5.19
rep cluster 219	0	3	0.4	2.6
rep cluster 2194	0	7	0.94	6.06
rep cluster 220	0	3	0.4	2.6
rep cluster 2201	0	2	0.27	1.73
rep cluster 2203	0	30	4.04	25.96
rep cluster 2214	86	117	27.35	175.65
rep cluster 2215	0	19	2.56	16.44
rep cluster 2218	0	10	1.35	8.65
rep cluster 2221	0	8	1.08	6.92
rep cluster 2222	0	33	4.45	28.55
rep cluster 2229	0	3	0.4	2.6
rep cluster 2232	0	78	10.51	67.49
rep cluster 2234	1	2	0.4	2.6
rep cluster 2238	2	7	1.21	7.79
rep cluster 224	0	5	0.67	4.33
rep cluster 2241	0	1	0.13	0.87
rep cluster 2244	26	488	69.24	444.76
rep cluster 2246	0	9	1.21	7.79
rep cluster 225	0	2	0.27	1.73
rep cluster 2252	0	2	0.27	1.73
rep cluster 2254	0	2	0.27	1.73
rep cluster 2256	0	1	0.13	0.87
rep cluster 2257	0	1	0.13	0.87
rep cluster 2263	0	1	0.13	0.87
rep cluster 2264	0	1	0.13	0.87
rep cluster 2268	0	17	2.29	14.71
rep cluster 2272	33	98	17.65	113.35
rep cluster 2273	0	10	1.35	8.65
rep cluster 2281	0	1	0.13	0.87
rep cluster 2282	0	2	0.27	1.73
rep cluster 2283	0	5	0.67	4.33
rep cluster 2287	0	1	0.13	0.87
rep cluster 2288	0	1	0.13	0.87
rep cluster 229	0	12	1.62	10.38
rep cluster 2294	0	2	0.27	1.73
rep cluster 2295	0	2	0.27	1.73
rep cluster 2298	0	1	0.13	0.87
rep cluster 2299	0	2	0.27	1.73
rep cluster 2300	2	6	1.08	6.92
rep cluster 2301	38	31	9.29	59.71

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 2302	0	9	1.21	7.79
rep cluster 2307	0	3	0.4	2.6
rep cluster 231	0	5	0.67	4.33
rep cluster 2310	0	3	0.4	2.6
rep cluster 2313	0	94	12.66	81.34
rep cluster 2314	0	10	1.35	8.65
rep cluster 2315	0	6	0.81	5.19
rep cluster 2318	0	1	0.13	0.87
rep cluster 2319	0	7	0.94	6.06
rep cluster 2324	11	15	3.5	22.5
rep cluster 2327	8	44	7	45
rep cluster 2329	0	1	0.13	0.87
rep cluster 2335	724	422	154.37	991.63
rep cluster 234	0	5	0.67	4.33
rep cluster 2341	3	19	2.96	19.04
rep cluster 2344	0	2	0.27	1.73
rep cluster 235	0	1	0.13	0.87
rep cluster 2350	384	811	160.97	1034.03
rep cluster 2351	10	52	8.35	53.65
rep cluster 2358	136	203	45.66	293.34
rep cluster 2359	0	13	1.75	11.25
rep cluster 236	0	2	0.27	1.73
rep cluster 2360	0	2	0.27	1.73
rep cluster 2361	1	1	0.27	1.73
rep cluster 2366	0	2	0.27	1.73
rep cluster 237	5	31	4.85	31.15
rep cluster 2370	36	266	40.68	261.32
rep cluster 2371	1	8	1.21	7.79
rep cluster 2373	63	156	29.5	189.5
rep cluster 2374	6	27	4.45	28.55
rep cluster 239	0	1	0.13	0.87
rep cluster 2392	8	55	8.49	54.51
rep cluster 2398	0	1	0.13	0.87
rep cluster 240	0	3	0.4	2.6
rep cluster 2401	81	63	19.4	124.6
rep cluster 2405	6	2	1.08	6.92
rep cluster 2406	5	3	1.08	6.92
rep cluster 2409	0	14	1.89	12.11
rep cluster 241	0	1	0.13	0.87
rep cluster 2414	0	1	0.13	0.87
rep cluster 2419	0	2	0.27	1.73
rep cluster 242	0	1	0.13	0.87
rep cluster 2422	0	46	6.2	39.8
rep cluster 2425	0	1	0.13	0.87
rep cluster 2427	0	1	0.13	0.87
rep cluster 2431	0	1	0.13	0.87
rep cluster 2439	0	19	2.56	16.44
rep cluster 244	0	1	0.13	0.87
rep cluster 2442	0	4	0.54	3.46
rep cluster 2446	0	39	5.25	33.75
rep cluster 2448	0	1	0.13	0.87
rep cluster 2455	0	1	0.13	0.87
rep cluster 2464	0	3	0.4	2.6
rep cluster 2466	2	1	0.4	2.6
rep cluster 2468	0	2	0.27	1.73

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 2469	0	20	2.69	17.31
rep cluster 2470	0	3	0.4	2.6
rep cluster 2472	0	1	0.13	0.87
rep cluster 2475	0	4	0.54	3.46
rep cluster 2479	0	1	0.13	0.87
rep cluster 2483	0	6	0.81	5.19
rep cluster 2484	0	3	0.4	2.6
rep cluster 2486	0	10	1.35	8.65
rep cluster 25	0	1	0.13	0.87
rep cluster 251	0	5	0.67	4.33
rep cluster 252	4	73	10.37	66.63
rep cluster 256	0	13	1.75	11.25
rep cluster 260	0	1	0.13	0.87
rep cluster 261	0	3	0.4	2.6
rep cluster 263	0	5	0.67	4.33
rep cluster 267	0	1	0.13	0.87
rep cluster 268	0	10	1.35	8.65
rep cluster 271	0	2	0.27	1.73
rep cluster 273	0	11	1.48	9.52
rep cluster 275	0	1	0.13	0.87
rep cluster 279	0	1	0.13	0.87
rep cluster 28	0	10	1.35	8.65
rep cluster 280	0	1	0.13	0.87
rep cluster 281	0	1	0.13	0.87
rep cluster 282	0	17	2.29	14.71
rep cluster 283	0	152	20.48	131.52
rep cluster 284	0	1	0.13	0.87
rep cluster 287	4	60	8.62	55.38
rep cluster 288	0	57	7.68	49.32
rep cluster 289	0	6	0.81	5.19
rep cluster 29	0	1	0.13	0.87
rep cluster 290	0	1	0.13	0.87
rep cluster 292	0	4	0.54	3.46
rep cluster 293	0	1	0.13	0.87
rep cluster 294	0	7	0.94	6.06
rep cluster 296	0	3	0.4	2.6
rep cluster 3	0	18	2.42	15.58
rep cluster 30	0	8	1.08	6.92
rep cluster 300	0	3	0.4	2.6
rep cluster 304	0	1	0.13	0.87
rep cluster 305	0	5	0.67	4.33
rep cluster 306	0	2	0.27	1.73
rep cluster 308	0	29	3.91	25.09
rep cluster 309	0	1	0.13	0.87
rep cluster 310	0	2	0.27	1.73
rep cluster 312	2	25	3.64	23.36
rep cluster 313	0	6	0.81	5.19
rep cluster 314	0	1	0.13	0.87
rep cluster 317	0	6	0.81	5.19
rep cluster 32	0	1	0.13	0.87
rep cluster 321	0	4	0.54	3.46
rep cluster 322	0	5	0.67	4.33
rep cluster 323	0	8	1.08	6.92
rep cluster 324	0	3	0.4	2.6
rep cluster 327	2	4	0.81	5.19

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 328	0	7	0.94	6.06
rep cluster 329	0	3	0.4	2.6
rep cluster 33	0	3	0.4	2.6
rep cluster 332	0	6	0.81	5.19
rep cluster 334	0	1	0.13	0.87
rep cluster 335	0	1	0.13	0.87
rep cluster 337	0	8	1.08	6.92
rep cluster 339	0	5	0.67	4.33
rep cluster 341	0	1	0.13	0.87
rep cluster 342	0	18	2.42	15.58
rep cluster 345	0	1	0.13	0.87
rep cluster 346	0	1	0.13	0.87
rep cluster 349	0	3	0.4	2.6
rep cluster 352	7	7	1.89	12.11
rep cluster 355	0	4	0.54	3.46
rep cluster 356	0	7	0.94	6.06
rep cluster 359	0	6	0.81	5.19
rep cluster 360	0	1	0.13	0.87
rep cluster 361	0	50	6.74	43.26
rep cluster 364	0	5	0.67	4.33
rep cluster 366	0	1	0.13	0.87
rep cluster 369	2	8	1.35	8.65
rep cluster 37	2	0	0.27	1.73
rep cluster 371	0	2	0.27	1.73
rep cluster 373	0	1	0.13	0.87
rep cluster 377	0	54	7.27	46.73
rep cluster 378	0	15	2.02	12.98
rep cluster 38	0	3	0.4	2.6
rep cluster 380	0	7	0.94	6.06
rep cluster 382	0	16	2.16	13.84
rep cluster 383	0	1	0.13	0.87
rep cluster 385	8	144	20.48	131.52
rep cluster 386	0	7	0.94	6.06
rep cluster 389	0	1	0.13	0.87
rep cluster 390	0	1	0.13	0.87
rep cluster 392	2	23	3.37	21.63
rep cluster 398	0	22	2.96	19.04
rep cluster 4	0	1	0.13	0.87
rep cluster 40	0	13	1.75	11.25
rep cluster 402	0	6	0.81	5.19
rep cluster 405	0	5	0.67	4.33
rep cluster 408	0	1	0.13	0.87
rep cluster 409	0	2	0.27	1.73
rep cluster 41	0	4	0.54	3.46
rep cluster 412	0	1	0.13	0.87
rep cluster 413	0	2	0.27	1.73
rep cluster 414	2	37	5.25	33.75
rep cluster 416	0	2	0.27	1.73
rep cluster 418	0	19	2.56	16.44
rep cluster 419	0	1	0.13	0.87
rep cluster 420	3	2	0.67	4.33
rep cluster 421	0	1	0.13	0.87
rep cluster 427	0	25	3.37	21.63
rep cluster 428	2	4	0.81	5.19
rep cluster 429	8	70	10.51	67.49

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 43	0	2	0.27	1.73
rep cluster 433	0	2	0.27	1.73
rep cluster 435	2	20	2.96	19.04
rep cluster 439	0	27	3.64	23.36
rep cluster 440	0	10	1.35	8.65
rep cluster 441	0	5	0.67	4.33
rep cluster 443	4	3	0.94	6.06
rep cluster 446	2	2	0.54	3.46
rep cluster 453	0	1	0.13	0.87
rep cluster 455	0	3	0.4	2.6
rep cluster 457	0	4	0.54	3.46
rep cluster 459	0	51	6.87	44.13
rep cluster 46	0	1	0.13	0.87
rep cluster 460	0	1	0.13	0.87
rep cluster 461	5	42	6.33	40.67
rep cluster 463	0	1	0.13	0.87
rep cluster 465	0	1	0.13	0.87
rep cluster 466	0	3	0.4	2.6
rep cluster 471	0	33	4.45	28.55
rep cluster 473	0	7	0.94	6.06
rep cluster 475	28	78	14.28	91.72
rep cluster 477	0	1	0.13	0.87
rep cluster 478	0	17	2.29	14.71
rep cluster 480	0	1	0.13	0.87
rep cluster 481	0	17	2.29	14.71
rep cluster 482	0	1	0.13	0.87
rep cluster 484	0	30	4.04	25.96
rep cluster 485	2	9	1.48	9.52
rep cluster 488	2	116	15.9	102.1
rep cluster 49	0	2	0.27	1.73
rep cluster 490	49	58	14.41	92.59
rep cluster 491	0	4	0.54	3.46
rep cluster 492	0	8	1.08	6.92
rep cluster 493	0	3	0.4	2.6
rep cluster 494	0	2	0.27	1.73
rep cluster 496	0	3	0.4	2.6
rep cluster 497	0	7	0.94	6.06
rep cluster 5	0	4	0.54	3.46
rep cluster 50	0	1	0.13	0.87
rep cluster 500	0	3	0.4	2.6
rep cluster 502	0	1	0.13	0.87
rep cluster 505	0	7	0.94	6.06
rep cluster 506	0	13	1.75	11.25
rep cluster 507	0	5	0.67	4.33
rep cluster 508	0	21	2.83	18.17
rep cluster 509	0	1	0.13	0.87
rep cluster 512	0	8	1.08	6.92
rep cluster 513	0	4	0.54	3.46
rep cluster 514	0	1	0.13	0.87
rep cluster 515	0	1	0.13	0.87
rep cluster 516	0	13	1.75	11.25
rep cluster 517	2	10	1.62	10.38
rep cluster 519	0	16	2.16	13.84
rep cluster 520	0	14	1.89	12.11
rep cluster 521	0	1	0.13	0.87

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 522	0	9	1.21	7.79
rep cluster 525	0	1	0.13	0.87
rep cluster 526	0	3	0.4	2.6
rep cluster 528	0	2	0.27	1.73
rep cluster 529	0	1	0.13	0.87
rep cluster 530	0	11	1.48	9.52
rep cluster 533	0	1	0.13	0.87
rep cluster 538	2	1	0.4	2.6
rep cluster 541	0	3	0.4	2.6
rep cluster 542	0	2	0.27	1.73
rep cluster 543	0	5	0.67	4.33
rep cluster 545	0	3	0.4	2.6
rep cluster 546	0	9	1.21	7.79
rep cluster 547	0	2	0.27	1.73
rep cluster 548	0	1	0.13	0.87
rep cluster 551	0	1	0.13	0.87
rep cluster 553	3	26	3.91	25.09
rep cluster 554	10	119	17.38	111.62
rep cluster 555	0	2	0.27	1.73
rep cluster 556	4	96	13.47	86.53
rep cluster 557	0	23	3.1	19.9
rep cluster 558	13	62	10.1	64.9
rep cluster 559	0	6	0.81	5.19
rep cluster 56	0	1	0.13	0.87
rep cluster 560	0	1	0.13	0.87
rep cluster 562	0	1	0.13	0.87
rep cluster 564	0	7	0.94	6.06
rep cluster 565	0	8	1.08	6.92
rep cluster 567	0	6	0.81	5.19
rep cluster 570	14	36	6.74	43.26
rep cluster 571	0	5	0.67	4.33
rep cluster 573	4	20	3.23	20.77
rep cluster 574	0	28	3.77	24.23
rep cluster 577	0	2	0.27	1.73
rep cluster 578	9	23	4.31	27.69
rep cluster 579	0	69	9.29	59.71
rep cluster 580	0	1	0.13	0.87
rep cluster 581	0	1	0.13	0.87
rep cluster 583	0	1	0.13	0.87
rep cluster 586	0	1	0.13	0.87
rep cluster 589	2	94	12.93	83.07
rep cluster 590	0	8	1.08	6.92
rep cluster 591	0	4	0.54	3.46
rep cluster 592	3	14	2.29	14.71
rep cluster 593	0	13	1.75	11.25
rep cluster 596	0	1	0.13	0.87
rep cluster 597	0	2	0.27	1.73
rep cluster 599	0	6	0.81	5.19
rep cluster 60	0	7	0.94	6.06
rep cluster 600	0	4	0.54	3.46
rep cluster 601	0	1	0.13	0.87
rep cluster 602	0	6	0.81	5.19
rep cluster 604	0	1	0.13	0.87
rep cluster 605	0	7	0.94	6.06
rep cluster 607	0	1	0.13	0.87

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 608	0	2	0.27	1.73
rep cluster 61	0	1	0.13	0.87
rep cluster 610	0	15	2.02	12.98
rep cluster 611	0	1	0.13	0.87
rep cluster 613	0	9	1.21	7.79
rep cluster 617	17	39	7.54	48.46
rep cluster 618	0	23	3.1	19.9
rep cluster 619	0	9	1.21	7.79
rep cluster 622	0	7	0.94	6.06
rep cluster 625	0	1	0.13	0.87
rep cluster 626	0	4	0.54	3.46
rep cluster 628	0	5	0.67	4.33
rep cluster 633	3	11	1.89	12.11
rep cluster 634	0	2	0.27	1.73
rep cluster 635	0	2	0.27	1.73
rep cluster 636	0	18	2.42	15.58
rep cluster 637	0	2	0.27	1.73
rep cluster 639	8	58	8.89	57.11
rep cluster 64	0	10	1.35	8.65
rep cluster 640	0	3	0.4	2.6
rep cluster 641	0	3	0.4	2.6
rep cluster 645	0	79	10.64	68.36
rep cluster 647	0	16	2.16	13.84
rep cluster 648	0	2	0.27	1.73
rep cluster 650	0	23	3.1	19.9
rep cluster 655	0	2	0.27	1.73
rep cluster 656	0	14	1.89	12.11
rep cluster 657	0	1	0.13	0.87
rep cluster 658	4	18	2.96	19.04
rep cluster 660	0	7	0.94	6.06
rep cluster 661	0	2	0.27	1.73
rep cluster 662	0	1	0.13	0.87
rep cluster 663	9	20	3.91	25.09
rep cluster 664	0	10	1.35	8.65
rep cluster 667	6	47	7.14	45.86
rep cluster 669	0	2	0.27	1.73
rep cluster 67	0	1	0.13	0.87
rep cluster 672	0	1	0.13	0.87
rep cluster 673	0	3	0.4	2.6
rep cluster 674	0	3	0.4	2.6
rep cluster 675	0	1	0.13	0.87
rep cluster 678	0	3	0.4	2.6
rep cluster 680	0	2	0.27	1.73
rep cluster 681	0	5	0.67	4.33
rep cluster 684	0	2	0.27	1.73
rep cluster 685	0	4	0.54	3.46
rep cluster 687	0	2	0.27	1.73
rep cluster 688	0	19	2.56	16.44
rep cluster 689	3	4	0.94	6.06
rep cluster 694	0	1	0.13	0.87
rep cluster 698	0	1	0.13	0.87
rep cluster 703	0	1	0.13	0.87
rep cluster 704	0	6	0.81	5.19
rep cluster 707	2	307	41.62	267.38
rep cluster 71	0	11	1.48	9.52

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 711	0	5	0.67	4.33
rep cluster 715	0	17	2.29	14.71
rep cluster 718	0	105	14.14	90.86
rep cluster 719	0	1	0.13	0.87
rep cluster 721	0	1	0.13	0.87
rep cluster 722	0	1	0.13	0.87
rep cluster 723	0	3	0.4	2.6
rep cluster 724	0	7	0.94	6.06
rep cluster 725	0	2	0.27	1.73
rep cluster 728	0	5	0.67	4.33
rep cluster 731	0	168	22.63	145.37
rep cluster 734	18	157	23.57	151.43
rep cluster 735	0	63	8.49	54.51
rep cluster 736	0	16	2.16	13.84
rep cluster 737	0	39	5.25	33.75
rep cluster 741	0	1	0.13	0.87
rep cluster 746	0	6	0.81	5.19
rep cluster 748	0	2	0.27	1.73
rep cluster 749	0	1	0.13	0.87
rep cluster 750	0	1	0.13	0.87
rep cluster 754	0	8	1.08	6.92
rep cluster 755	0	3	0.4	2.6
rep cluster 757	0	14	1.89	12.11
rep cluster 759	0	5	0.67	4.33
rep cluster 762	0	8	1.08	6.92
rep cluster 763	0	10	1.35	8.65
rep cluster 765	0	59	7.95	51.05
rep cluster 766	0	2	0.27	1.73
rep cluster 768	0	7	0.94	6.06
rep cluster 77	0	2	0.27	1.73
rep cluster 771	0	9	1.21	7.79
rep cluster 776	0	6	0.81	5.19
rep cluster 778	0	40	5.39	34.61
rep cluster 779	0	2	0.27	1.73
rep cluster 780	0	1	0.13	0.87
rep cluster 782	0	1	0.13	0.87
rep cluster 783	0	1	0.13	0.87
rep cluster 784	0	4	0.54	3.46
rep cluster 787	0	3	0.4	2.6
rep cluster 792	0	5	0.67	4.33
rep cluster 793	0	6	0.81	5.19
rep cluster 795	6	5	1.48	9.52
rep cluster 796	0	8	1.08	6.92
rep cluster 799	0	20	2.69	17.31
rep cluster 80	0	2	0.27	1.73
rep cluster 800	0	5	0.67	4.33
rep cluster 802	2	0	0.27	1.73
rep cluster 803	0	9	1.21	7.79
rep cluster 806	2	5	0.94	6.06
rep cluster 807	0	36	4.85	31.15
rep cluster 81	0	3	0.4	2.6
rep cluster 814	0	8	1.08	6.92
rep cluster 816	0	6	0.81	5.19
rep cluster 822	0	1	0.13	0.87
rep cluster 824	0	4	0.54	3.46

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 827	0	11	1.48	9.52
rep cluster 83	0	1	0.13	0.87
rep cluster 831	0	3	0.4	2.6
rep cluster 832	0	1	0.13	0.87
rep cluster 833	0	1	0.13	0.87
rep cluster 835	0	1	0.13	0.87
rep cluster 837	0	1	0.13	0.87
rep cluster 84	0	1	0.13	0.87
rep cluster 840	0	1	0.13	0.87
rep cluster 841	2	74	10.24	65.76
rep cluster 842	0	2	0.27	1.73
rep cluster 843	0	2	0.27	1.73
rep cluster 845	0	5	0.67	4.33
rep cluster 848	0	10	1.35	8.65
rep cluster 849	0	1	0.13	0.87
rep cluster 850	0	13	1.75	11.25
rep cluster 853	0	1	0.13	0.87
rep cluster 855	0	2	0.27	1.73
rep cluster 856	0	1	0.13	0.87
rep cluster 858	2	7	1.21	7.79
rep cluster 862	0	1	0.13	0.87
rep cluster 865	0	3	0.4	2.6
rep cluster 866	0	1	0.13	0.87
rep cluster 867	0	8	1.08	6.92
rep cluster 870	12	12	3.23	20.77
rep cluster 871	0	2	0.27	1.73
rep cluster 873	0	3	0.4	2.6
rep cluster 881	0	2	0.27	1.73
rep cluster 887	0	2	0.27	1.73
rep cluster 889	0	283	38.12	244.88
rep cluster 891	0	1	0.13	0.87
rep cluster 892	0	5	0.67	4.33
rep cluster 893	0	320	43.11	276.89
rep cluster 896	0	8	1.08	6.92
rep cluster 899	0	6	0.81	5.19
rep cluster 901	0	1	0.13	0.87
rep cluster 902	0	1	0.13	0.87
rep cluster 904	0	3	0.4	2.6
rep cluster 905	0	4	0.54	3.46
rep cluster 906	0	3	0.4	2.6
rep cluster 909	0	3	0.4	2.6
rep cluster 91	0	1	0.13	0.87
rep cluster 910	0	54	7.27	46.73
rep cluster 911	0	1	0.13	0.87
rep cluster 914	0	11	1.48	9.52
rep cluster 916	0	4	0.54	3.46
rep cluster 918	0	1	0.13	0.87
rep cluster 919	0	1	0.13	0.87
rep cluster 92	0	1	0.13	0.87
rep cluster 920	0	36	4.85	31.15
rep cluster 922	0	1	0.13	0.87
rep cluster 923	0	1	0.13	0.87
rep cluster 93	0	1	0.13	0.87
rep cluster 935	0	14	1.89	12.11
rep cluster 936	0	8	1.08	6.92

Replicon type	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 943	2	106	14.55	93.45
rep cluster 946	0	6	0.81	5.19
rep cluster 949	0	2	0.27	1.73
rep cluster 950	0	3	0.4	2.6
rep cluster 957	0	3	0.4	2.6
rep cluster 958	0	9	1.21	7.79
rep cluster 96	0	2	0.27	1.73
rep cluster 961	0	3	0.4	2.6
rep cluster 962	0	1	0.13	0.87
rep cluster 966	0	2	0.27	1.73
rep cluster 969	0	4	0.54	3.46
rep cluster 971	0	6	0.81	5.19
rep cluster 974	0	3	0.4	2.6
rep cluster 976	0	75	10.1	64.9
rep cluster 977	0	1	0.13	0.87
rep cluster 980	0	177	23.84	153.16
rep cluster 982	2	3	0.67	4.33
rep cluster 986	0	1	0.13	0.87
rep cluster 991	0	15	2.02	12.98
rep cluster 992	5	167	23.17	148.83
rep cluster 993	0	1	0.13	0.87
rep cluster 994	0	4	0.54	3.46
rep cluster 995	0	20	2.69	17.31
rep cluster 996	0	2	0.27	1.73
rep cluster 998	0	11	1.48	9.52

Table 6: Replicon types as detected by MOBtyper and the number of plasmids with said replicon type that were observed to be BHR on the level of species, observed to be related to a single host, and the number of plasmids that were expected to be BHR and SH under the assumption that host range classification is independent of replicon type.

Relaxase type	Observed BHR	Observed SH	Expected BHR	Expected SH
No relaxase type	1577	18959	3099.61	17436.39
MOBB	14	43	8.6	48.4
MOBC	468	803	191.84	1079.16
MOBF	1217	9877	1674.48	9419.52
MOBH	723	2898	546.54	3074.46
MOBM	0	60	9.06	50.94
MOBP	3902	11480	2321.69	13060.31
MOBQ	517	3216	563.44	3169.56
MOBT	9	179	28.38	159.62
MOBV	321	1604	290.55	1634.45
MOB UNKNOWN	7	131	20.83	117.17

Table 7: Relaxase types as detected by MOBtyper and the number of plasmids with said relaxase type that were observed to be BHR on the level of species, observed to be related to a single host, and the number of plasmids that were expected to be BHR and SH under the assumption that host range classification is independent of relaxase type.

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
No replicon type	9957	468	8837.6	1587.4
Col(BS512)	1	0	0.85	0.15
Col(MG828)	86	24	93.25	16.75
Col(MG828),IncK2/Z	2	0	1.7	0.3
Col(MG828),IncX3	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
Col(MG828),rep cluster 2350	2	0	1.7	0.3
Col(MG828),rep cluster 2373	1	0	0.85	0.15
Col(MG828),rep cluster 2392	3	0	2.54	0.46
Col(MP18)	7	0	5.93	1.07
Col(MP18),rep cluster 2350	1	0	0.85	0.15
Col(SD853)	72	9	68.67	12.33
Col(SD853),Inc13	1	0	0.85	0.15
Col(SD853),rep cluster 1017	2	0	1.7	0.3
Col(VCM04)	67	7	62.73	11.27
Col(VCM04),IncC	1	0	0.85	0.15
Col(VCM04),IncFII	1	0	0.85	0.15
Col(VCM04),IncP	1	0	0.85	0.15
Col(VCM04),IncQ1	1	0	0.85	0.15
Col(VCM04),IncR,rep cluster 1418	1	0	0.85	0.15
Col(YF27601)	1	0	0.85	0.15
Col(Ye4449)	33	8	34.76	6.24
Col156	113	92	173.78	31.22
Col156,IncFIA,IncFIB,IncFIC,IncFII	1	0	0.85	0.15
Col156,IncFIA,IncFIB,IncFIC,rep cluster 2232,rep cluster 2313	1	0	0.85	0.15
Col156,IncFIA,IncFIB,IncFII	2	0	1.7	0.3
Col156,IncFIA,IncFIC,rep cluster 2232,rep cluster 2313	1	0	0.85	0.15
Col156,IncFIA,IncFII,rep cluster 2131	1	0	0.85	0.15
Col156,IncFIB,rep cluster 2313	1	0	0.85	0.15
Col156,rep cluster 2131	5	4	7.63	1.37
Col156,rep cluster 2361	0	1	0.85	0.15
Col3M	16	2	15.26	2.74
Col3M,IncC,rep cluster 1220	1	0	0.85	0.15
Col3M,rep cluster 1220	2	0	1.7	0.3
Col3M,rep cluster 1254	1	0	0.85	0.15
Col8282,IncFIA,IncFIC,rep cluster 1778	1	0	0.85	0.15
Col8282,rep cluster 1778	1	1	1.7	0.3
ColE10	24	9	27.98	5.02
ColE10,IncX1	1	0	0.85	0.15
ColE10,rep cluster 2350	2	0	1.7	0.3
ColKP3	2	1	2.54	0.46
ColRNAI rep cluster 1857	185	398	494.23	88.77
ColRNAI rep cluster 1857,ColRNAI rep cluster 1987,IncFIB,IncHI1B,rep cluster 1195	1	0	0.85	0.15
ColRNAI rep cluster 1857,ColRNAI rep cluster 1987,IncFIB,rep cluster 1195	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncFIA	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncFIA,IncFIB	4	0	3.39	0.61
ColRNAI rep cluster 1857,IncFIA,IncFIB,IncFII	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncFIA,IncFIC,IncFII	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncFIA,IncFII,IncR,IncU	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncFIA,IncFII,IncR,rep cluster 1254,rep cluster 2183	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncFIA,IncFII,rep cluster 2183	2	0	1.7	0.3
ColRNAI rep cluster 1857,IncFIA,IncP,IncR,IncU	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncFIA,IncU	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
ColRNAI rep cluster 1857,IncFIA,IncU,rep cluster 1418	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncFIA,IncU,rep cluster 2327	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncFIB	2	0	1.7	0.3
ColRNAI rep cluster 1857,IncFIB,IncFII	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncFIB,IncFII,rep cluster 2183	4	0	3.39	0.61
ColRNAI rep cluster 1857,IncFIB,IncHI1B	3	0	2.54	0.46
ColRNAI rep cluster 1857,IncFIC,IncR,IncU,rep cluster 2183	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncFII	5	0	4.24	0.76
ColRNAI rep cluster 1857,IncFII,IncU	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncFII,rep cluster 1418	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncK2/Z	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncN	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncN,IncU	0	2	1.7	0.3
ColRNAI rep cluster 1857,IncR	5	0	4.24	0.76
ColRNAI rep cluster 1857,IncR,IncU	2	0	1.7	0.3
ColRNAI rep cluster 1857,IncU	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncU,rep cluster 1418	1	0	0.85	0.15
ColRNAI rep cluster 1857,IncX3	2	0	1.7	0.3
ColRNAI rep cluster 1857,rep cluster 1155	1	0	0.85	0.15
ColRNAI rep cluster 1857,rep cluster 1232	1	0	0.85	0.15
ColRNAI rep cluster 1857,rep cluster 1367	1	0	0.85	0.15
ColRNAI rep cluster 1857,rep cluster 2324	1	0	0.85	0.15
ColRNAI rep cluster 1857,rep cluster 2350	7	7	11.87	2.13
ColRNAI rep cluster 1857,rep cluster 2358	0	1	0.85	0.15
ColRNAI rep cluster 1857,rep cluster 2370	1	0	0.85	0.15
ColRNAI rep cluster 1857,rep cluster 2373	1	0	0.85	0.15
ColRNAI rep cluster 1857,rep cluster 2401	1	0	0.85	0.15
ColRNAI rep cluster 1987	759	463	1035.93	186.07
ColRNAI rep cluster 1987,ColpVC	2	0	1.7	0.3
ColRNAI rep cluster 1987,IncC,rep cluster 1254	2	0	1.7	0.3
ColRNAI rep cluster 1987,IncFIA,IncFIB	2	0	1.7	0.3
ColRNAI rep cluster 1987,IncFIB	2	0	1.7	0.3
ColRNAI rep cluster 1987,IncFIB,IncFII	1	0	0.85	0.15
ColRNAI rep cluster 1987,IncFIB,IncFII,IncQ1,rep cluster 2183	1	0	0.85	0.15
ColRNAI rep cluster 1987,IncFIB,IncFII,rep cluster 2183	1	0	0.85	0.15
ColRNAI rep cluster 1987,IncFIB,IncHI1B,rep cluster 1195	1	0	0.85	0.15
ColRNAI rep cluster 1987,IncFIB,IncHI1B,rep cluster 1254	1	0	0.85	0.15
ColRNAI rep cluster 1987,IncFIB,IncX3	1	0	0.85	0.15
ColRNAI rep cluster 1987,IncFII	3	0	2.54	0.46
ColRNAI rep cluster 1987,IncFII,rep cluster 2373	1	0	0.85	0.15
ColRNAI rep cluster 1987,IncI-gamma/K1	1	0	0.85	0.15
ColRNAI rep cluster 1987,IncN,IncR,IncX1,rep cluster 1195	1	0	0.85	0.15
ColRNAI rep cluster 1987,IncN,IncX1	1	0	0.85	0.15
ColRNAI rep cluster 1987,IncP,IncU	2	0	1.7	0.3
ColRNAI rep cluster 1987,IncQ1	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
ColRNAI rep cluster 1987,IncR	3	0	2.54	0.46
ColRNAI rep cluster 1987,IncR,rep cluster 1195	1	0	0.85	0.15
ColRNAI rep cluster 1987,rep cluster 1155,rep cluster 2374	1	0	0.85	0.15
ColRNAI rep cluster 1987,rep cluster 2327	2	0	1.7	0.3
ColRNAI rep cluster 1987,rep cluster 2327,rep cluster 2335	1	0	0.85	0.15
ColRNAI rep cluster 1987,rep cluster 2335	6	1	5.93	1.07
ColRNAI rep cluster 1987,rep cluster 2335,rep cluster 2358	3	0	2.54	0.46
ColRNAI rep cluster 1987,rep cluster 2350	0	1	0.85	0.15
ColRNAI rep cluster 1987,rep cluster 2358	23	21	37.3	6.7
ColRNAI rep cluster 1987,rep cluster 2370	1	0	0.85	0.15
ColRNAI rep cluster 1987,rep cluster 2373	8	3	9.33	1.67
ColRNAI rep cluster 1987,rep cluster 2374	10	1	9.33	1.67
ColRNAI rep cluster 1987,rep cluster 2405	0	4	3.39	0.61
ColRNAI rep cluster 1987,rep cluster 2406	2	3	4.24	0.76
ColRNAI rep cluster 1998	2	0	1.7	0.3
ColpVC	21	3	20.35	3.65
ColpVC,IncFIA,IncFII,IncN,IncU	1	0	0.85	0.15
Inc11	43	5	40.69	7.31
Inc11,rep cluster 1118,rep cluster 185	4	0	3.39	0.61
Inc11,rep cluster 1118,rep cluster 185,rep cluster 589	2	0	1.7	0.3
Inc11,rep cluster 1118,rep cluster 992	1	0	0.85	0.15
Inc11,rep cluster 1378,rep cluster 556	2	0	1.7	0.3
Inc11,rep cluster 2350	2	0	1.7	0.3
Inc11,rep cluster 342	3	0	2.54	0.46
Inc11,rep cluster 342,rep cluster 554,rep cluster 799	1	0	0.85	0.15
Inc11,rep cluster 414,rep cluster 554,rep cluster 556	1	0	0.85	0.15
Inc11,rep cluster 520,rep cluster 799	1	0	0.85	0.15
Inc11,rep cluster 556	1	0	0.85	0.15
Inc11,rep cluster 556,rep cluster 570	1	0	0.85	0.15
Inc11,rep cluster 589	1	0	0.85	0.15
Inc13	49	0	41.54	7.46
Inc18	60	0	50.86	9.14
Inc18,rep cluster 1017,rep cluster 1018,rep cluster 185	2	0	1.7	0.3
Inc18,rep cluster 1027	1	0	0.85	0.15
Inc18,rep cluster 1118	2	0	1.7	0.3
Inc18,rep cluster 1118,rep cluster 185	14	0	11.87	2.13
Inc18,rep cluster 1118,rep cluster 185,rep cluster 943	1	0	0.85	0.15
Inc18,rep cluster 1118,rep cluster 976	1	0	0.85	0.15
Inc18,rep cluster 1118,rep cluster 980	1	0	0.85	0.15
Inc18,rep cluster 1294,rep cluster 992	1	0	0.85	0.15
Inc18,rep cluster 180	2	0	1.7	0.3
Inc18,rep cluster 183	1	0	0.85	0.15
Inc18,rep cluster 185	9	0	7.63	1.37
Inc18,rep cluster 185,rep cluster 1968	3	0	2.54	0.46
Inc18,rep cluster 185,rep cluster 1968,rep cluster 943	1	0	0.85	0.15
Inc18,rep cluster 185,rep cluster 893	1	0	0.85	0.15
Inc18,rep cluster 1861,rep cluster 893	1	0	0.85	0.15
Inc18,rep cluster 2101	1	0	0.85	0.15
Inc18,rep cluster 2350	1	0	0.85	0.15
Inc18,rep cluster 554,rep cluster 893	4	0	3.39	0.61

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
Inc18,rep cluster 889,rep cluster 893	1	0	0.85	0.15
Inc18,rep cluster 893	32	0	27.13	4.87
Inc18,rep cluster 980	1	0	0.85	0.15
IncA	28	11	33.06	5.94
IncA,IncC	1	1	1.7	0.3
IncA,IncFII	1	0	0.85	0.15
IncA,IncP	3	0	2.54	0.46
IncA,IncX3	1	0	0.85	0.15
IncA,rep cluster 1254	1	0	0.85	0.15
IncA,rep cluster 1332,rep cluster 658	1	0	0.85	0.15
IncA,rep cluster 658	1	0	0.85	0.15
IncC	479	179	557.81	100.19
IncC,IncFIA	1	0	0.85	0.15
IncC,IncFIA,IncFIB	1	0	0.85	0.15
IncC,IncFIA,IncFIB,IncFIC,IncFII	1	0	0.85	0.15
IncC,IncFIA,IncFIB,IncFII,rep cluster 2131	1	0	0.85	0.15
IncC,IncFIA,IncFIC	3	0	2.54	0.46
IncC,IncFIA,IncFIC,IncFII,IncU	1	0	0.85	0.15
IncC,IncFIA,IncFIC,rep cluster 1254	3	0	2.54	0.46
IncC,IncFIA,IncFII	1	0	0.85	0.15
IncC,IncFIA,IncFII,IncR,IncU	1	0	0.85	0.15
IncC,IncFIA,IncFII,IncR,IncU,rep cluster 1254	1	0	0.85	0.15
IncC,IncFIA,IncFII,rep cluster 1254	2	0	1.7	0.3
IncC,IncFIA,IncFII,rep cluster 2183	2	0	1.7	0.3
IncC,IncFIA,IncR	1	0	0.85	0.15
IncC,IncFIA,IncR,IncU	2	0	1.7	0.3
IncC,IncFIA,IncR,rep cluster 1254	1	0	0.85	0.15
IncC,IncFIB	4	0	3.39	0.61
IncC,IncFIB,IncFIC,IncFII	7	0	5.93	1.07
IncC,IncFIB,IncFII	2	0	1.7	0.3
IncC,IncFIB,IncFII,IncX1	1	0	0.85	0.15
IncC,IncFIB,IncFII,rep cluster 2183	6	0	5.09	0.91
IncC,IncFIB,IncHI1B	2	0	1.7	0.3
IncC,IncFIB,IncU	1	0	0.85	0.15
IncC,IncFIB,IncX3	1	0	0.85	0.15
IncC,IncFIB,rep cluster 1254	1	0	0.85	0.15
IncC,IncFIB,rep cluster 995	1	0	0.85	0.15
IncC,IncFIC	1	0	0.85	0.15
IncC,IncFII	4	0	3.39	0.61
IncC,IncFII,IncX1	4	0	3.39	0.61
IncC,IncFII,rep cluster 1418,rep cluster 2183	1	0	0.85	0.15
IncC,IncFII,rep cluster 2183	1	0	0.85	0.15
IncC,IncHI1B	1	0	0.85	0.15
IncC,IncHI1B,IncR	1	0	0.85	0.15
IncC,IncHI2A,IncN,rep cluster 1088	1	0	0.85	0.15
IncC,IncHI2A,rep cluster 1088	3	0	2.54	0.46
IncC,IncI-gamma/K1	2	0	1.7	0.3
IncC,IncN	6	2	6.78	1.22
IncC,IncP	2	0	1.7	0.3
IncC,IncQ1	3	4	5.93	1.07
IncC,IncQ1,IncU	4	0	3.39	0.61
IncC,IncQ1,rep cluster 1418	1	0	0.85	0.15
IncC,IncR	8	0	6.78	1.22
IncC,IncR,IncU	2	0	1.7	0.3

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
IncC,IncR,rep cluster 1254	2	0	1.7	0.3
IncC,IncT	1	0	0.85	0.15
IncC,IncU,rep cluster 1254	1	0	0.85	0.15
IncC,IncX1	1	0	0.85	0.15
IncC,IncX3	1	0	0.85	0.15
IncC,IncY	1	0	0.85	0.15
IncC,rep cluster 1254	51	18	58.49	10.51
IncC,rep cluster 1254,rep cluster 1506	1	0	0.85	0.15
IncC,rep cluster 1254,rep cluster 2175	1	0	0.85	0.15
IncC,rep cluster 1367	1	0	0.85	0.15
IncC,rep cluster 1506	2	0	1.7	0.3
IncC,rep cluster 1527	2	2	3.39	0.61
IncC,rep cluster 2238	0	2	1.7	0.3
IncC,rep cluster 2294	1	0	0.85	0.15
IncC,rep cluster 435	1	0	0.85	0.15
IncC,rep cluster 71	1	0	0.85	0.15
IncFIA	510	20	449.3	80.7
IncFIA,IncFIB	519	7	445.91	80.09
IncFIA,IncFIB,IncFIC	407	5	349.27	62.73
IncFIA,IncFIB,IncFIC,IncFII	138	0	116.99	21.01
IncFIA,IncFIB,IncFIC,IncFII,IncHI1B	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,IncFII,IncI-gamma/K1	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,IncFII,IncQ1	9	0	7.63	1.37
IncFIA,IncFIB,IncFIC,IncFII,IncQ1,rep cluster 1418,rep cluster 2183	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,IncFII,IncX3	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,IncFII,rep cluster 2131	5	1	5.09	0.91
IncFIA,IncFIB,IncFIC,IncFII,rep cluster 2232	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,IncFII,rep cluster 2244,rep cluster 2313	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,IncHI1A	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,IncHI2A,IncN,rep cluster 1088	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,IncI-gamma/K1	6	0	5.09	0.91
IncFIA,IncFIB,IncFIC,IncI1	2	0	1.7	0.3
IncFIA,IncFIB,IncFIC,IncN	11	0	9.33	1.67
IncFIA,IncFIB,IncFIC,IncN,rep cluster 312	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,IncQ1	32	0	27.13	4.87
IncFIA,IncFIB,IncFIC,IncQ1,rep cluster 2131	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,IncQ1,rep cluster 2232	2	0	1.7	0.3
IncFIA,IncFIB,IncFIC,IncQ1,rep cluster 2244	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,IncR,rep cluster 2244	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,IncX3	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,rep cluster 1195	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,rep cluster 1304	2	0	1.7	0.3
IncFIA,IncFIB,IncFIC,rep cluster 2131	79	16	80.53	14.47
IncFIA,IncFIB,IncFIC,rep cluster 2131,rep cluster 2232,rep cluster 2313	2	0	1.7	0.3
IncFIA,IncFIB,IncFIC,rep cluster 2232	8	0	6.78	1.22
IncFIA,IncFIB,IncFIC,rep cluster 2232,rep cluster 2313	3	0	2.54	0.46
IncFIA,IncFIB,IncFIC,rep cluster 2244	33	0	27.98	5.02
IncFIA,IncFIB,IncFIC,rep cluster 2244,rep cluster 2313	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
IncFIA,IncFIB,IncFIC,rep cluster 2244,rep cluster 2350	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,rep cluster 2313	5	0	4.24	0.76
IncFIA,IncFIB,IncFIC,rep cluster 2350,rep cluster 2370	1	0	0.85	0.15
IncFIA,IncFIB,IncFIC,rep cluster 488	1	0	0.85	0.15
IncFIA,IncFIB,IncFII	198	0	167.85	30.15
IncFIA,IncFIB,IncFII,IncHI1B	2	0	1.7	0.3
IncFIA,IncFIB,IncFII,IncHI1B,IncR,IncU	1	0	0.85	0.15
IncFIA,IncFIB,IncFII,IncHI1B,rep cluster 1254,rep cluster 2183	1	0	0.85	0.15
IncFIA,IncFIB,IncFII,IncHI1B,rep cluster 2183	1	0	0.85	0.15
IncFIA,IncFIB,IncFII,IncN	1	0	0.85	0.15
IncFIA,IncFIB,IncFII,IncQ1	38	7	38.15	6.85
IncFIA,IncFIB,IncFII,IncR,rep cluster 1418,rep cluster 2131	1	0	0.85	0.15
IncFIA,IncFIB,IncFII,IncX3	1	0	0.85	0.15
IncFIA,IncFIB,IncFII,IncX3,rep cluster 2183	1	0	0.85	0.15
IncFIA,IncFIB,IncFII,rep cluster 1195	2	0	1.7	0.3
IncFIA,IncFIB,IncFII,rep cluster 1418,rep cluster 2183	1	0	0.85	0.15
IncFIA,IncFIB,IncFII,rep cluster 2131	65	2	56.8	10.2
IncFIA,IncFIB,IncFII,rep cluster 2131,rep cluster 2232	1	0	0.85	0.15
IncFIA,IncFIB,IncFII,rep cluster 2131,rep cluster 2344	1	0	0.85	0.15
IncFIA,IncFIB,IncFII,rep cluster 2183	12	0	10.17	1.83
IncFIA,IncFIB,IncFII,rep cluster 2232	3	0	2.54	0.46
IncFIA,IncFIB,IncFII,rep cluster 2244	1	0	0.85	0.15
IncFIA,IncFIB,IncHI1B	1	0	0.85	0.15
IncFIA,IncFIB,IncHI1B,IncR	1	0	0.85	0.15
IncFIA,IncFIB,IncHI1B,IncR,IncU	1	0	0.85	0.15
IncFIA,IncFIB,IncHI1B,IncR,rep cluster 1254	2	0	1.7	0.3
IncFIA,IncFIB,IncHI1B,rep cluster 1254	1	0	0.85	0.15
IncFIA,IncFIB,IncHI2A,rep cluster 1088	3	0	2.54	0.46
IncFIA,IncFIB,IncN	9	1	8.48	1.52
IncFIA,IncFIB,IncN,IncX1	1	0	0.85	0.15
IncFIA,IncFIB,IncN,rep cluster 2131	1	0	0.85	0.15
IncFIA,IncFIB,IncQ1	2	0	1.7	0.3
IncFIA,IncFIB,IncQ1,IncU,rep cluster 1418	1	0	0.85	0.15
IncFIA,IncFIB,IncQ1,rep cluster 2131	4	0	3.39	0.61
IncFIA,IncFIB,IncQ1,rep cluster 2232	1	0	0.85	0.15
IncFIA,IncFIB,IncR	3	0	2.54	0.46
IncFIA,IncFIB,IncR,IncU	1	0	0.85	0.15
IncFIA,IncFIB,IncR,IncU,rep cluster 1418	1	0	0.85	0.15
IncFIA,IncFIB,IncR,rep cluster 1418	1	0	0.85	0.15
IncFIA,IncFIB,IncU	4	0	3.39	0.61
IncFIA,IncFIB,IncX1	36	3	33.06	5.94
IncFIA,IncFIB,IncX1,IncX3	2	0	1.7	0.3
IncFIA,IncFIB,IncX1,rep cluster 2268	1	0	0.85	0.15
IncFIA,IncFIB,IncY	1	0	0.85	0.15
IncFIA,IncFIB,IncY,rep cluster 2131	1	0	0.85	0.15
IncFIA,IncFIB,rep cluster 1418	1	0	0.85	0.15
IncFIA,IncFIB,rep cluster 1760	5	0	4.24	0.76

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
IncFIA,IncFIB,rep cluster 2131	143	23	140.72	25.28
IncFIA,IncFIB,rep cluster 2131,rep cluster 2232,rep cluster 2313	1	0	0.85	0.15
IncFIA,IncFIB,rep cluster 2232	1	0	0.85	0.15
IncFIA,IncFIB,rep cluster 2244	23	0	19.5	3.5
IncFIA,IncFIB,rep cluster 2313	2	0	1.7	0.3
IncFIA,IncFIB,rep cluster 2350	1	0	0.85	0.15
IncFIA,IncFIC	589	148	624.78	112.22
IncFIA,IncFIC,IncFII	60	0	50.86	9.14
IncFIA,IncFIC,IncFII,IncN	1	0	0.85	0.15
IncFIA,IncFIC,IncFII,IncR,rep cluster 2183	1	0	0.85	0.15
IncFIA,IncFIC,IncFII,IncY	1	0	0.85	0.15
IncFIA,IncFIC,IncHI1B	3	0	2.54	0.46
IncFIA,IncFIC,IncHI2A,IncN,rep cluster 1088	1	0	0.85	0.15
IncFIA,IncFIC,IncHI2A,rep cluster 1088	4	0	3.39	0.61
IncFIA,IncFIC,IncI-gamma/K1	2	0	1.7	0.3
IncFIA,IncFIC,IncI1/B/O	1	0	0.85	0.15
IncFIA,IncFIC,IncN	9	0	7.63	1.37
IncFIA,IncFIC,IncN,IncR	1	0	0.85	0.15
IncFIA,IncFIC,IncQ1	2	0	1.7	0.3
IncFIA,IncFIC,IncQ1,IncU,rep cluster 1418,rep cluster 2183	1	0	0.85	0.15
IncFIA,IncFIC,IncR	17	0	14.41	2.59
IncFIA,IncFIC,IncX1	12	0	10.17	1.83
IncFIA,IncFIC,IncX4	1	0	0.85	0.15
IncFIA,IncFIC,IncY	2	0	1.7	0.3
IncFIA,IncFIC,rep cluster 1088	1	0	0.85	0.15
IncFIA,IncFIC,rep cluster 1195	1	0	0.85	0.15
IncFIA,IncFIC,rep cluster 1254	3	0	2.54	0.46
IncFIA,IncFIC,rep cluster 2232	1	0	0.85	0.15
IncFIA,IncFIC,rep cluster 2232,rep cluster 2313	1	0	0.85	0.15
IncFIA,IncFIC,rep cluster 2244	4	0	3.39	0.61
IncFIA,IncFIC,rep cluster 2313	3	0	2.54	0.46
IncFIA,IncFIC,rep cluster 2335,rep cluster 2370	1	0	0.85	0.15
IncFIA,IncFIC,rep cluster 2358,rep cluster 2370	1	0	0.85	0.15
IncFIA,IncFII	357	36	333.16	59.84
IncFIA,IncFII,IncHI1B	1	0	0.85	0.15
IncFIA,IncFII,IncHI1B,IncN	1	0	0.85	0.15
IncFIA,IncFII,IncHI1B,IncX1	1	0	0.85	0.15
IncFIA,IncFII,IncI1,IncR	1	0	0.85	0.15
IncFIA,IncFII,IncL/M	1	0	0.85	0.15
IncFIA,IncFII,IncN	21	2	19.5	3.5
IncFIA,IncFII,IncN,IncR	2	0	1.7	0.3
IncFIA,IncFII,IncN,IncR,IncU	7	0	5.93	1.07
IncFIA,IncFII,IncN,IncR,IncX1	5	0	4.24	0.76
IncFIA,IncFII,IncN,IncU	2	0	1.7	0.3
IncFIA,IncFII,IncN,IncU,rep cluster 1418	1	0	0.85	0.15
IncFIA,IncFII,IncN,IncX1	11	0	9.33	1.67
IncFIA,IncFII,IncN,rep cluster 2183	1	0	0.85	0.15
IncFIA,IncFII,IncQ1,rep cluster 1254,rep cluster 2183	1	0	0.85	0.15
IncFIA,IncFII,IncQ1,rep cluster 2183	2	0	1.7	0.3
IncFIA,IncFII,IncR	76	0	64.43	11.57
IncFIA,IncFII,IncR,IncU	138	7	122.92	22.08
IncFIA,IncFII,IncR,IncU,rep cluster 2370	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
IncFIA,IncFII,IncR,rep cluster 1	1	0	0.85	0.15
IncFIA,IncFII,IncR,rep cluster 1254,rep cluster 2183	6	0	5.09	0.91
IncFIA,IncFII,IncR,rep cluster 2183	3	0	2.54	0.46
IncFIA,IncFII,IncU	22	4	22.04	3.96
IncFIA,IncFII,IncU,rep cluster 1418	15	0	12.72	2.28
IncFIA,IncFII,IncU,rep cluster 1760	1	0	0.85	0.15
IncFIA,IncFII,IncX1	9	5	11.87	2.13
IncFIA,IncFII,rep cluster 1418	3	0	2.54	0.46
IncFIA,IncFII,rep cluster 2131	3	0	2.54	0.46
IncFIA,IncFII,rep cluster 2165	2	0	1.7	0.3
IncFIA,IncFII,rep cluster 2183	86	0	72.9	13.1
IncFIA,IncFII,rep cluster 2244	1	0	0.85	0.15
IncFIA,IncFII,rep cluster 2313	2	0	1.7	0.3
IncFIA,IncHI1B	4	0	3.39	0.61
IncFIA,IncHI1B,IncN	1	0	0.85	0.15
IncFIA,IncHI2A,IncN,IncR,rep cluster 1088,rep cluster 1254	1	0	0.85	0.15
IncFIA,IncHI2A,IncR,rep cluster 1088	1	0	0.85	0.15
IncFIA,IncI-gamma/K1	4	0	3.39	0.61
IncFIA,IncI1	2	0	1.7	0.3
IncFIA,IncL/M,IncR	2	0	1.7	0.3
IncFIA,IncL/M,IncR,rep cluster 1418	1	0	0.85	0.15
IncFIA,IncN	12	0	10.17	1.83
IncFIA,IncN,IncR	3	0	2.54	0.46
IncFIA,IncN,IncR,IncU	3	0	2.54	0.46
IncFIA,IncN,IncR,IncX1	1	0	0.85	0.15
IncFIA,IncN,IncX1	2	0	1.7	0.3
IncFIA,IncQ1	29	0	24.58	4.42
IncFIA,IncQ1,IncR	2	0	1.7	0.3
IncFIA,IncR	114	7	102.58	18.42
IncFIA,IncR,IncU	44	2	39	7
IncFIA,IncR,IncU,rep cluster 1	1	0	0.85	0.15
IncFIA,IncR,IncU,rep cluster 2370	1	0	0.85	0.15
IncFIA,IncR,rep cluster 1195	2	0	1.7	0.3
IncFIA,IncR,rep cluster 1254	2	0	1.7	0.3
IncFIA,IncR,rep cluster 1418	52	0	44.08	7.92
IncFIA,IncR,rep cluster 2127	1	0	0.85	0.15
IncFIA,IncR,rep cluster 2392	1	0	0.85	0.15
IncFIA,IncR,rep cluster 282	1	0	0.85	0.15
IncFIA,IncU	11	0	9.33	1.67
IncFIA,IncU,rep cluster 1418	16	0	13.56	2.44
IncFIA,IncX1	29	0	24.58	4.42
IncFIA,IncX3	2	0	1.7	0.3
IncFIA,IncY	4	0	3.39	0.61
IncFIA,rep cluster 1418	39	3	35.6	6.4
IncFIA,rep cluster 1527	1	0	0.85	0.15
IncFIA,rep cluster 1704	1	0	0.85	0.15
IncFIA,rep cluster 1798	1	0	0.85	0.15
IncFIA,rep cluster 2077,rep cluster 2203	1	0	0.85	0.15
IncFIA,rep cluster 2131	1	0	0.85	0.15
IncFIA,rep cluster 2183	5	0	4.24	0.76
IncFIA,rep cluster 2232	1	0	0.85	0.15
IncFIA,rep cluster 2268	1	0	0.85	0.15
IncFIA,rep cluster 2313	3	0	2.54	0.46

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
IncFIA,rep cluster 2327	2	0	1.7	0.3
IncFIA,rep cluster 2335	1	0	0.85	0.15
IncFIA,rep cluster 2344	1	0	0.85	0.15
IncFIA,rep cluster 488	2	0	1.7	0.3
IncFIB	1324	70	1181.74	212.26
IncFIB,IncFIC	9	0	7.63	1.37
IncFIB,IncFIC,IncFII	21	18	33.06	5.94
IncFIB,IncFIC,IncFII,IncI-gamma/K1	1	0	0.85	0.15
IncFIB,IncFIC,IncFII,IncI1	1	0	0.85	0.15
IncFIB,IncFIC,IncFII,rep cluster 2244	5	0	4.24	0.76
IncFIB,IncFIC,IncFII,rep cluster 2272	1	0	0.85	0.15
IncFIB,IncFIC,IncHI2A,IncN,rep cluster 1088,rep cluster 2244	1	0	0.85	0.15
IncFIB,IncFIC,IncHI2A,rep cluster 1088,rep cluster 2232	1	0	0.85	0.15
IncFIB,IncFIC,IncI-gamma/K1	4	0	3.39	0.61
IncFIB,IncFIC,IncI-gamma/K1,rep cluster 2244,rep cluster 2313	1	0	0.85	0.15
IncFIB,IncFIC,IncI-gamma/K1,rep cluster 2313	1	0	0.85	0.15
IncFIB,IncFIC,IncI1	12	0	10.17	1.83
IncFIB,IncFIC,IncK2/Z	3	0	2.54	0.46
IncFIB,IncFIC,IncN,rep cluster 1088,rep cluster 2244	3	0	2.54	0.46
IncFIB,IncFIC,IncN,rep cluster 2244	2	0	1.7	0.3
IncFIB,IncFIC,IncQ1	1	0	0.85	0.15
IncFIB,IncFIC,IncQ1,rep cluster 2244	21	0	17.8	3.2
IncFIB,IncFIC,IncR,rep cluster 2244	1	0	0.85	0.15
IncFIB,IncFIC,IncU,rep cluster 2183,rep cluster 2392	1	0	0.85	0.15
IncFIB,IncFIC,IncX4,rep cluster 2244	1	0	0.85	0.15
IncFIB,IncFIC,rep cluster 1195,rep cluster 2244,rep cluster 2358	1	0	0.85	0.15
IncFIB,IncFIC,rep cluster 2131,rep cluster 2244,rep cluster 2350,rep cluster 2370	1	0	0.85	0.15
IncFIB,IncFIC,rep cluster 2183	2	0	1.7	0.3
IncFIB,IncFIC,rep cluster 2203,rep cluster 2283	4	0	3.39	0.61
IncFIB,IncFIC,rep cluster 2232	7	0	5.93	1.07
IncFIB,IncFIC,rep cluster 2244	175	15	161.07	28.93
IncFIB,IncFIC,rep cluster 2244,rep cluster 2313	54	0	45.78	8.22
IncFIB,IncFIC,rep cluster 2371	1	0	0.85	0.15
IncFIB,IncFII	726	102	701.92	126.08
IncFIB,IncFII,IncHI1A,IncP	1	0	0.85	0.15
IncFIB,IncFII,IncHI1B	3	0	2.54	0.46
IncFIB,IncFII,IncHI1B,IncR,IncU	1	0	0.85	0.15
IncFIB,IncFII,IncHI1B,IncR,rep cluster 1254,rep cluster 2183	1	0	0.85	0.15
IncFIB,IncFII,IncHI1B,IncR,rep cluster 1418,rep cluster 2183	1	0	0.85	0.15
IncFIB,IncFII,IncHI1B,IncU,rep cluster 1418,rep cluster 2183	2	0	1.7	0.3
IncFIB,IncFII,IncHI1B,rep cluster 1254,rep cluster 2183	2	0	1.7	0.3
IncFIB,IncFII,IncHI1B,rep cluster 1418	1	0	0.85	0.15
IncFIB,IncFII,IncHI1B,rep cluster 1418,rep cluster 2183	2	0	1.7	0.3
IncFIB,IncFII,IncHI1B,rep cluster 2183	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
IncFIB,IncFII,IncHI2A,rep cluster 1088	4	0	3.39	0.61
IncFIB,IncFII,IncHI2A,rep cluster 1088,rep cluster 2272	1	0	0.85	0.15
IncFIB,IncFII,IncI-gamma/K1	22	0	18.65	3.35
IncFIB,IncFII,IncI1	16	0	13.56	2.44
IncFIB,IncFII,IncI1/B/O,IncQ1,rep cluster 2131	1	0	0.85	0.15
IncFIB,IncFII,IncK2/Z	1	0	0.85	0.15
IncFIB,IncFII,IncN	1	0	0.85	0.15
IncFIB,IncFII,IncN,rep cluster 2183	3	0	2.54	0.46
IncFIB,IncFII,IncQ1	34	0	28.82	5.18
IncFIB,IncFII,IncQ1,rep cluster 2131	11	3	11.87	2.13
IncFIB,IncFII,IncQ1,rep cluster 2183	12	0	10.17	1.83
IncFIB,IncFII,IncR	4	0	3.39	0.61
IncFIB,IncFII,IncR,IncU,rep cluster 2183	1	0	0.85	0.15
IncFIB,IncFII,IncR,rep cluster 1,rep cluster 1254,rep cluster 2183	1	0	0.85	0.15
IncFIB,IncFII,IncR,rep cluster 1418,rep cluster 2183	4	0	3.39	0.61
IncFIB,IncFII,IncR,rep cluster 2183	11	0	9.33	1.67
IncFIB,IncFII,IncU	3	0	2.54	0.46
IncFIB,IncFII,IncU,rep cluster 1254	1	0	0.85	0.15
IncFIB,IncFII,IncX1	5	0	4.24	0.76
IncFIB,IncFII,IncX1,rep cluster 2272	1	0	0.85	0.15
IncFIB,IncFII,IncX3	3	0	2.54	0.46
IncFIB,IncFII,IncX3,rep cluster 1195	1	0	0.85	0.15
IncFIB,IncFII,IncX3,rep cluster 1195,rep cluster 2183	1	0	0.85	0.15
IncFIB,IncFII,IncX3,rep cluster 2183	2	0	1.7	0.3
IncFIB,IncFII,rep cluster 1,rep cluster 2183	1	0	0.85	0.15
IncFIB,IncFII,rep cluster 1254	1	0	0.85	0.15
IncFIB,IncFII,rep cluster 1254,rep cluster 2183	4	0	3.39	0.61
IncFIB,IncFII,rep cluster 1418	5	0	4.24	0.76
IncFIB,IncFII,rep cluster 1418,rep cluster 2183	11	0	9.33	1.67
IncFIB,IncFII,rep cluster 1418,rep cluster 2272	2	0	1.7	0.3
IncFIB,IncFII,rep cluster 1418,rep cluster 2272,rep cluster 2370	1	0	0.85	0.15
IncFIB,IncFII,rep cluster 1804	1	0	0.85	0.15
IncFIB,IncFII,rep cluster 2131	34	0	28.82	5.18
IncFIB,IncFII,rep cluster 2131,rep cluster 2232	2	0	1.7	0.3
IncFIB,IncFII,rep cluster 2183	775	67	713.79	128.21
IncFIB,IncFII,rep cluster 2183,rep cluster 2272	0	1	0.85	0.15
IncFIB,IncFII,rep cluster 2183,rep cluster 2358	2	0	1.7	0.3
IncFIB,IncFII,rep cluster 2183,rep cluster 2373	1	0	0.85	0.15
IncFIB,IncFII,rep cluster 2232	11	0	9.33	1.67
IncFIB,IncFII,rep cluster 2244	7	0	5.93	1.07
IncFIB,IncFII,rep cluster 2268	1	0	0.85	0.15
IncFIB,IncFII,rep cluster 2272	66	32	83.08	14.92
IncFIB,IncFII,rep cluster 2350	1	0	0.85	0.15
IncFIB,IncFII,rep cluster 2358	4	0	3.39	0.61
IncFIB,IncFII,rep cluster 2371	1	0	0.85	0.15
IncFIB,IncFII,rep cluster 2392	29	7	30.52	5.48
IncFIB,IncHI1A	1	0	0.85	0.15
IncFIB,IncHI1A,rep cluster 1304	3	0	2.54	0.46
IncFIB,IncHI1B	367	120	412.85	74.15
IncFIB,IncHI1B,IncI-gamma/K1	1	4	4.24	0.76
IncFIB,IncHI1B,IncK2/Z	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
IncFIB,IncHI1B,IncN	10	0	8.48	1.52
IncFIB,IncHI1B,IncN,IncU	1	0	0.85	0.15
IncFIB,IncHI1B,IncQ1	11	0	9.33	1.67
IncFIB,IncHI1B,IncQ1,IncU	1	0	0.85	0.15
IncFIB,IncHI1B,IncR	4	0	3.39	0.61
IncFIB,IncHI1B,IncR,IncX1,rep cluster 1254	1	0	0.85	0.15
IncFIB,IncHI1B,IncR,rep cluster 1254	6	0	5.09	0.91
IncFIB,IncHI1B,IncU	21	2	19.5	3.5
IncFIB,IncHI1B,IncU,IncX3	1	0	0.85	0.15
IncFIB,IncHI1B,IncU,rep cluster 1254	5	0	4.24	0.76
IncFIB,IncHI1B,IncX3	1	0	0.85	0.15
IncFIB,IncHI1B,IncY	1	0	0.85	0.15
IncFIB,IncHI1B,IncY,rep cluster 2373	1	0	0.85	0.15
IncFIB,IncHI1B,rep cluster 1254	82	2	71.21	12.79
IncFIB,IncHI1B,rep cluster 1304	3	0	2.54	0.46
IncFIB,IncHI1B,rep cluster 2392	1	0	0.85	0.15
IncFIB,IncHI2A,rep cluster 1088	5	0	4.24	0.76
IncFIB,IncHI2A,rep cluster 1088,rep cluster 2244	2	0	1.7	0.3
IncFIB,IncI-gamma/K1	105	0	89.01	15.99
IncFIB,IncI-gamma/K1,IncQ1	1	0	0.85	0.15
IncFIB,IncI-gamma/K1,rep cluster 1088	1	0	0.85	0.15
IncFIB,IncI-gamma/K1,rep cluster 2371	2	0	1.7	0.3
IncFIB,IncI1	3	0	2.54	0.46
IncFIB,IncI1/B/O	1	0	0.85	0.15
IncFIB,IncI2	1	0	0.85	0.15
IncFIB,IncI2,rep cluster 2244	1	0	0.85	0.15
IncFIB,IncK2/Z	9	0	7.63	1.37
IncFIB,IncL/M	2	0	1.7	0.3
IncFIB,IncL/M,IncU	1	0	0.85	0.15
IncFIB,IncN	10	0	8.48	1.52
IncFIB,IncN,IncR,IncX1	1	0	0.85	0.15
IncFIB,IncN,IncR,IncX1,rep cluster 2244	1	0	0.85	0.15
IncFIB,IncN,IncU	1	0	0.85	0.15
IncFIB,IncN,IncX1	3	0	2.54	0.46
IncFIB,IncN,IncX1,rep cluster 2358	1	0	0.85	0.15
IncFIB,IncN,rep cluster 2244	1	0	0.85	0.15
IncFIB,IncQ1	5	0	4.24	0.76
IncFIB,IncQ1,IncY,rep cluster 312	1	0	0.85	0.15
IncFIB,IncQ1,rep cluster 2244	4	0	3.39	0.61
IncFIB,IncR	21	0	17.8	3.2
IncFIB,IncR,IncU	4	0	3.39	0.61
IncFIB,IncR,rep cluster 2183	1	0	0.85	0.15
IncFIB,IncR,rep cluster 2187	1	0	0.85	0.15
IncFIB,IncU	40	4	37.3	6.7
IncFIB,IncX1	8	0	6.78	1.22
IncFIB,IncX1,rep cluster 2244	2	0	1.7	0.3
IncFIB,IncX3	4	0	3.39	0.61
IncFIB,IncX3,rep cluster 1804	1	0	0.85	0.15
IncFIB,IncY	3	0	2.54	0.46
IncFIB,rep cluster 1	1	0	0.85	0.15
IncFIB,rep cluster 1088	1	0	0.85	0.15
IncFIB,rep cluster 1254	14	0	11.87	2.13
IncFIB,rep cluster 1254,rep cluster 1418	4	0	3.39	0.61
IncFIB,rep cluster 1418	3	0	2.54	0.46

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
IncFIB,rep cluster 1444	2	0	1.7	0.3
IncFIB,rep cluster 1760	1	0	0.85	0.15
IncFIB,rep cluster 1804	8	0	6.78	1.22
IncFIB,rep cluster 2131	2	0	1.7	0.3
IncFIB,rep cluster 2131,rep cluster 2313	5	0	4.24	0.76
IncFIB,rep cluster 2183	8	0	6.78	1.22
IncFIB,rep cluster 2183,rep cluster 995	1	0	0.85	0.15
IncFIB,rep cluster 2187	2	0	1.7	0.3
IncFIB,rep cluster 2232	8	0	6.78	1.22
IncFIB,rep cluster 2244	127	9	115.29	20.71
IncFIB,rep cluster 2313	1	0	0.85	0.15
IncFIB,rep cluster 2350	2	0	1.7	0.3
IncFIB,rep cluster 2358	2	0	1.7	0.3
IncFIB,rep cluster 2370	1	0	0.85	0.15
IncFIB,rep cluster 2392	1	0	0.85	0.15
IncFIB,rep cluster 312	2	0	1.7	0.3
IncFIB,rep cluster 488	1	0	0.85	0.15
IncFIB,rep cluster 505	1	0	0.85	0.15
IncFIB,rep cluster 574	28	0	23.74	4.26
IncFIC	15	0	12.72	2.28
IncFIC,IncFII	36	4	33.91	6.09
IncFIC,IncFII,IncI1	1	0	0.85	0.15
IncFIC,IncFII,rep cluster 2165	1	0	0.85	0.15
IncFIC,IncI-gamma/K1	1	0	0.85	0.15
IncFIC,IncI-gamma/K1,IncN,rep cluster 2244	1	0	0.85	0.15
IncFIC,IncI-gamma/K1,rep cluster 2313	1	0	0.85	0.15
IncFIC,IncI1	12	0	10.17	1.83
IncFIC,IncK2/Z	2	0	1.7	0.3
IncFIC,IncN	1	0	0.85	0.15
IncFIC,IncN,rep cluster 2183	2	0	1.7	0.3
IncFIC,IncQ1,IncU,rep cluster 1418,rep cluster 2183	3	4	5.93	1.07
IncFIC,IncR,IncU,rep cluster 2183	1	0	0.85	0.15
IncFIC,IncU,rep cluster 1418,rep cluster 2183	1	0	0.85	0.15
IncFIC,IncU,rep cluster 2183	1	0	0.85	0.15
IncFIC,rep cluster 2183	11	2	11.02	1.98
IncFIC,rep cluster 2203,rep cluster 2283	1	0	0.85	0.15
IncFIC,rep cluster 2232	1	0	0.85	0.15
IncFIC,rep cluster 2244	3	2	4.24	0.76
IncFIC,rep cluster 2244,rep cluster 2313	1	0	0.85	0.15
IncFIC,rep cluster 2272	1	0	0.85	0.15
IncFII	352	42	334.01	59.99
IncFII,IncHI1B,IncP	1	0	0.85	0.15
IncFII,IncHI1B,IncP,rep cluster 2183	2	0	1.7	0.3
IncFII,IncHI1B,IncP,rep cluster 2272	1	0	0.85	0.15
IncFII,IncHI2A	1	0	0.85	0.15
IncFII,IncHI2A,rep cluster 1088	1	0	0.85	0.15
IncFII,IncHI2A,rep cluster 1088,rep cluster 1418	1	0	0.85	0.15
IncFII,IncI-gamma/K1	21	0	17.8	3.2
IncFII,IncI-gamma/K1,IncX1	1	0	0.85	0.15
IncFII,IncI1	2	0	1.7	0.3
IncFII,IncK2/Z	2	0	1.7	0.3
IncFII,IncL/M,IncQ1,rep cluster 1418,rep cluster 2183	2	0	1.7	0.3
IncFII,IncN	3	0	2.54	0.46

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
IncFII,IncQ1	1	0	0.85	0.15
IncFII,IncQ1,IncR,rep cluster 1418,rep cluster 2183	3	0	2.54	0.46
IncFII,IncQ1,IncR,rep cluster 2183	3	0	2.54	0.46
IncFII,IncQ1,IncX3,rep cluster 1254,rep cluster 1418,rep cluster 2183	1	0	0.85	0.15
IncFII,IncQ1,rep cluster 1418	1	0	0.85	0.15
IncFII,IncQ1,rep cluster 1418,rep cluster 2183	26	9	29.67	5.33
IncFII,IncQ1,rep cluster 1418,rep cluster 2183,rep cluster 2350	1	0	0.85	0.15
IncFII,IncR	48	0	40.69	7.31
IncFII,IncR,IncU	3	0	2.54	0.46
IncFII,IncR,IncU,rep cluster 1254,rep cluster 2183	1	0	0.85	0.15
IncFII,IncR,IncU,rep cluster 1418,rep cluster 2183	2	0	1.7	0.3
IncFII,IncR,IncU,rep cluster 2183	3	8	9.33	1.67
IncFII,IncR,rep cluster 1254	1	0	0.85	0.15
IncFII,IncR,rep cluster 1254,rep cluster 2183	6	0	5.09	0.91
IncFII,IncR,rep cluster 1418	1	0	0.85	0.15
IncFII,IncR,rep cluster 1418,rep cluster 2183	3	0	2.54	0.46
IncFII,IncR,rep cluster 2183	47	4	43.23	7.77
IncFII,IncR,rep cluster 2183,rep cluster 2350	1	0	0.85	0.15
IncFII,IncR,rep cluster 2183,rep cluster 2358	1	0	0.85	0.15
IncFII,IncR,rep cluster 2183,rep cluster 2373	1	0	0.85	0.15
IncFII,IncR,rep cluster 2358	1	0	0.85	0.15
IncFII,IncU	21	0	17.8	3.2
IncFII,IncU,IncX6	1	0	0.85	0.15
IncFII,IncU,rep cluster 1254,rep cluster 1418,rep cluster 2183	1	0	0.85	0.15
IncFII,IncU,rep cluster 1254,rep cluster 2183	2	0	1.7	0.3
IncFII,IncU,rep cluster 1418	8	2	8.48	1.52
IncFII,IncU,rep cluster 1418,rep cluster 2183	9	5	11.87	2.13
IncFII,IncU,rep cluster 2183	2	0	1.7	0.3
IncFII,IncU,rep cluster 2370	1	0	0.85	0.15
IncFII,IncX1	21	0	17.8	3.2
IncFII,IncX1,rep cluster 1418,rep cluster 2183	1	0	0.85	0.15
IncFII,IncX3	2	0	1.7	0.3
IncFII,IncX3,rep cluster 2183	3	0	2.54	0.46
IncFII,IncX5	1	0	0.85	0.15
IncFII,rep cluster 1195	1	0	0.85	0.15
IncFII,rep cluster 1195,rep cluster 2183	1	0	0.85	0.15
IncFII,rep cluster 1254	1	0	0.85	0.15
IncFII,rep cluster 1254,rep cluster 1418,rep cluster 2183	1	0	0.85	0.15
IncFII,rep cluster 1254,rep cluster 1798,rep cluster 2183	1	0	0.85	0.15
IncFII,rep cluster 1254,rep cluster 2183	2	0	1.7	0.3
IncFII,rep cluster 1304	1	0	0.85	0.15
IncFII,rep cluster 1418	151	4	131.4	23.6
IncFII,rep cluster 1418,rep cluster 1704	1	0	0.85	0.15
IncFII,rep cluster 1418,rep cluster 2183	74	10	71.21	12.79
IncFII,rep cluster 1418,rep cluster 2183,rep cluster 2401	1	0	0.85	0.15
IncFII,rep cluster 1418,rep cluster 2272	1	0	0.85	0.15
IncFII,rep cluster 1418,rep cluster 2350	1	0	0.85	0.15
IncFII,rep cluster 2078	3	0	2.54	0.46

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
IncFII,rep cluster 2165	2	0	1.7	0.3
IncFII,rep cluster 2183	66	4	59.34	10.66
IncFII,rep cluster 2183,rep cluster 2373	1	0	0.85	0.15
IncFII,rep cluster 2183,rep cluster 658	2	0	1.7	0.3
IncFII,rep cluster 2203	14	0	11.87	2.13
IncFII,rep cluster 2203,rep cluster 645	7	0	5.93	1.07
IncFII,rep cluster 2218	2	0	1.7	0.3
IncFII,rep cluster 2232	16	0	13.56	2.44
IncFII,rep cluster 2268	9	0	7.63	1.37
IncFII,rep cluster 2272	17	0	14.41	2.59
IncFII,rep cluster 2307	1	0	0.85	0.15
IncFII,rep cluster 2335,rep cluster 2358	1	0	0.85	0.15
IncFII,rep cluster 2350	1	0	0.85	0.15
IncFII,rep cluster 2358	1	0	0.85	0.15
IncFII,rep cluster 312	1	0	0.85	0.15
IncFII,rep cluster 645	72	0	61.04	10.96
IncFII,rep cluster 658	2	0	1.7	0.3
IncFII,rep cluster 904	1	0	0.85	0.15
IncHI1A	39	19	49.17	8.83
IncHI1A,IncN	1	0	0.85	0.15
IncHI1A,IncP	5	3	6.78	1.22
IncHI1A,IncQ1	22	10	27.13	4.87
IncHI1A,IncQ1,IncX1,IncX3	1	0	0.85	0.15
IncHI1A,IncX1	2	0	1.7	0.3
IncHI1A,IncX3	1	0	0.85	0.15
IncHI1A,rep cluster 1254	2	0	1.7	0.3
IncHI1A,rep cluster 1760	1	0	0.85	0.15
IncHI1A,rep cluster 2327	0	1	0.85	0.15
IncHI1B	324	10	283.14	50.86
IncHI1B,IncN	1	0	0.85	0.15
IncHI1B,IncN,IncX1	1	0	0.85	0.15
IncHI1B,IncP	15	4	16.11	2.89
IncHI1B,IncP,rep cluster 1304	1	0	0.85	0.15
IncHI1B,IncQ1	5	2	5.93	1.07
IncHI1B,IncR	1	0	0.85	0.15
IncHI1B,IncX1	4	0	3.39	0.61
IncHI1B,IncX4	1	0	0.85	0.15
IncHI1B,IncY,rep cluster 1304	1	0	0.85	0.15
IncHI1B,rep cluster 1254	5	0	4.24	0.76
IncHI1B,rep cluster 2077	3	0	2.54	0.46
IncHI1B,rep cluster 2358	1	0	0.85	0.15
IncHI1B,rep cluster 2392	1	0	0.85	0.15
IncHI2A	4	0	3.39	0.61
IncHI2A,IncI-gamma/K1,IncQ1,rep cluster 1088	1	0	0.85	0.15
IncHI2A,IncI-gamma/K1,rep cluster 1088	1	0	0.85	0.15
IncHI2A,IncI2,rep cluster 1088	1	0	0.85	0.15
IncHI2A,IncL/M,rep cluster 1088	1	0	0.85	0.15
IncHI2A,IncN	1	0	0.85	0.15
IncHI2A,IncN,IncQ1,rep cluster 1088	9	0	7.63	1.37
IncHI2A,IncN,IncY,rep cluster 1088	1	0	0.85	0.15
IncHI2A,IncN,rep cluster 1088	58	29	73.75	13.25
IncHI2A,IncP,rep cluster 1088	12	0	10.17	1.83
IncHI2A,IncQ1	1	0	0.85	0.15
IncHI2A,IncQ1,rep cluster 1088	41	2	36.45	6.55

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
IncHI2A,IncR,rep cluster 1088	3	0	2.54	0.46
IncHI2A,IncR,rep cluster 1088,rep cluster 1254	2	0	1.7	0.3
IncHI2A,IncX1,rep cluster 1088	1	0	0.85	0.15
IncHI2A,IncY,rep cluster 1088	1	0	0.85	0.15
IncHI2A,rep cluster 1088	574	179	638.34	114.66
IncHI2A,rep cluster 1088,rep cluster 1254	15	2	14.41	2.59
IncHI2A,rep cluster 1088,rep cluster 1798	1	0	0.85	0.15
IncHI2A,rep cluster 1088,rep cluster 1804	1	0	0.85	0.15
IncHI2A,rep cluster 1088,rep cluster 2318	1	0	0.85	0.15
IncHI2A,rep cluster 1088,rep cluster 2335	0	1	0.85	0.15
IncHI2A,rep cluster 1088,rep cluster 2350	1	0	0.85	0.15
IncHI2A,rep cluster 1088,rep cluster 827	1	0	0.85	0.15
IncI-gamma/K1	773	194	819.76	147.24
IncI-gamma/K1,IncI1/B/O	1	0	0.85	0.15
IncI-gamma/K1,IncN	3	0	2.54	0.46
IncI-gamma/K1,IncN,IncX1	1	0	0.85	0.15
IncI-gamma/K1,IncQ1	15	0	12.72	2.28
IncI-gamma/K1,IncX1	3	0	2.54	0.46
IncI-gamma/K1,rep cluster 2350	3	0	2.54	0.46
IncI-gamma/K1,rep cluster 2371	2	0	1.7	0.3
IncI-gamma/K1,rep cluster 435	1	0	0.85	0.15
IncI1	46	2	40.69	7.31
IncI1,IncU	1	0	0.85	0.15
IncI1,rep cluster 1057	10	0	8.48	1.52
IncI1,rep cluster 2273	1	0	0.85	0.15
IncI1,rep cluster 998	1	0	0.85	0.15
IncI1/B/O	128	53	153.44	27.56
IncI1/B/O,IncN	1	0	0.85	0.15
IncI1/B/O,rep cluster 1254	1	0	0.85	0.15
IncI1/B/O,rep cluster 1527	1	0	0.85	0.15
IncI1/B/O,rep cluster 2371	1	1	1.7	0.3
IncI2	215	240	385.72	69.28
IncI2,IncN,IncR,IncX1	1	0	0.85	0.15
IncI2,IncN,IncX1	1	0	0.85	0.15
IncI2,IncX4	1	0	0.85	0.15
IncK2/Z	228	25	214.48	38.52
IncK2/Z,IncN	1	0	0.85	0.15
IncK2/Z,IncQ1	9	0	7.63	1.37
IncK2/Z,IncX1	1	0	0.85	0.15
IncK2/Z,rep cluster 1527	1	0	0.85	0.15
IncK2/Z,rep cluster 2131	1	0	0.85	0.15
IncL/M	176	373	465.4	83.6
IncL/M,IncN	1	0	0.85	0.15
IncL/M,IncN,IncR	1	0	0.85	0.15
IncL/M,IncP	1	0	0.85	0.15
IncL/M,IncR	3	0	2.54	0.46
IncL/M,IncR,rep cluster 1418	1	0	0.85	0.15
IncL/M,IncU	3	0	2.54	0.46
IncL/M,rep cluster 1254	8	15	19.5	3.5
IncL/M,rep cluster 1367	1	0	0.85	0.15
IncL/M,rep cluster 2335	1	0	0.85	0.15
IncN	321	201	442.52	79.48
IncN,IncP	1	0	0.85	0.15
IncN,IncQ1,rep cluster 1088	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
IncN,IncR	26	11	31.37	5.63
IncN,IncR,IncU,rep cluster 1418	2	0	1.7	0.3
IncN,IncR,IncX1	5	0	4.24	0.76
IncN,IncR,IncX1,IncY	1	0	0.85	0.15
IncN,IncU	9	4	11.02	1.98
IncN,IncU,IncX1	1	0	0.85	0.15
IncN,IncX1	14	0	11.87	2.13
IncN,IncX1,rep cluster 2370	1	0	0.85	0.15
IncN,IncX2	1	0	0.85	0.15
IncN,IncX3	1	0	0.85	0.15
IncN,IncY	2	0	1.7	0.3
IncN,rep cluster 1195	1	0	0.85	0.15
IncN,rep cluster 1254	1	0	0.85	0.15
IncN,rep cluster 1418	2	0	1.7	0.3
IncN,rep cluster 1506	1	0	0.85	0.15
IncN,rep cluster 2183,rep cluster 2327	1	0	0.85	0.15
IncN,rep cluster 2203	1	0	0.85	0.15
IncN,rep cluster 2335	1	0	0.85	0.15
IncN,rep cluster 2335,rep cluster 2373	1	0	0.85	0.15
IncN,rep cluster 2350	2	0	1.7	0.3
IncN,rep cluster 2392	1	0	0.85	0.15
IncN,rep cluster 435	2	0	1.7	0.3
IncP	242	34	233.97	42.03
IncP,IncQ2,rep cluster 573,rep cluster 995	1	0	0.85	0.15
IncP,IncR,IncU	1	0	0.85	0.15
IncP,IncU	15	45	50.86	9.14
IncP,IncU,rep cluster 1068,rep cluster 2037	1	0	0.85	0.15
IncP,IncU,rep cluster 1115	1	0	0.85	0.15
IncP,IncU,rep cluster 658	2	0	1.7	0.3
IncP,rep cluster 1088	1	0	0.85	0.15
IncP,rep cluster 1115	19	0	16.11	2.89
IncP,rep cluster 156	1	0	0.85	0.15
IncP,rep cluster 1875	1	0	0.85	0.15
IncP,rep cluster 2037	1	0	0.85	0.15
IncP,rep cluster 2350	3	0	2.54	0.46
IncP,rep cluster 2358	1	0	0.85	0.15
IncP,rep cluster 282,rep cluster 312	1	0	0.85	0.15
IncP,rep cluster 283	1	0	0.85	0.15
IncP,rep cluster 398	14	0	11.87	2.13
IncP,rep cluster 398,rep cluster 688	1	0	0.85	0.15
IncQ1	75	56	111.05	19.95
IncQ1,IncU	8	0	6.78	1.22
IncQ1,IncU,rep cluster 1254	2	0	1.7	0.3
IncQ1,IncU,rep cluster 1418	1	0	0.85	0.15
IncQ1,IncX1	2	0	1.7	0.3
IncQ1,IncY	10	0	8.48	1.52
IncQ1,rep cluster 1088	1	0	0.85	0.15
IncQ1,rep cluster 2131	1	0	0.85	0.15
IncQ1,rep cluster 2335	1	0	0.85	0.15
IncQ2	27	24	43.23	7.77
IncQ2,rep cluster 1115	0	5	4.24	0.76
IncR	415	38	384.02	68.98
IncR,IncU	17	2	16.11	2.89
IncR,IncU,rep cluster 2183	2	1	2.54	0.46

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
IncR,IncX1	24	0	20.35	3.65
IncR,IncX3	1	0	0.85	0.15
IncR,IncY	4	0	3.39	0.61
IncR,rep cluster 1195	3	0	2.54	0.46
IncR,rep cluster 1254	4	0	3.39	0.61
IncR,rep cluster 1254,rep cluster 1418	1	0	0.85	0.15
IncR,rep cluster 1304	2	0	1.7	0.3
IncR,rep cluster 1418	15	0	12.72	2.28
IncR,rep cluster 1418,rep cluster 2272	1	0	0.85	0.15
IncR,rep cluster 2244	1	0	0.85	0.15
IncR,rep cluster 2327	1	0	0.85	0.15
IncR,rep cluster 2358	1	0	0.85	0.15
IncR,rep cluster 2401	1	0	0.85	0.15
IncT	5	0	4.24	0.76
IncT,rep cluster 1254,rep cluster 1478	1	0	0.85	0.15
IncT,rep cluster 1478	1	0	0.85	0.15
IncU	109	16	105.97	19.03
IncU,IncX3	3	4	5.93	1.07
IncU,IncX5	4	0	3.39	0.61
IncU,IncX6	0	12	10.17	1.83
IncU,rep cluster 1115	8	0	6.78	1.22
IncU,rep cluster 1172	2	0	1.7	0.3
IncU,rep cluster 1254	20	4	20.35	3.65
IncU,rep cluster 1289	1	0	0.85	0.15
IncU,rep cluster 1418	2	0	1.7	0.3
IncU,rep cluster 1418,rep cluster 2183	10	16	22.04	3.96
IncU,rep cluster 1444	1	0	0.85	0.15
IncU,rep cluster 1648	1	0	0.85	0.15
IncU,rep cluster 1760	1	0	0.85	0.15
IncU,rep cluster 2025	1	0	0.85	0.15
IncU,rep cluster 2037	1	0	0.85	0.15
IncU,rep cluster 2175	1	0	0.85	0.15
IncU,rep cluster 2183	2	0	1.7	0.3
IncU,rep cluster 2350	1	1	1.7	0.3
IncU,rep cluster 2373	1	0	0.85	0.15
IncU,rep cluster 658	3	0	2.54	0.46
IncW	12	0	10.17	1.83
IncX1	302	95	336.55	60.45
IncX1,IncX3	30	21	43.23	7.77
IncX1,IncX3,rep cluster 1527	1	0	0.85	0.15
IncX1,IncX9	4	0	3.39	0.61
IncX1,IncY	3	0	2.54	0.46
IncX1,rep cluster 1195	1	0	0.85	0.15
IncX1,rep cluster 1254	1	0	0.85	0.15
IncX1,rep cluster 2335	2	0	1.7	0.3
IncX1,rep cluster 312	1	0	0.85	0.15
IncX2	10	2	10.17	1.83
IncX3	101	482	494.23	88.77
IncX3,IncY	1	0	0.85	0.15
IncX3,rep cluster 1195	11	88	83.93	15.07
IncX3,rep cluster 1254	1	0	0.85	0.15
IncX3,rep cluster 1418	1	0	0.85	0.15
IncX4	152	250	340.79	61.21
IncX4,rep cluster 991	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
IncX5	30	10	33.91	6.09
IncX6	2	0	1.7	0.3
IncX7	4	0	3.39	0.61
IncX8	4	0	3.39	0.61
IncY	264	2	225.5	40.5
IncY,rep cluster 1195	1	0	0.85	0.15
IncY,rep cluster 1254	1	0	0.85	0.15
IncY,rep cluster 1704	1	0	0.85	0.15
IncY,rep cluster 1804	1	0	0.85	0.15
IncY,rep cluster 2232	1	0	0.85	0.15
rep cluster 1	124	3	107.66	19.34
rep cluster 100	4	0	3.39	0.61
rep cluster 1002	1	0	0.85	0.15
rep cluster 1005	1	0	0.85	0.15
rep cluster 1006	11	0	9.33	1.67
rep cluster 1006,rep cluster 1299,rep cluster 1607,rep cluster 1956	2	0	1.7	0.3
rep cluster 1006,rep cluster 1328,rep cluster 1465	3	0	2.54	0.46
rep cluster 1006,rep cluster 1465,rep cluster 731	4	0	3.39	0.61
rep cluster 1006,rep cluster 1567,rep cluster 707	1	0	0.85	0.15
rep cluster 1006,rep cluster 167	1	0	0.85	0.15
rep cluster 1006,rep cluster 707	8	0	6.78	1.22
rep cluster 1006,rep cluster 731	6	0	5.09	0.91
rep cluster 1006,rep cluster 778	1	0	0.85	0.15
rep cluster 101,rep cluster 287	4	0	3.39	0.61
rep cluster 1010	14	0	11.87	2.13
rep cluster 1010,rep cluster 2037	2	0	1.7	0.3
rep cluster 1011	2	0	1.7	0.3
rep cluster 1011,rep cluster 1420	1	0	0.85	0.15
rep cluster 1012	3	0	2.54	0.46
rep cluster 1012,rep cluster 282	1	0	0.85	0.15
rep cluster 1015	1	0	0.85	0.15
rep cluster 1016	2	0	1.7	0.3
rep cluster 1016,rep cluster 252	1	0	0.85	0.15
rep cluster 1017	74	30	88.16	15.84
rep cluster 1017,rep cluster 1018,rep cluster 1110	1	0	0.85	0.15
rep cluster 1017,rep cluster 1018,rep cluster 1118,rep cluster 1142	1	0	0.85	0.15
rep cluster 1017,rep cluster 1018,rep cluster 1142,rep cluster 1281	3	0	2.54	0.46
rep cluster 1017,rep cluster 1018,rep cluster 1215,rep cluster 1230,rep cluster 2104	1	0	0.85	0.15
rep cluster 1017,rep cluster 1018,rep cluster 1733	2	0	1.7	0.3
rep cluster 1017,rep cluster 1110	16	0	13.56	2.44
rep cluster 1017,rep cluster 1110,rep cluster 1118,rep cluster 1439	1	0	0.85	0.15
rep cluster 1017,rep cluster 1110,rep cluster 1302	1	0	0.85	0.15
rep cluster 1017,rep cluster 1110,rep cluster 1442	1	0	0.85	0.15
rep cluster 1017,rep cluster 1110,rep cluster 1442,rep cluster 910	1	0	0.85	0.15
rep cluster 1017,rep cluster 1110,rep cluster 155	1	0	0.85	0.15
rep cluster 1017,rep cluster 1110,rep cluster 2341	1	0	0.85	0.15
rep cluster 1017,rep cluster 1118	2	0	1.7	0.3
rep cluster 1017,rep cluster 1118,rep cluster 1215	7	0	5.93	1.07

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1017,rep cluster 1118,rep cluster 1215,rep cluster 2133,rep cluster 2214	4	0	3.39	0.61
rep cluster 1017,rep cluster 1118,rep cluster 1215,rep cluster 2234	1	0	0.85	0.15
rep cluster 1017,rep cluster 1118,rep cluster 1300	1	0	0.85	0.15
rep cluster 1017,rep cluster 1118,rep cluster 1733,rep cluster 2100	6	12	15.26	2.74
rep cluster 1017,rep cluster 1118,rep cluster 1733,rep cluster 2100,rep cluster 765	2	0	1.7	0.3
rep cluster 1017,rep cluster 1118,rep cluster 2215	1	0	0.85	0.15
rep cluster 1017,rep cluster 1142	42	0	35.6	6.4
rep cluster 1017,rep cluster 1142,rep cluster 1281	1	0	0.85	0.15
rep cluster 1017,rep cluster 1142,rep cluster 992	2	0	1.7	0.3
rep cluster 1017,rep cluster 1215	205	0	173.78	31.22
rep cluster 1017,rep cluster 1215,rep cluster 1230	1	0	0.85	0.15
rep cluster 1017,rep cluster 1215,rep cluster 1256	1	0	0.85	0.15
rep cluster 1017,rep cluster 1215,rep cluster 1281,rep cluster 1733,rep cluster 2100	1	0	0.85	0.15
rep cluster 1017,rep cluster 1215,rep cluster 1442,rep cluster 1947,rep cluster 2409,rep cluster 910	1	0	0.85	0.15
rep cluster 1017,rep cluster 1215,rep cluster 155,rep cluster 2176	1	0	0.85	0.15
rep cluster 1017,rep cluster 1215,rep cluster 1733,rep cluster 1947,rep cluster 2100,rep cluster 910	1	0	0.85	0.15
rep cluster 1017,rep cluster 1215,rep cluster 1733,rep cluster 2100	2	0	1.7	0.3
rep cluster 1017,rep cluster 1215,rep cluster 2176	1	0	0.85	0.15
rep cluster 1017,rep cluster 1215,rep cluster 2214	15	0	12.72	2.28
rep cluster 1017,rep cluster 1215,rep cluster 910	1	0	0.85	0.15
rep cluster 1017,rep cluster 1216,rep cluster 1733	5	0	4.24	0.76
rep cluster 1017,rep cluster 1230,rep cluster 1281,rep cluster 1596	4	0	3.39	0.61
rep cluster 1017,rep cluster 1281	1	0	0.85	0.15
rep cluster 1017,rep cluster 1281,rep cluster 1596	2	0	1.7	0.3
rep cluster 1017,rep cluster 1300	1	0	0.85	0.15
rep cluster 1017,rep cluster 1302	1	0	0.85	0.15
rep cluster 1017,rep cluster 1442,rep cluster 1733,rep cluster 2214,rep cluster 2409,rep cluster 910	1	0	0.85	0.15
rep cluster 1017,rep cluster 155	7	0	5.93	1.07
rep cluster 1017,rep cluster 155,rep cluster 1733	3	0	2.54	0.46
rep cluster 1017,rep cluster 155,rep cluster 1733,rep cluster 2100	6	0	5.09	0.91
rep cluster 1017,rep cluster 155,rep cluster 1733,rep cluster 2100,rep cluster 2176	2	0	1.7	0.3
rep cluster 1017,rep cluster 155,rep cluster 1733,rep cluster 2214	1	0	0.85	0.15
rep cluster 1017,rep cluster 155,rep cluster 2176	5	0	4.24	0.76
rep cluster 1017,rep cluster 1596,rep cluster 1944	1	0	0.85	0.15
rep cluster 1017,rep cluster 1733	1	0	0.85	0.15
rep cluster 1017,rep cluster 1733,rep cluster 1947,rep cluster 2100	1	0	0.85	0.15
rep cluster 1017,rep cluster 1733,rep cluster 2100	306	0	259.41	46.59
rep cluster 1017,rep cluster 1733,rep cluster 2100,rep cluster 2176	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1017,rep cluster 1733,rep cluster 2100,rep cluster 2214	11	0	9.33	1.67
rep cluster 1017,rep cluster 1733,rep cluster 2100,rep cluster 765	3	0	2.54	0.46
rep cluster 1017,rep cluster 1733,rep cluster 2100,rep cluster 910	1	0	0.85	0.15
rep cluster 1017,rep cluster 1733,rep cluster 2214	27	0	22.89	4.11
rep cluster 1017,rep cluster 2100	3	0	2.54	0.46
rep cluster 1017,rep cluster 2120	2	0	1.7	0.3
rep cluster 1017,rep cluster 2246	1	0	0.85	0.15
rep cluster 1017,rep cluster 2409	1	0	0.85	0.15
rep cluster 1017,rep cluster 765	1	0	0.85	0.15
rep cluster 1017,rep cluster 910	1	0	0.85	0.15
rep cluster 1017,rep cluster 936	1	0	0.85	0.15
rep cluster 1018	12	5	14.41	2.59
rep cluster 1018,rep cluster 1093,rep cluster 185,rep cluster 2351	1	0	0.85	0.15
rep cluster 1018,rep cluster 1110,rep cluster 1442,rep cluster 2120	2	0	1.7	0.3
rep cluster 1018,rep cluster 1118	1	0	0.85	0.15
rep cluster 1018,rep cluster 1118,rep cluster 1442,rep cluster 910	2	0	1.7	0.3
rep cluster 1018,rep cluster 1118,rep cluster 2120	2	0	1.7	0.3
rep cluster 1018,rep cluster 1142	1	0	0.85	0.15
rep cluster 1018,rep cluster 1142,rep cluster 1215	1	0	0.85	0.15
rep cluster 1018,rep cluster 1142,rep cluster 1281	3	0	2.54	0.46
rep cluster 1018,rep cluster 1197,rep cluster 185,rep cluster 2146	1	0	0.85	0.15
rep cluster 1018,rep cluster 1215	1	0	0.85	0.15
rep cluster 1018,rep cluster 1281,rep cluster 1442,rep cluster 910	1	0	0.85	0.15
rep cluster 1018,rep cluster 1442,rep cluster 910	1	0	0.85	0.15
rep cluster 1018,rep cluster 155	3	0	2.54	0.46
rep cluster 1018,rep cluster 155,rep cluster 2439	1	0	0.85	0.15
rep cluster 1018,rep cluster 1650,rep cluster 2120	1	0	0.85	0.15
rep cluster 1018,rep cluster 1763,rep cluster 185	3	0	2.54	0.46
rep cluster 1018,rep cluster 1763,rep cluster 185,rep cluster 1968	1	0	0.85	0.15
rep cluster 1018,rep cluster 183,rep cluster 185,rep cluster 2146	1	0	0.85	0.15
rep cluster 1018,rep cluster 185	13	0	11.02	1.98
rep cluster 1018,rep cluster 185,rep cluster 1968	1	0	0.85	0.15
rep cluster 1018,rep cluster 185,rep cluster 889	3	0	2.54	0.46
rep cluster 1018,rep cluster 185,rep cluster 893	1	0	0.85	0.15
rep cluster 1018,rep cluster 1947,rep cluster 2246	1	0	0.85	0.15
rep cluster 1018,rep cluster 2104	2	1	2.54	0.46
rep cluster 1018,rep cluster 2120	5	0	4.24	0.76
rep cluster 1018,rep cluster 2301	1	0	0.85	0.15
rep cluster 1018,rep cluster 2350	2	0	1.7	0.3
rep cluster 1018,rep cluster 889	1	0	0.85	0.15
rep cluster 1018,rep cluster 893	3	0	2.54	0.46
rep cluster 1018,rep cluster 976,rep cluster 992	1	0	0.85	0.15
rep cluster 1019	4	2	5.09	0.91
rep cluster 102	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1021	1	0	0.85	0.15
rep cluster 1022	2	0	1.7	0.3
rep cluster 1023	4	0	3.39	0.61
rep cluster 1027	81	0	68.67	12.33
rep cluster 1027,rep cluster 1118	1	0	0.85	0.15
rep cluster 1027,rep cluster 1197	1	0	0.85	0.15
rep cluster 1027,rep cluster 183	1	0	0.85	0.15
rep cluster 1027,rep cluster 980	1	0	0.85	0.15
rep cluster 1027,rep cluster 992	1	0	0.85	0.15
rep cluster 103	1	0	0.85	0.15
rep cluster 1030	4	0	3.39	0.61
rep cluster 1031	1	0	0.85	0.15
rep cluster 1039	1	0	0.85	0.15
rep cluster 104,rep cluster 1581,rep cluster 80	2	0	1.7	0.3
rep cluster 1041	7	2	7.63	1.37
rep cluster 1041,rep cluster 1254	1	0	0.85	0.15
rep cluster 1042	3	0	2.54	0.46
rep cluster 1044	1	0	0.85	0.15
rep cluster 1046	6	2	6.78	1.22
rep cluster 1049	4	0	3.39	0.61
rep cluster 105	1	0	0.85	0.15
rep cluster 1050	2	0	1.7	0.3
rep cluster 1053	43	0	36.45	6.55
rep cluster 1053,rep cluster 737	1	0	0.85	0.15
rep cluster 1057	39	13	44.08	7.92
rep cluster 1058	32	0	27.13	4.87
rep cluster 1058,rep cluster 1229	1	0	0.85	0.15
rep cluster 1058,rep cluster 172,rep cluster 1948,rep cluster 707	1	0	0.85	0.15
rep cluster 1058,rep cluster 1948	3	0	2.54	0.46
rep cluster 1058,rep cluster 556	1	0	0.85	0.15
rep cluster 1058,rep cluster 893	2	0	1.7	0.3
rep cluster 1059	1	0	0.85	0.15
rep cluster 106	2	0	1.7	0.3
rep cluster 106,rep cluster 287	1	0	0.85	0.15
rep cluster 1060	2	0	1.7	0.3
rep cluster 1061	14	0	11.87	2.13
rep cluster 1061,rep cluster 429	1	0	0.85	0.15
rep cluster 1061,rep cluster 461	1	0	0.85	0.15
rep cluster 1066	2	0	1.7	0.3
rep cluster 1067	10	0	8.48	1.52
rep cluster 1068	11	4	12.72	2.28
rep cluster 1068,rep cluster 2022	1	0	0.85	0.15
rep cluster 1068,rep cluster 2037	2	0	1.7	0.3
rep cluster 1069	27	0	22.89	4.11
rep cluster 1069,rep cluster 2021	2	0	1.7	0.3
rep cluster 107	6	0	5.09	0.91
rep cluster 1070	6	0	5.09	0.91
rep cluster 1071	8	4	10.17	1.83
rep cluster 1071,rep cluster 1265	1	0	0.85	0.15
rep cluster 1071,rep cluster 1265,rep cluster 1280	1	0	0.85	0.15
rep cluster 1072	15	0	12.72	2.28
rep cluster 1074	1	0	0.85	0.15
rep cluster 1074,rep cluster 1294	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1077	13	0	11.02	1.98
rep cluster 1078	12	0	10.17	1.83
rep cluster 1078,rep cluster 1118	1	1	1.7	0.3
rep cluster 1078,rep cluster 180	1	0	0.85	0.15
rep cluster 1079	7	2	7.63	1.37
rep cluster 1081	3	0	2.54	0.46
rep cluster 1085	4	0	3.39	0.61
rep cluster 1087	2	0	1.7	0.3
rep cluster 1088	13	0	11.02	1.98
rep cluster 1092	1	0	0.85	0.15
rep cluster 1093	68	0	57.65	10.35
rep cluster 1093,rep cluster 1197	1	0	0.85	0.15
rep cluster 1093,rep cluster 1197,rep cluster 185	1	0	0.85	0.15
rep cluster 1093,rep cluster 1197,rep cluster 889	1	0	0.85	0.15
rep cluster 1093,rep cluster 1249,rep cluster 185	1	0	0.85	0.15
rep cluster 1093,rep cluster 1291	1	0	0.85	0.15
rep cluster 1093,rep cluster 1742	1	0	0.85	0.15
rep cluster 1093,rep cluster 1763,rep cluster 185	1	0	0.85	0.15
rep cluster 1093,rep cluster 1763,rep cluster 889	4	0	3.39	0.61
rep cluster 1093,rep cluster 1763,rep cluster 889,rep cluster 943	1	0	0.85	0.15
rep cluster 1093,rep cluster 183	5	0	4.24	0.76
rep cluster 1093,rep cluster 185	2	0	1.7	0.3
rep cluster 1093,rep cluster 185,rep cluster 2351,rep cluster 943	1	0	0.85	0.15
rep cluster 1093,rep cluster 889	16	0	13.56	2.44
rep cluster 1093,rep cluster 889,rep cluster 893	2	0	1.7	0.3
rep cluster 1093,rep cluster 893	1	0	0.85	0.15
rep cluster 1096	3	0	2.54	0.46
rep cluster 1097	3	9	10.17	1.83
rep cluster 1099	25	0	21.19	3.81
rep cluster 110	42	18	50.86	9.14
rep cluster 110,rep cluster 1118	1	2	2.54	0.46
rep cluster 110,rep cluster 2446	2	0	1.7	0.3
rep cluster 110,rep cluster 554	5	0	4.24	0.76
rep cluster 110,rep cluster 556	2	0	1.7	0.3
rep cluster 1100	1	0	0.85	0.15
rep cluster 1100,rep cluster 219	1	0	0.85	0.15
rep cluster 1101	1	0	0.85	0.15
rep cluster 1103	1	0	0.85	0.15
rep cluster 1104	1	0	0.85	0.15
rep cluster 1110	20	0	16.95	3.05
rep cluster 1110,rep cluster 1118	1	0	0.85	0.15
rep cluster 1110,rep cluster 1442	6	0	5.09	0.91
rep cluster 1110,rep cluster 1448,rep cluster 2222	1	0	0.85	0.15
rep cluster 1110,rep cluster 1733,rep cluster 2214	1	0	0.85	0.15
rep cluster 1110,rep cluster 1986	4	0	3.39	0.61
rep cluster 1110,rep cluster 2341	6	3	7.63	1.37
rep cluster 1111	2	0	1.7	0.3
rep cluster 1112	3	0	2.54	0.46
rep cluster 1113	2	0	1.7	0.3
rep cluster 1115	36	0	30.52	5.48
rep cluster 1115,rep cluster 1254	19	2	17.8	3.2
rep cluster 1115,rep cluster 2238	2	0	1.7	0.3

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1117	2	0	1.7	0.3
rep cluster 1118	83	60	121.23	21.77
rep cluster 1118,rep cluster 1142,rep cluster 1215,rep cluster 2133	1	0	0.85	0.15
rep cluster 1118,rep cluster 1215	6	0	5.09	0.91
rep cluster 1118,rep cluster 1215,rep cluster 1281	4	3	5.93	1.07
rep cluster 1118,rep cluster 1215,rep cluster 1281,rep cluster 2214	1	0	0.85	0.15
rep cluster 1118,rep cluster 1215,rep cluster 2214	3	0	2.54	0.46
rep cluster 1118,rep cluster 1257	1	0	0.85	0.15
rep cluster 1118,rep cluster 1281,rep cluster 1442,rep cluster 765	1	0	0.85	0.15
rep cluster 1118,rep cluster 1442	1	0	0.85	0.15
rep cluster 1118,rep cluster 1442,rep cluster 2215	1	0	0.85	0.15
rep cluster 1118,rep cluster 1448,rep cluster 418	1	0	0.85	0.15
rep cluster 1118,rep cluster 180	3	0	2.54	0.46
rep cluster 1118,rep cluster 180,rep cluster 185	1	0	0.85	0.15
rep cluster 1118,rep cluster 185	12	0	10.17	1.83
rep cluster 1118,rep cluster 185,rep cluster 2301	2	0	1.7	0.3
rep cluster 1118,rep cluster 185,rep cluster 589	1	0	0.85	0.15
rep cluster 1118,rep cluster 185,rep cluster 889	1	0	0.85	0.15
rep cluster 1118,rep cluster 185,rep cluster 980	1	0	0.85	0.15
rep cluster 1118,rep cluster 1947	1	0	0.85	0.15
rep cluster 1118,rep cluster 2215	1	0	0.85	0.15
rep cluster 1118,rep cluster 2222	1	0	0.85	0.15
rep cluster 1118,rep cluster 2301,rep cluster 980	1	0	0.85	0.15
rep cluster 1118,rep cluster 2409,rep cluster 910	1	0	0.85	0.15
rep cluster 1118,rep cluster 418	1	0	0.85	0.15
rep cluster 1118,rep cluster 530	4	0	3.39	0.61
rep cluster 1118,rep cluster 554,rep cluster 976,rep cluster 992	1	0	0.85	0.15
rep cluster 1118,rep cluster 554,rep cluster 992	1	0	0.85	0.15
rep cluster 1118,rep cluster 765	1	0	0.85	0.15
rep cluster 1118,rep cluster 796	2	0	1.7	0.3
rep cluster 1118,rep cluster 889	1	0	0.85	0.15
rep cluster 1118,rep cluster 893	2	0	1.7	0.3
rep cluster 1118,rep cluster 910	3	0	2.54	0.46
rep cluster 1118,rep cluster 976	1	0	0.85	0.15
rep cluster 1118,rep cluster 980	18	0	15.26	2.74
rep cluster 1118,rep cluster 992	44	3	39.84	7.16
rep cluster 112	2	0	1.7	0.3
rep cluster 1121	22	0	18.65	3.35
rep cluster 1123	2	0	1.7	0.3
rep cluster 1124	1	0	0.85	0.15
rep cluster 1125	1	0	0.85	0.15
rep cluster 1127	3	0	2.54	0.46
rep cluster 1128	3	0	2.54	0.46
rep cluster 1129	1	0	0.85	0.15
rep cluster 1131	2	0	1.7	0.3
rep cluster 1134	16	2	15.26	2.74
rep cluster 1134,rep cluster 1254	2	0	1.7	0.3
rep cluster 1136	3	0	2.54	0.46
rep cluster 1138	2	0	1.7	0.3
rep cluster 1139	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 114	1	0	0.85	0.15
rep cluster 1140	2	0	1.7	0.3
rep cluster 1140,rep cluster 848	1	0	0.85	0.15
rep cluster 1142	39	16	46.63	8.37
rep cluster 1142,rep cluster 1215	1	0	0.85	0.15
rep cluster 1142,rep cluster 1230,rep cluster 1947	1	0	0.85	0.15
rep cluster 1142,rep cluster 1230,rep cluster 1947,rep cluster 2246	1	0	0.85	0.15
rep cluster 1142,rep cluster 1281	4	0	3.39	0.61
rep cluster 1142,rep cluster 1281,rep cluster 992	2	0	1.7	0.3
rep cluster 1142,rep cluster 1947	3	0	2.54	0.46
rep cluster 1142,rep cluster 2222	1	0	0.85	0.15
rep cluster 1143	2	0	1.7	0.3
rep cluster 1143,rep cluster 1152,rep cluster 1320	1	0	0.85	0.15
rep cluster 1144	1	0	0.85	0.15
rep cluster 1148	1	0	0.85	0.15
rep cluster 1149	3	0	2.54	0.46
rep cluster 1149,rep cluster 1319,rep cluster 855	1	0	0.85	0.15
rep cluster 1149,rep cluster 543	1	0	0.85	0.15
rep cluster 1151	9	0	7.63	1.37
rep cluster 1152,rep cluster 1320	4	2	5.09	0.91
rep cluster 1154	7	0	5.93	1.07
rep cluster 1154,rep cluster 1172	2	0	1.7	0.3
rep cluster 1155	21	6	22.89	4.11
rep cluster 1156	2	0	1.7	0.3
rep cluster 1159	5	0	4.24	0.76
rep cluster 1164	17	0	14.41	2.59
rep cluster 1166	2	0	1.7	0.3
rep cluster 1167	1	0	0.85	0.15
rep cluster 1168	4	0	3.39	0.61
rep cluster 1168,rep cluster 1319	1	0	0.85	0.15
rep cluster 1168,rep cluster 541	1	0	0.85	0.15
rep cluster 1168,rep cluster 647	2	0	1.7	0.3
rep cluster 1169	39	2	34.76	6.24
rep cluster 1169,rep cluster 2021	2	0	1.7	0.3
rep cluster 1171	19	0	16.11	2.89
rep cluster 1172	161	2	138.18	24.82
rep cluster 1172,rep cluster 1254	1	0	0.85	0.15
rep cluster 1172,rep cluster 1259	1	0	0.85	0.15
rep cluster 1172,rep cluster 1282	1	0	0.85	0.15
rep cluster 1172,rep cluster 1364	3	0	2.54	0.46
rep cluster 1172,rep cluster 1656	1	0	0.85	0.15
rep cluster 1172,rep cluster 1903	5	0	4.24	0.76
rep cluster 1172,rep cluster 1906	4	0	3.39	0.61
rep cluster 1172,rep cluster 1919	1	0	0.85	0.15
rep cluster 1172,rep cluster 734	1	0	0.85	0.15
rep cluster 1173	2	0	1.7	0.3
rep cluster 1174	1	0	0.85	0.15
rep cluster 1175	2	0	1.7	0.3
rep cluster 1176	1	0	0.85	0.15
rep cluster 1178	2	0	1.7	0.3
rep cluster 1180	1	0	0.85	0.15
rep cluster 1181	4	0	3.39	0.61
rep cluster 1186	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1187	2	0	1.7	0.3
rep cluster 1188	2	0	1.7	0.3
rep cluster 1189	1	0	0.85	0.15
rep cluster 119	2	0	1.7	0.3
rep cluster 119,rep cluster 287	3	0	2.54	0.46
rep cluster 1190	1	0	0.85	0.15
rep cluster 1192	1	0	0.85	0.15
rep cluster 1193	4	0	3.39	0.61
rep cluster 1194	2	0	1.7	0.3
rep cluster 1195	24	25	41.54	7.46
rep cluster 1197	113	0	95.79	17.21
rep cluster 1197,rep cluster 185	5	0	4.24	0.76
rep cluster 1197,rep cluster 185,rep cluster 943	1	0	0.85	0.15
rep cluster 1197,rep cluster 889	2	0	1.7	0.3
rep cluster 12,rep cluster 1879,rep cluster 1910,rep cluster 1922,rep cluster 40	11	0	9.33	1.67
rep cluster 12,rep cluster 1922	5	0	4.24	0.76
rep cluster 120	3	0	2.54	0.46
rep cluster 1200	1	0	0.85	0.15
rep cluster 1202	1	0	0.85	0.15
rep cluster 1203	9	0	7.63	1.37
rep cluster 1203,rep cluster 1441	1	0	0.85	0.15
rep cluster 1205	3	0	2.54	0.46
rep cluster 1205,rep cluster 1855	1	0	0.85	0.15
rep cluster 1206	1	0	0.85	0.15
rep cluster 1207	1	0	0.85	0.15
rep cluster 1208	1	0	0.85	0.15
rep cluster 1209	3	0	2.54	0.46
rep cluster 1212	1	0	0.85	0.15
rep cluster 1213	1	0	0.85	0.15
rep cluster 1214	1	0	0.85	0.15
rep cluster 1215	107	0	90.71	16.29
rep cluster 1215,rep cluster 1256	1	0	0.85	0.15
rep cluster 1215,rep cluster 1256,rep cluster 765	1	0	0.85	0.15
rep cluster 1215,rep cluster 1281	1	0	0.85	0.15
rep cluster 1215,rep cluster 1281,rep cluster 1442	1	0	0.85	0.15
rep cluster 1215,rep cluster 1281,rep cluster 2214	1	0	0.85	0.15
rep cluster 1215,rep cluster 155	1	0	0.85	0.15
rep cluster 1215,rep cluster 1733	1	0	0.85	0.15
rep cluster 1215,rep cluster 1947	1	0	0.85	0.15
rep cluster 1215,rep cluster 2222	1	0	0.85	0.15
rep cluster 1215,rep cluster 765	1	0	0.85	0.15
rep cluster 1215,rep cluster 910	1	0	0.85	0.15
rep cluster 1216	1	0	0.85	0.15
rep cluster 1216,rep cluster 1733	6	0	5.09	0.91
rep cluster 1216,rep cluster 1733,rep cluster 2133	2	0	1.7	0.3
rep cluster 1218	65	0	55.1	9.9
rep cluster 1218,rep cluster 1226	1	0	0.85	0.15
rep cluster 1218,rep cluster 1254	1	0	0.85	0.15
rep cluster 1218,rep cluster 1364	1	0	0.85	0.15
rep cluster 1218,rep cluster 1481	1	0	0.85	0.15
rep cluster 1218,rep cluster 639	3	0	2.54	0.46
rep cluster 1219	14	0	11.87	2.13
rep cluster 1219,rep cluster 1280	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 122	1	0	0.85	0.15
rep cluster 1220	34	0	28.82	5.18
rep cluster 1222	1	0	0.85	0.15
rep cluster 1223	4	0	3.39	0.61
rep cluster 1224	2	0	1.7	0.3
rep cluster 1225	1	0	0.85	0.15
rep cluster 1226	56	11	56.8	10.2
rep cluster 1226,rep cluster 1233	1	0	0.85	0.15
rep cluster 1226,rep cluster 1364	2	0	1.7	0.3
rep cluster 1229	26	2	23.74	4.26
rep cluster 1229,rep cluster 1328	1	0	0.85	0.15
rep cluster 1229,rep cluster 1880	2	0	1.7	0.3
rep cluster 1229,rep cluster 1880,rep cluster 1948	1	0	0.85	0.15
rep cluster 1229,rep cluster 1880,rep cluster 2119	1	0	0.85	0.15
rep cluster 1229,rep cluster 1948	7	0	5.93	1.07
rep cluster 1229,rep cluster 1948,rep cluster 589	2	0	1.7	0.3
rep cluster 1229,rep cluster 2119	2	0	1.7	0.3
rep cluster 1229,rep cluster 589	1	0	0.85	0.15
rep cluster 1229,rep cluster 707	2	0	1.7	0.3
rep cluster 123	1	0	0.85	0.15
rep cluster 1230	11	7	15.26	2.74
rep cluster 1230,rep cluster 2234	0	1	0.85	0.15
rep cluster 1231	1	0	0.85	0.15
rep cluster 1232	68	0	57.65	10.35
rep cluster 1233	17	0	14.41	2.59
rep cluster 1233,rep cluster 1282	1	0	0.85	0.15
rep cluster 1233,rep cluster 1364	1	0	0.85	0.15
rep cluster 1233,rep cluster 2359	1	0	0.85	0.15
rep cluster 1234	1	0	0.85	0.15
rep cluster 1239	2	0	1.7	0.3
rep cluster 124	1	0	0.85	0.15
rep cluster 1241	3	0	2.54	0.46
rep cluster 1244	30	0	25.43	4.57
rep cluster 1244,rep cluster 1511	1	0	0.85	0.15
rep cluster 1244,rep cluster 1880	3	0	2.54	0.46
rep cluster 1244,rep cluster 2119	2	0	1.7	0.3
rep cluster 1244,rep cluster 707	1	0	0.85	0.15
rep cluster 1246	2	0	1.7	0.3
rep cluster 1247	4	0	3.39	0.61
rep cluster 1247,rep cluster 1254	1	0	0.85	0.15
rep cluster 1249	32	0	27.13	4.87
rep cluster 1249,rep cluster 183	1	0	0.85	0.15
rep cluster 1249,rep cluster 185	2	0	1.7	0.3
rep cluster 1249,rep cluster 185,rep cluster 2351,rep cluster 943	1	0	0.85	0.15
rep cluster 125	145	61	174.63	31.37
rep cluster 125,rep cluster 163	1	0	0.85	0.15
rep cluster 125,rep cluster 163,rep cluster 169	1	0	0.85	0.15
rep cluster 125,rep cluster 163,rep cluster 1721	1	0	0.85	0.15
rep cluster 125,rep cluster 164	11	0	9.33	1.67
rep cluster 125,rep cluster 169,rep cluster 459	1	0	0.85	0.15
rep cluster 125,rep cluster 1698	8	4	10.17	1.83
rep cluster 125,rep cluster 1698,rep cluster 1751	1	0	0.85	0.15
rep cluster 125,rep cluster 1721	58	0	49.17	8.83

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 125,rep cluster 1738,rep cluster 1821	1	0	0.85	0.15
rep cluster 125,rep cluster 1869	4	0	3.39	0.61
rep cluster 125,rep cluster 459	7	0	5.93	1.07
rep cluster 125,rep cluster 459,rep cluster 920	1	0	0.85	0.15
rep cluster 1254	62	5	56.8	10.2
rep cluster 1254,rep cluster 1259	1	0	0.85	0.15
rep cluster 1254,rep cluster 1357	1	0	0.85	0.15
rep cluster 1254,rep cluster 1364	5	0	4.24	0.76
rep cluster 1255	21	0	17.8	3.2
rep cluster 1255,rep cluster 1328	2	0	1.7	0.3
rep cluster 1255,rep cluster 1511,rep cluster 556,rep cluster 707	1	0	0.85	0.15
rep cluster 1255,rep cluster 1880	2	0	1.7	0.3
rep cluster 1255,rep cluster 2119	3	0	2.54	0.46
rep cluster 1255,rep cluster 556,rep cluster 707	1	0	0.85	0.15
rep cluster 1255,rep cluster 707	1	0	0.85	0.15
rep cluster 1255,rep cluster 778	1	0	0.85	0.15
rep cluster 1256	19	4	19.5	3.5
rep cluster 1256,rep cluster 1442	1	0	0.85	0.15
rep cluster 1256,rep cluster 2366	2	0	1.7	0.3
rep cluster 1257	2	0	1.7	0.3
rep cluster 1258	1	0	0.85	0.15
rep cluster 1259	12	2	11.87	2.13
rep cluster 1259,rep cluster 1364	1	0	0.85	0.15
rep cluster 1259,rep cluster 1903	7	0	5.93	1.07
rep cluster 1259,rep cluster 639	1	0	0.85	0.15
rep cluster 1260	2	0	1.7	0.3
rep cluster 1261	11	0	9.33	1.67
rep cluster 1262	9	0	7.63	1.37
rep cluster 1264	3	0	2.54	0.46
rep cluster 1265	21	4	21.19	3.81
rep cluster 1269	3	0	2.54	0.46
rep cluster 1274	1	0	0.85	0.15
rep cluster 1277	1	0	0.85	0.15
rep cluster 1278	3	0	2.54	0.46
rep cluster 1280	11	0	9.33	1.67
rep cluster 1281	10	0	8.48	1.52
rep cluster 1281,rep cluster 910	2	0	1.7	0.3
rep cluster 1282	27	0	22.89	4.11
rep cluster 1282,rep cluster 1656	2	0	1.7	0.3
rep cluster 1283	19	0	16.11	2.89
rep cluster 1285	1	0	0.85	0.15
rep cluster 1286	1	0	0.85	0.15
rep cluster 1287	2	0	1.7	0.3
rep cluster 1288	9	0	7.63	1.37
rep cluster 1289	8	0	6.78	1.22
rep cluster 1289,rep cluster 156,rep cluster 461	1	0	0.85	0.15
rep cluster 1290	1	0	0.85	0.15
rep cluster 1291	15	0	12.72	2.28
rep cluster 1291,rep cluster 185	1	0	0.85	0.15
rep cluster 1294	2	0	1.7	0.3
rep cluster 1294,rep cluster 992	3	0	2.54	0.46
rep cluster 1295	2	0	1.7	0.3
rep cluster 1297	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1299	4	0	3.39	0.61
rep cluster 1299,rep cluster 1956	1	0	0.85	0.15
rep cluster 1300	1	0	0.85	0.15
rep cluster 1301	1	0	0.85	0.15
rep cluster 1302	9	0	7.63	1.37
rep cluster 1304	7	2	7.63	1.37
rep cluster 1306	2	0	1.7	0.3
rep cluster 1307	42	0	35.6	6.4
rep cluster 1309	1	0	0.85	0.15
rep cluster 1310	2	0	1.7	0.3
rep cluster 1311	12	0	10.17	1.83
rep cluster 1313	1	0	0.85	0.15
rep cluster 1316	1	0	0.85	0.15
rep cluster 1317	11	0	9.33	1.67
rep cluster 1319	9	0	7.63	1.37
rep cluster 1319,rep cluster 543	1	0	0.85	0.15
rep cluster 1319,rep cluster 736	1	0	0.85	0.15
rep cluster 1319,rep cluster 754	3	0	2.54	0.46
rep cluster 1320,rep cluster 1351	1	0	0.85	0.15
rep cluster 1322	1	0	0.85	0.15
rep cluster 1323	20	0	16.95	3.05
rep cluster 1325	1	0	0.85	0.15
rep cluster 1328	49	0	41.54	7.46
rep cluster 1328,rep cluster 1465	8	0	6.78	1.22
rep cluster 1328,rep cluster 1465,rep cluster 1607,rep cluster 167	1	0	0.85	0.15
rep cluster 1328,rep cluster 1465,rep cluster 707	1	0	0.85	0.15
rep cluster 1328,rep cluster 1607	3	0	2.54	0.46
rep cluster 1328,rep cluster 1607,rep cluster 707	3	0	2.54	0.46
rep cluster 1328,rep cluster 1618	2	0	1.7	0.3
rep cluster 1328,rep cluster 167	1	0	0.85	0.15
rep cluster 1328,rep cluster 167,rep cluster 707	1	0	0.85	0.15
rep cluster 1328,rep cluster 172	1	0	0.85	0.15
rep cluster 1328,rep cluster 1827	1	0	0.85	0.15
rep cluster 1328,rep cluster 707	2	0	1.7	0.3
rep cluster 1328,rep cluster 731	1	0	0.85	0.15
rep cluster 1331	1	0	0.85	0.15
rep cluster 1332	9	3	10.17	1.83
rep cluster 1333	9	2	9.33	1.67
rep cluster 1336	8	0	6.78	1.22
rep cluster 1336,rep cluster 848	1	0	0.85	0.15
rep cluster 1337	1	2	2.54	0.46
rep cluster 1340	1	0	0.85	0.15
rep cluster 1341	1	0	0.85	0.15
rep cluster 1341,rep cluster 647	1	0	0.85	0.15
rep cluster 1344	1	0	0.85	0.15
rep cluster 1345	3	0	2.54	0.46
rep cluster 1346	1	0	0.85	0.15
rep cluster 1347	1	0	0.85	0.15
rep cluster 135	1	0	0.85	0.15
rep cluster 1355	1	0	0.85	0.15
rep cluster 1357	6	0	5.09	0.91
rep cluster 1359	1	0	0.85	0.15
rep cluster 136	3	0	2.54	0.46

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1362	3	0	2.54	0.46
rep cluster 1362,rep cluster 1874	2	0	1.7	0.3
rep cluster 1364	64	3	56.8	10.2
rep cluster 1367	23	33	47.47	8.53
rep cluster 1367,rep cluster 1804	1	0	0.85	0.15
rep cluster 1368	10	2	10.17	1.83
rep cluster 137	49	0	41.54	7.46
rep cluster 137,rep cluster 146	1	0	0.85	0.15
rep cluster 137,rep cluster 60	6	0	5.09	0.91
rep cluster 1370	1	0	0.85	0.15
rep cluster 1371	6	0	5.09	0.91
rep cluster 1374	29	0	24.58	4.42
rep cluster 1375	2	0	1.7	0.3
rep cluster 1376	2	2	3.39	0.61
rep cluster 1377	3	0	2.54	0.46
rep cluster 1378	1	2	2.54	0.46
rep cluster 1378,rep cluster 554	1	0	0.85	0.15
rep cluster 1379	1	0	0.85	0.15
rep cluster 138	8	0	6.78	1.22
rep cluster 1380	4	0	3.39	0.61
rep cluster 1381	5	0	4.24	0.76
rep cluster 1392	2	0	1.7	0.3
rep cluster 1393	2	0	1.7	0.3
rep cluster 1394	1	0	0.85	0.15
rep cluster 1396	14	0	11.87	2.13
rep cluster 1396,rep cluster 845	1	0	0.85	0.15
rep cluster 1398	9	0	7.63	1.37
rep cluster 1399	1	0	0.85	0.15
rep cluster 1403	9	0	7.63	1.37
rep cluster 1407	1	0	0.85	0.15
rep cluster 141	4	0	3.39	0.61
rep cluster 141,rep cluster 471	1	0	0.85	0.15
rep cluster 1410	41	0	34.76	6.24
rep cluster 1416	1	0	0.85	0.15
rep cluster 1417	10	0	8.48	1.52
rep cluster 1418	70	3	61.88	11.12
rep cluster 1418,rep cluster 1804	1	0	0.85	0.15
rep cluster 1418,rep cluster 2183	8	0	6.78	1.22
rep cluster 1418,rep cluster 3	1	0	0.85	0.15
rep cluster 142	1	0	0.85	0.15
rep cluster 1420	17	0	14.41	2.59
rep cluster 1426	3	0	2.54	0.46
rep cluster 1431	1	0	0.85	0.15
rep cluster 1433	102	0	86.47	15.53
rep cluster 1433,rep cluster 1749	2	0	1.7	0.3
rep cluster 1435	1	0	0.85	0.15
rep cluster 1436	4	0	3.39	0.61
rep cluster 1439	13	0	11.02	1.98
rep cluster 1439,rep cluster 1448	5	0	4.24	0.76
rep cluster 144	1	0	0.85	0.15
rep cluster 1440	2	0	1.7	0.3
rep cluster 1441	25	0	21.19	3.81
rep cluster 1441,rep cluster 2315	2	0	1.7	0.3
rep cluster 1441,rep cluster 2319	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1442	16	2	15.26	2.74
rep cluster 1442,rep cluster 1448	1	0	0.85	0.15
rep cluster 1442,rep cluster 2215	10	0	8.48	1.52
rep cluster 1442,rep cluster 765	1	0	0.85	0.15
rep cluster 1442,rep cluster 910	2	0	1.7	0.3
rep cluster 1443	7	0	5.93	1.07
rep cluster 1444	15	17	27.13	4.87
rep cluster 1445	1	0	0.85	0.15
rep cluster 1445,rep cluster 1822	1	0	0.85	0.15
rep cluster 1448	20	0	16.95	3.05
rep cluster 1448,rep cluster 2194	4	0	3.39	0.61
rep cluster 1448,rep cluster 418	1	0	0.85	0.15
rep cluster 1449	22	0	18.65	3.35
rep cluster 1449,rep cluster 1509	1	0	0.85	0.15
rep cluster 1449,rep cluster 382	1	0	0.85	0.15
rep cluster 1450	1	0	0.85	0.15
rep cluster 1451	8	0	6.78	1.22
rep cluster 1454	23	0	19.5	3.5
rep cluster 1455	0	2	1.7	0.3
rep cluster 1456	4	0	3.39	0.61
rep cluster 146	3	0	2.54	0.46
rep cluster 1461	1	0	0.85	0.15
rep cluster 1462	5	0	4.24	0.76
rep cluster 1464	3	0	2.54	0.46
rep cluster 1465	46	0	39	7
rep cluster 1465,rep cluster 1607	1	0	0.85	0.15
rep cluster 1465,rep cluster 167	2	0	1.7	0.3
rep cluster 1465,rep cluster 707	2	0	1.7	0.3
rep cluster 1465,rep cluster 731	9	0	7.63	1.37
rep cluster 1467	52	0	44.08	7.92
rep cluster 1469	1	0	0.85	0.15
rep cluster 1474	4	0	3.39	0.61
rep cluster 1478	1	0	0.85	0.15
rep cluster 1480	2	0	1.7	0.3
rep cluster 1481	4	0	3.39	0.61
rep cluster 1481,rep cluster 1656	1	0	0.85	0.15
rep cluster 1485	4	0	3.39	0.61
rep cluster 1486	28	25	44.93	8.07
rep cluster 1486,rep cluster 557	4	0	3.39	0.61
rep cluster 1489	1	0	0.85	0.15
rep cluster 1493	1	0	0.85	0.15
rep cluster 1495	1	0	0.85	0.15
rep cluster 1497	1	0	0.85	0.15
rep cluster 1498	1	0	0.85	0.15
rep cluster 150	19	0	16.11	2.89
rep cluster 150,rep cluster 185	2	0	1.7	0.3
rep cluster 150,rep cluster 889	1	0	0.85	0.15
rep cluster 150,rep cluster 893	1	0	0.85	0.15
rep cluster 1500	7	0	5.93	1.07
rep cluster 1501	1	0	0.85	0.15
rep cluster 1502	14	0	11.87	2.13
rep cluster 1503	8	3	9.33	1.67
rep cluster 1503,rep cluster 1510	1	0	0.85	0.15
rep cluster 1503,rep cluster 589	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1506	32	2	28.82	5.18
rep cluster 1506,rep cluster 282	1	0	0.85	0.15
rep cluster 1508	1	0	0.85	0.15
rep cluster 1509	5	0	4.24	0.76
rep cluster 151	2	0	1.7	0.3
rep cluster 1510	7	0	5.93	1.07
rep cluster 1511	4	0	3.39	0.61
rep cluster 1513	1	0	0.85	0.15
rep cluster 1515	1	0	0.85	0.15
rep cluster 1516	5	0	4.24	0.76
rep cluster 1519	60	0	50.86	9.14
rep cluster 152	1	0	0.85	0.15
rep cluster 1520	9	0	7.63	1.37
rep cluster 1520,rep cluster 1575	1	0	0.85	0.15
rep cluster 1520,rep cluster 163,rep cluster 459	1	0	0.85	0.15
rep cluster 1520,rep cluster 1645	1	0	0.85	0.15
rep cluster 1520,rep cluster 1698	8	0	6.78	1.22
rep cluster 1520,rep cluster 1698,rep cluster 1912	2	0	1.7	0.3
rep cluster 1520,rep cluster 1912	1	0	0.85	0.15
rep cluster 1522	3	0	2.54	0.46
rep cluster 1523	3	0	2.54	0.46
rep cluster 1524	2	0	1.7	0.3
rep cluster 1527	5	0	4.24	0.76
rep cluster 1529	12	0	10.17	1.83
rep cluster 153	2	0	1.7	0.3
rep cluster 153,rep cluster 185	1	0	0.85	0.15
rep cluster 1533	2	0	1.7	0.3
rep cluster 1534	1	0	0.85	0.15
rep cluster 1535	2	0	1.7	0.3
rep cluster 1536	10	0	8.48	1.52
rep cluster 1537	3	0	2.54	0.46
rep cluster 154	3	0	2.54	0.46
rep cluster 1542	1	0	0.85	0.15
rep cluster 1546	2	0	1.7	0.3
rep cluster 1547	4	0	3.39	0.61
rep cluster 1548	1	0	0.85	0.15
rep cluster 155	27	3	25.43	4.57
rep cluster 155,rep cluster 1733	3	0	2.54	0.46
rep cluster 155,rep cluster 2439	16	0	13.56	2.44
rep cluster 1551	1	0	0.85	0.15
rep cluster 1552	1	0	0.85	0.15
rep cluster 1556	1	0	0.85	0.15
rep cluster 156	13	6	16.11	2.89
rep cluster 156,rep cluster 1991	2	0	1.7	0.3
rep cluster 1561	1	0	0.85	0.15
rep cluster 1566	7	0	5.93	1.07
rep cluster 1567	1	0	0.85	0.15
rep cluster 1567,rep cluster 167	1	0	0.85	0.15
rep cluster 1567,rep cluster 172,rep cluster 731	1	0	0.85	0.15
rep cluster 1567,rep cluster 707,rep cluster 731	1	0	0.85	0.15
rep cluster 1567,rep cluster 731	16	0	13.56	2.44
rep cluster 1570	1	0	0.85	0.15
rep cluster 1573	1	0	0.85	0.15
rep cluster 1574	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1575	3	0	2.54	0.46
rep cluster 1575,rep cluster 164,rep cluster 1751	1	0	0.85	0.15
rep cluster 1575,rep cluster 1850,rep cluster 1869	1	0	0.85	0.15
rep cluster 1575,rep cluster 1869	1	0	0.85	0.15
rep cluster 1578	1	0	0.85	0.15
rep cluster 158	1	0	0.85	0.15
rep cluster 1582	1	0	0.85	0.15
rep cluster 159	1	0	0.85	0.15
rep cluster 1590	4	0	3.39	0.61
rep cluster 1592	1	0	0.85	0.15
rep cluster 1593	2	0	1.7	0.3
rep cluster 1596	4	0	3.39	0.61
rep cluster 1598	1	0	0.85	0.15
rep cluster 1599	3	0	2.54	0.46
rep cluster 16	4	0	3.39	0.61
rep cluster 160	3	0	2.54	0.46
rep cluster 160,rep cluster 392	2	0	1.7	0.3
rep cluster 1601	3	0	2.54	0.46
rep cluster 1603	2	0	1.7	0.3
rep cluster 1604	1	0	0.85	0.15
rep cluster 1605	1	0	0.85	0.15
rep cluster 1607	17	0	14.41	2.59
rep cluster 1607,rep cluster 167,rep cluster 707	1	0	0.85	0.15
rep cluster 1607,rep cluster 556	1	0	0.85	0.15
rep cluster 1607,rep cluster 707	4	0	3.39	0.61
rep cluster 1607,rep cluster 731	1	0	0.85	0.15
rep cluster 1607,rep cluster 778	2	0	1.7	0.3
rep cluster 1608	1	0	0.85	0.15
rep cluster 1609	1	0	0.85	0.15
rep cluster 1610	4	0	3.39	0.61
rep cluster 1613	4	0	3.39	0.61
rep cluster 1614	1	0	0.85	0.15
rep cluster 1617	8	0	6.78	1.22
rep cluster 1617,rep cluster 736	1	0	0.85	0.15
rep cluster 1618	2	0	1.7	0.3
rep cluster 1618,rep cluster 1640	2	0	1.7	0.3
rep cluster 1621	31	0	26.28	4.72
rep cluster 1621,rep cluster 1857	1	0	0.85	0.15
rep cluster 1622	6	0	5.09	0.91
rep cluster 1622,rep cluster 317	1	0	0.85	0.15
rep cluster 1625	14	0	11.87	2.13
rep cluster 1626	6	0	5.09	0.91
rep cluster 1628	11	0	9.33	1.67
rep cluster 1629	12	0	10.17	1.83
rep cluster 163	92	22	96.64	17.36
rep cluster 163,rep cluster 1720	1	0	0.85	0.15
rep cluster 163,rep cluster 1721	2	0	1.7	0.3
rep cluster 163,rep cluster 1721,rep cluster 459	1	0	0.85	0.15
rep cluster 163,rep cluster 459	10	0	8.48	1.52
rep cluster 163,rep cluster 757	1	0	0.85	0.15
rep cluster 1631	17	0	14.41	2.59
rep cluster 1631,rep cluster 279	1	0	0.85	0.15
rep cluster 1634	1	0	0.85	0.15
rep cluster 1636	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1637	1	0	0.85	0.15
rep cluster 1639	1	0	0.85	0.15
rep cluster 164	28	6	28.82	5.18
rep cluster 164,rep cluster 1698,rep cluster 1751	1	0	0.85	0.15
rep cluster 164,rep cluster 1698,rep cluster 1751,rep cluster 1912	1	0	0.85	0.15
rep cluster 164,rep cluster 1698,rep cluster 1912	1	0	0.85	0.15
rep cluster 164,rep cluster 1751	9	0	7.63	1.37
rep cluster 1642	3	0	2.54	0.46
rep cluster 1645	2	0	1.7	0.3
rep cluster 1645,rep cluster 1721	1	0	0.85	0.15
rep cluster 1648	2	0	1.7	0.3
rep cluster 1652,rep cluster 841	1	0	0.85	0.15
rep cluster 1656	56	0	47.47	8.53
rep cluster 1656,rep cluster 1903	2	0	1.7	0.3
rep cluster 1657	2	2	3.39	0.61
rep cluster 1657,rep cluster 776	1	0	0.85	0.15
rep cluster 166	1	0	0.85	0.15
rep cluster 1662	19	0	16.11	2.89
rep cluster 1664	1	0	0.85	0.15
rep cluster 1664,rep cluster 2002	1	0	0.85	0.15
rep cluster 1665	7	4	9.33	1.67
rep cluster 167	79	0	66.97	12.03
rep cluster 167,rep cluster 1948	12	0	10.17	1.83
rep cluster 167,rep cluster 707	23	0	19.5	3.5
rep cluster 167,rep cluster 707,rep cluster 731	2	0	1.7	0.3
rep cluster 167,rep cluster 731	3	0	2.54	0.46
rep cluster 1671	3	2	4.24	0.76
rep cluster 1673	1	0	0.85	0.15
rep cluster 1674	1	0	0.85	0.15
rep cluster 1675	9	0	7.63	1.37
rep cluster 1676	2	0	1.7	0.3
rep cluster 1678	3	0	2.54	0.46
rep cluster 1678,rep cluster 776	1	0	0.85	0.15
rep cluster 1679	1	0	0.85	0.15
rep cluster 168	1	0	0.85	0.15
rep cluster 1685	31	0	26.28	4.72
rep cluster 1688	1	0	0.85	0.15
rep cluster 169	5	0	4.24	0.76
rep cluster 169,rep cluster 1721	27	0	22.89	4.11
rep cluster 169,rep cluster 1721,rep cluster 484	1	0	0.85	0.15
rep cluster 1693	1	0	0.85	0.15
rep cluster 1695	1	0	0.85	0.15
rep cluster 1696	1	0	0.85	0.15
rep cluster 1698	8	3	9.33	1.67
rep cluster 1698,rep cluster 1912	3	0	2.54	0.46
rep cluster 1698,rep cluster 845	2	0	1.7	0.3
rep cluster 170	2	0	1.7	0.3
rep cluster 1701	3	0	2.54	0.46
rep cluster 1702	19	12	26.28	4.72
rep cluster 1704	86	8	79.69	14.31
rep cluster 1705	3	0	2.54	0.46
rep cluster 1706	3	0	2.54	0.46
rep cluster 1707	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1710	3	0	2.54	0.46
rep cluster 1712	1	0	0.85	0.15
rep cluster 1713	32	0	27.13	4.87
rep cluster 1713,rep cluster 427	1	0	0.85	0.15
rep cluster 1714	1	0	0.85	0.15
rep cluster 1717	3	0	2.54	0.46
rep cluster 1719	1	0	0.85	0.15
rep cluster 172	21	3	20.35	3.65
rep cluster 172,rep cluster 707	3	0	2.54	0.46
rep cluster 172,rep cluster 707,rep cluster 731	1	0	0.85	0.15
rep cluster 172,rep cluster 731	1	0	0.85	0.15
rep cluster 1720	1	0	0.85	0.15
rep cluster 1720,rep cluster 1869	1	0	0.85	0.15
rep cluster 1721	25	0	21.19	3.81
rep cluster 1721,rep cluster 1767	2	0	1.7	0.3
rep cluster 1721,rep cluster 1869	2	0	1.7	0.3
rep cluster 1721,rep cluster 459,rep cluster 920	2	0	1.7	0.3
rep cluster 1721,rep cluster 484	16	0	13.56	2.44
rep cluster 1721,rep cluster 484,rep cluster 899	1	0	0.85	0.15
rep cluster 1721,rep cluster 887	1	0	0.85	0.15
rep cluster 1722	1	0	0.85	0.15
rep cluster 1722,rep cluster 731	1	0	0.85	0.15
rep cluster 1723	3	0	2.54	0.46
rep cluster 1725	1	0	0.85	0.15
rep cluster 1729	1	0	0.85	0.15
rep cluster 1730	3	0	2.54	0.46
rep cluster 1731	1	0	0.85	0.15
rep cluster 1732	3	0	2.54	0.46
rep cluster 1733	8	0	6.78	1.22
rep cluster 1733,rep cluster 2100	10	0	8.48	1.52
rep cluster 1733,rep cluster 2100,rep cluster 2176	14	0	11.87	2.13
rep cluster 1733,rep cluster 2100,rep cluster 2176,rep cluster 765	1	0	0.85	0.15
rep cluster 1733,rep cluster 2214	48	86	113.6	20.4
rep cluster 1733,rep cluster 2214,rep cluster 2439	1	0	0.85	0.15
rep cluster 1734	1	0	0.85	0.15
rep cluster 1735	2	0	1.7	0.3
rep cluster 1737	1	0	0.85	0.15
rep cluster 1738	8	0	6.78	1.22
rep cluster 1738,rep cluster 1913	1	0	0.85	0.15
rep cluster 1738,rep cluster 920	1	0	0.85	0.15
rep cluster 1739	6	0	5.09	0.91
rep cluster 1740	4	0	3.39	0.61
rep cluster 1741	1	0	0.85	0.15
rep cluster 1742	55	9	54.25	9.75
rep cluster 1742,rep cluster 183	1	0	0.85	0.15
rep cluster 1742,rep cluster 185	2	0	1.7	0.3
rep cluster 1742,rep cluster 185,rep cluster 2146	1	0	0.85	0.15
rep cluster 1744	4	0	3.39	0.61
rep cluster 1747	1	0	0.85	0.15
rep cluster 1749	1	0	0.85	0.15
rep cluster 1756	7	2	7.63	1.37
rep cluster 1760	10	0	8.48	1.52
rep cluster 1763	26	0	22.04	3.96

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1763,rep cluster 185	10	0	8.48	1.52
rep cluster 1763,rep cluster 185,rep cluster 2351	2	0	1.7	0.3
rep cluster 1763,rep cluster 185,rep cluster 889	2	0	1.7	0.3
rep cluster 1763,rep cluster 185,rep cluster 889,rep cluster 935	3	0	2.54	0.46
rep cluster 1763,rep cluster 2146,rep cluster 889	1	0	0.85	0.15
rep cluster 1763,rep cluster 2300	1	0	0.85	0.15
rep cluster 1763,rep cluster 889	25	0	21.19	3.81
rep cluster 1763,rep cluster 889,rep cluster 935	1	0	0.85	0.15
rep cluster 1763,rep cluster 889,rep cluster 943	1	0	0.85	0.15
rep cluster 1763,rep cluster 893	2	0	1.7	0.3
rep cluster 1767	6	0	5.09	0.91
rep cluster 1769	1	0	0.85	0.15
rep cluster 177	2	0	1.7	0.3
rep cluster 1775	3	3	5.09	0.91
rep cluster 1776	1	0	0.85	0.15
rep cluster 1776,rep cluster 215	1	0	0.85	0.15
rep cluster 1776,rep cluster 946	1	0	0.85	0.15
rep cluster 1777	2	0	1.7	0.3
rep cluster 1778	141	125	225.5	40.5
rep cluster 1779	2	0	1.7	0.3
rep cluster 1780	6	0	5.09	0.91
rep cluster 1787	5	0	4.24	0.76
rep cluster 179	2	0	1.7	0.3
rep cluster 1791	4	0	3.39	0.61
rep cluster 1798	2	0	1.7	0.3
rep cluster 18	1	0	0.85	0.15
rep cluster 180	116	0	98.34	17.66
rep cluster 180,rep cluster 185	1	0	0.85	0.15
rep cluster 180,rep cluster 185,rep cluster 2301	1	0	0.85	0.15
rep cluster 180,rep cluster 2101	1	0	0.85	0.15
rep cluster 180,rep cluster 570	1	0	0.85	0.15
rep cluster 180,rep cluster 980	1	0	0.85	0.15
rep cluster 180,rep cluster 992	4	0	3.39	0.61
rep cluster 1800	2	0	1.7	0.3
rep cluster 1801	1	0	0.85	0.15
rep cluster 1804	8	0	6.78	1.22
rep cluster 1805	1	0	0.85	0.15
rep cluster 1806	4	0	3.39	0.61
rep cluster 1806,rep cluster 771	3	0	2.54	0.46
rep cluster 1809	1	0	0.85	0.15
rep cluster 181	4	0	3.39	0.61
rep cluster 1814	3	0	2.54	0.46
rep cluster 1821	12	0	10.17	1.83
rep cluster 1821,rep cluster 1950	1	0	0.85	0.15
rep cluster 1825	5	0	4.24	0.76
rep cluster 1826	1	0	0.85	0.15
rep cluster 1827	10	0	8.48	1.52
rep cluster 183	32	0	27.13	4.87
rep cluster 183,rep cluster 185	9	4	11.02	1.98
rep cluster 183,rep cluster 185,rep cluster 2146	1	0	0.85	0.15
rep cluster 183,rep cluster 1891	2	0	1.7	0.3
rep cluster 183,rep cluster 1968	1	0	0.85	0.15
rep cluster 183,rep cluster 570	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 183,rep cluster 889	1	0	0.85	0.15
rep cluster 1830	13	0	11.02	1.98
rep cluster 1830,rep cluster 1895	1	0	0.85	0.15
rep cluster 1832	17	0	14.41	2.59
rep cluster 1833	1	2	2.54	0.46
rep cluster 1836	23	0	19.5	3.5
rep cluster 1836,rep cluster 1850,rep cluster 1958	1	0	0.85	0.15
rep cluster 1837	1	0	0.85	0.15
rep cluster 1838	1	0	0.85	0.15
rep cluster 1840	1	0	0.85	0.15
rep cluster 1848	6	0	5.09	0.91
rep cluster 1848,rep cluster 1895	7	0	5.93	1.07
rep cluster 185	176	33	177.18	31.82
rep cluster 185,rep cluster 1968	1	0	0.85	0.15
rep cluster 185,rep cluster 2146	2	0	1.7	0.3
rep cluster 185,rep cluster 2301	1	0	0.85	0.15
rep cluster 185,rep cluster 2351	13	0	11.02	1.98
rep cluster 185,rep cluster 2351,rep cluster 943	13	0	11.02	1.98
rep cluster 185,rep cluster 282	3	0	2.54	0.46
rep cluster 185,rep cluster 556	1	0	0.85	0.15
rep cluster 185,rep cluster 589	2	0	1.7	0.3
rep cluster 185,rep cluster 889	5	0	4.24	0.76
rep cluster 185,rep cluster 893	1	0	0.85	0.15
rep cluster 185,rep cluster 935	2	0	1.7	0.3
rep cluster 185,rep cluster 943	8	0	6.78	1.22
rep cluster 185,rep cluster 992	3	0	2.54	0.46
rep cluster 1850	12	0	10.17	1.83
rep cluster 1852	3	0	2.54	0.46
rep cluster 1855	6	0	5.09	0.91
rep cluster 1857	3	0	2.54	0.46
rep cluster 1858	1	0	0.85	0.15
rep cluster 1861	1	0	0.85	0.15
rep cluster 1863	4	0	3.39	0.61
rep cluster 1863,rep cluster 1958	1	0	0.85	0.15
rep cluster 1864	4	0	3.39	0.61
rep cluster 1868	1	0	0.85	0.15
rep cluster 1868,rep cluster 870	1	0	0.85	0.15
rep cluster 1869	31	2	27.98	5.02
rep cluster 1869,rep cluster 459	4	0	3.39	0.61
rep cluster 187	5	0	4.24	0.76
rep cluster 1871	5	0	4.24	0.76
rep cluster 1874	3	0	2.54	0.46
rep cluster 1874,rep cluster 557	4	0	3.39	0.61
rep cluster 1875	8	0	6.78	1.22
rep cluster 1875,rep cluster 337	2	0	1.7	0.3
rep cluster 1879,rep cluster 1910	1	0	0.85	0.15
rep cluster 1879,rep cluster 1910,rep cluster 1975,rep cluster 40	2	0	1.7	0.3
rep cluster 1880	14	2	13.56	2.44
rep cluster 1880,rep cluster 2119	1	0	0.85	0.15
rep cluster 1880,rep cluster 707	1	0	0.85	0.15
rep cluster 1887	1	0	0.85	0.15
rep cluster 1892	5	0	4.24	0.76
rep cluster 1895	12	2	11.87	2.13

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1895,rep cluster 1912	3	0	2.54	0.46
rep cluster 1897	1	0	0.85	0.15
rep cluster 190	5	2	5.93	1.07
rep cluster 1902	4	0	3.39	0.61
rep cluster 1903	1	0	0.85	0.15
rep cluster 1905	1	0	0.85	0.15
rep cluster 1906	30	0	25.43	4.57
rep cluster 1907	3	2	4.24	0.76
rep cluster 1908	2	0	1.7	0.3
rep cluster 1909	10	0	8.48	1.52
rep cluster 1909,rep cluster 1958	1	0	0.85	0.15
rep cluster 1909,rep cluster 1966	1	0	0.85	0.15
rep cluster 1910	1	0	0.85	0.15
rep cluster 1911	1	0	0.85	0.15
rep cluster 1912	29	0	24.58	4.42
rep cluster 1913	1	0	0.85	0.15
rep cluster 1914	2	0	1.7	0.3
rep cluster 1919	4	0	3.39	0.61
rep cluster 1925	2	5	5.93	1.07
rep cluster 1927	3	0	2.54	0.46
rep cluster 1929	1	0	0.85	0.15
rep cluster 193	7	0	5.93	1.07
rep cluster 1930	1	0	0.85	0.15
rep cluster 1931	0	4	3.39	0.61
rep cluster 1931,rep cluster 1966	0	3	2.54	0.46
rep cluster 1934	3	0	2.54	0.46
rep cluster 1938	3	0	2.54	0.46
rep cluster 1943	3	0	2.54	0.46
rep cluster 1944	2	0	1.7	0.3
rep cluster 1945,rep cluster 2045	2	0	1.7	0.3
rep cluster 1947	3	0	2.54	0.46
rep cluster 1948	24	0	20.35	3.65
rep cluster 1948,rep cluster 2119	5	0	4.24	0.76
rep cluster 1949	1	0	0.85	0.15
rep cluster 195	8	0	6.78	1.22
rep cluster 1950	4	0	3.39	0.61
rep cluster 1951	2	0	1.7	0.3
rep cluster 1952	3	0	2.54	0.46
rep cluster 1953	3	0	2.54	0.46
rep cluster 1953,rep cluster 589	0	2	1.7	0.3
rep cluster 1956	1	0	0.85	0.15
rep cluster 1958	2	0	1.7	0.3
rep cluster 196	12	0	10.17	1.83
rep cluster 196,rep cluster 707	2	0	1.7	0.3
rep cluster 1964	2	0	1.7	0.3
rep cluster 1965	2	0	1.7	0.3
rep cluster 1966	2	0	1.7	0.3
rep cluster 1967	3	0	2.54	0.46
rep cluster 1967,rep cluster 2310	1	0	0.85	0.15
rep cluster 1968,rep cluster 889	1	0	0.85	0.15
rep cluster 197	7	0	5.93	1.07
rep cluster 197,rep cluster 707	1	0	0.85	0.15
rep cluster 197,rep cluster 731	4	0	3.39	0.61
rep cluster 1970	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 1971	17	0	14.41	2.59
rep cluster 1974	8	0	6.78	1.22
rep cluster 1975	1	0	0.85	0.15
rep cluster 1977	4	0	3.39	0.61
rep cluster 198	1	0	0.85	0.15
rep cluster 1980	8	0	6.78	1.22
rep cluster 1981	1	0	0.85	0.15
rep cluster 1985	1	0	0.85	0.15
rep cluster 1986	2	0	1.7	0.3
rep cluster 1988	2	0	1.7	0.3
rep cluster 199	7	0	5.93	1.07
rep cluster 1991	46	0	39	7
rep cluster 1992	3	2	4.24	0.76
rep cluster 1993	5	0	4.24	0.76
rep cluster 2002	3	0	2.54	0.46
rep cluster 2005	4	0	3.39	0.61
rep cluster 2006	13	0	11.02	1.98
rep cluster 2007	1	0	0.85	0.15
rep cluster 2011	1	0	0.85	0.15
rep cluster 2013	7	0	5.93	1.07
rep cluster 2014	1	0	0.85	0.15
rep cluster 2015	1	0	0.85	0.15
rep cluster 2020	13	0	11.02	1.98
rep cluster 2021	8	0	6.78	1.22
rep cluster 2021,rep cluster 2022	1	0	0.85	0.15
rep cluster 2022	17	4	17.8	3.2
rep cluster 2022,rep cluster 2373	1	0	0.85	0.15
rep cluster 2022,rep cluster 658	1	0	0.85	0.15
rep cluster 2023	5	3	6.78	1.22
rep cluster 2025	22	0	18.65	3.35
rep cluster 203	0	2	1.7	0.3
rep cluster 2037	15	0	12.72	2.28
rep cluster 204	11	0	9.33	1.67
rep cluster 2042	10	2	10.17	1.83
rep cluster 2042,rep cluster 429	1	0	0.85	0.15
rep cluster 2045	3	0	2.54	0.46
rep cluster 2046	4	0	3.39	0.61
rep cluster 2051	1	0	0.85	0.15
rep cluster 2061	2	0	1.7	0.3
rep cluster 2062	1	0	0.85	0.15
rep cluster 207	1	0	0.85	0.15
rep cluster 2077	1	0	0.85	0.15
rep cluster 2078	3	0	2.54	0.46
rep cluster 208	2	0	1.7	0.3
rep cluster 2082	2	0	1.7	0.3
rep cluster 209	4	0	3.39	0.61
rep cluster 210	1	0	0.85	0.15
rep cluster 2101	13	0	11.02	1.98
rep cluster 2101,rep cluster 976,rep cluster 980	1	0	0.85	0.15
rep cluster 2104	12	0	10.17	1.83
rep cluster 2104,rep cluster 282	1	0	0.85	0.15
rep cluster 2104,rep cluster 893	1	0	0.85	0.15
rep cluster 2104,rep cluster 910	1	0	0.85	0.15
rep cluster 2109	4	0	3.39	0.61

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 211	4	0	3.39	0.61
rep cluster 211,rep cluster 765	1	0	0.85	0.15
rep cluster 2117	5	0	4.24	0.76
rep cluster 2119	34	7	34.76	6.24
rep cluster 2119,rep cluster 707	2	0	1.7	0.3
rep cluster 2119,rep cluster 731	1	0	0.85	0.15
rep cluster 2120	5	0	4.24	0.76
rep cluster 2120,rep cluster 2214,rep cluster 418	1	0	0.85	0.15
rep cluster 2127	8	0	6.78	1.22
rep cluster 213	1	0	0.85	0.15
rep cluster 2131	84	12	81.38	14.62
rep cluster 2131,rep cluster 2361	1	0	0.85	0.15
rep cluster 2131,rep cluster 2370	1	0	0.85	0.15
rep cluster 2132	1	0	0.85	0.15
rep cluster 2133	1	0	0.85	0.15
rep cluster 2134	10	0	8.48	1.52
rep cluster 214	11	0	9.33	1.67
rep cluster 2140	1	0	0.85	0.15
rep cluster 2141	16	0	13.56	2.44
rep cluster 2141,rep cluster 2150	0	1	0.85	0.15
rep cluster 2146	11	5	13.56	2.44
rep cluster 2146,rep cluster 889	3	0	2.54	0.46
rep cluster 2147	1	0	0.85	0.15
rep cluster 215	2	0	1.7	0.3
rep cluster 2150	22	15	31.37	5.63
rep cluster 216	84	0	71.21	12.79
rep cluster 216,rep cluster 792	1	0	0.85	0.15
rep cluster 2164	2	0	1.7	0.3
rep cluster 2165	6	0	5.09	0.91
rep cluster 2167	1	0	0.85	0.15
rep cluster 2169	1	0	0.85	0.15
rep cluster 2172	1	0	0.85	0.15
rep cluster 2175	11	10	17.8	3.2
rep cluster 2176	1	0	0.85	0.15
rep cluster 2183	11	0	9.33	1.67
rep cluster 2183,rep cluster 2335,rep cluster 2358	1	0	0.85	0.15
rep cluster 2183,rep cluster 658	1	0	0.85	0.15
rep cluster 2187	3	0	2.54	0.46
rep cluster 219	2	0	1.7	0.3
rep cluster 2194	2	0	1.7	0.3
rep cluster 2194,rep cluster 2222	1	0	0.85	0.15
rep cluster 220	3	0	2.54	0.46
rep cluster 2201	2	0	1.7	0.3
rep cluster 2203	2	0	1.7	0.3
rep cluster 2214,rep cluster 2215,rep cluster 2222	2	0	1.7	0.3
rep cluster 2215	4	0	3.39	0.61
rep cluster 2218	8	0	6.78	1.22
rep cluster 2221	8	0	6.78	1.22
rep cluster 2222	25	0	21.19	3.81
rep cluster 2222,rep cluster 418	1	0	0.85	0.15
rep cluster 2229	3	0	2.54	0.46
rep cluster 2232	3	0	2.54	0.46
rep cluster 2234	1	0	0.85	0.15
rep cluster 2238	4	0	3.39	0.61

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 2238,rep cluster 402	1	0	0.85	0.15
rep cluster 224	5	0	4.24	0.76
rep cluster 2241	1	0	0.85	0.15
rep cluster 2244	3	0	2.54	0.46
rep cluster 2244,rep cluster 2313	1	0	0.85	0.15
rep cluster 2246	6	0	5.09	0.91
rep cluster 225	2	0	1.7	0.3
rep cluster 2252	2	0	1.7	0.3
rep cluster 2254	2	0	1.7	0.3
rep cluster 2256	1	0	0.85	0.15
rep cluster 2257	1	0	0.85	0.15
rep cluster 2263	1	0	0.85	0.15
rep cluster 2264	1	0	0.85	0.15
rep cluster 2268	5	0	4.24	0.76
rep cluster 2272	5	0	4.24	0.76
rep cluster 2273	8	0	6.78	1.22
rep cluster 2273,rep cluster 904	1	0	0.85	0.15
rep cluster 2281	1	0	0.85	0.15
rep cluster 2282	2	0	1.7	0.3
rep cluster 2287	1	0	0.85	0.15
rep cluster 2288	1	0	0.85	0.15
rep cluster 229	12	0	10.17	1.83
rep cluster 2294	1	0	0.85	0.15
rep cluster 2295	2	0	1.7	0.3
rep cluster 2298	1	0	0.85	0.15
rep cluster 2299	2	0	1.7	0.3
rep cluster 2300	5	2	5.93	1.07
rep cluster 2301	22	38	50.86	9.14
rep cluster 2301,rep cluster 980	3	0	2.54	0.46
rep cluster 2302	9	0	7.63	1.37
rep cluster 2307	2	0	1.7	0.3
rep cluster 231	5	0	4.24	0.76
rep cluster 2310	2	0	1.7	0.3
rep cluster 2313	2	0	1.7	0.3
rep cluster 2314	10	0	8.48	1.52
rep cluster 2315	4	0	3.39	0.61
rep cluster 2319	6	0	5.09	0.91
rep cluster 2324	14	11	21.19	3.81
rep cluster 2327	30	7	31.37	5.63
rep cluster 2327,rep cluster 2358	3	0	2.54	0.46
rep cluster 2327,rep cluster 2373	2	0	1.7	0.3
rep cluster 2327,rep cluster 2374	1	0	0.85	0.15
rep cluster 2329	1	0	0.85	0.15
rep cluster 2335	367	644	857.06	153.94
rep cluster 2335,rep cluster 2350	3	0	2.54	0.46
rep cluster 2335,rep cluster 2358	15	52	56.8	10.2
rep cluster 2335,rep cluster 2373	5	15	16.95	3.05
rep cluster 2335,rep cluster 2374	0	1	0.85	0.15
rep cluster 2335,rep cluster 2392	2	1	2.54	0.46
rep cluster 2335,rep cluster 2405	1	1	1.7	0.3
rep cluster 2335,rep cluster 2406	0	2	1.7	0.3
rep cluster 234	5	0	4.24	0.76
rep cluster 2341,rep cluster 418	12	0	10.17	1.83
rep cluster 235	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 2350	709	354	901.14	161.86
rep cluster 2350,rep cluster 2358	7	2	7.63	1.37
rep cluster 2350,rep cluster 2358,rep cluster 2370	1	0	0.85	0.15
rep cluster 2350,rep cluster 2370	36	19	46.63	8.37
rep cluster 2350,rep cluster 2370,rep cluster 2401	1	0	0.85	0.15
rep cluster 2350,rep cluster 2373	2	0	1.7	0.3
rep cluster 2350,rep cluster 2401	2	0	1.7	0.3
rep cluster 2350,rep cluster 2486	1	0	0.85	0.15
rep cluster 2350,rep cluster 352	2	0	1.7	0.3
rep cluster 2351	18	10	23.74	4.26
rep cluster 2351,rep cluster 889	1	0	0.85	0.15
rep cluster 2351,rep cluster 893	1	0	0.85	0.15
rep cluster 2351,rep cluster 943	1	0	0.85	0.15
rep cluster 2358	124	52	149.2	26.8
rep cluster 2358,rep cluster 2370	1	0	0.85	0.15
rep cluster 2358,rep cluster 2373	7	8	12.72	2.28
rep cluster 2359	12	0	10.17	1.83
rep cluster 236	2	0	1.7	0.3
rep cluster 2360	2	0	1.7	0.3
rep cluster 237	31	5	30.52	5.48
rep cluster 2370	212	17	194.13	34.87
rep cluster 2371	1	0	0.85	0.15
rep cluster 2373	113	37	127.16	22.84
rep cluster 2373,rep cluster 2374	8	0	6.78	1.22
rep cluster 2373,rep cluster 658	1	0	0.85	0.15
rep cluster 2374	7	4	9.33	1.67
rep cluster 239	1	0	0.85	0.15
rep cluster 2392	15	0	12.72	2.28
rep cluster 2398	1	0	0.85	0.15
rep cluster 240	3	0	2.54	0.46
rep cluster 2401	57	81	116.99	21.01
rep cluster 2405	1	1	1.7	0.3
rep cluster 2406	1	0	0.85	0.15
rep cluster 2409	3	0	2.54	0.46
rep cluster 2409,rep cluster 910	7	0	5.93	1.07
rep cluster 241	1	0	0.85	0.15
rep cluster 2414	1	0	0.85	0.15
rep cluster 2419	2	0	1.7	0.3
rep cluster 242	1	0	0.85	0.15
rep cluster 2422	41	0	34.76	6.24
rep cluster 2422,rep cluster 554	2	0	1.7	0.3
rep cluster 2422,rep cluster 554,rep cluster 667	1	0	0.85	0.15
rep cluster 2422,rep cluster 664	2	0	1.7	0.3
rep cluster 2425	1	0	0.85	0.15
rep cluster 2427	1	0	0.85	0.15
rep cluster 2431	1	0	0.85	0.15
rep cluster 2439	1	0	0.85	0.15
rep cluster 244	1	0	0.85	0.15
rep cluster 2442	4	0	3.39	0.61
rep cluster 2446	37	0	31.37	5.63
rep cluster 2448	1	0	0.85	0.15
rep cluster 2455	1	0	0.85	0.15
rep cluster 2464	3	0	2.54	0.46
rep cluster 2466	1	2	2.54	0.46

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 2468	2	0	1.7	0.3
rep cluster 2469	20	0	16.95	3.05
rep cluster 2470	3	0	2.54	0.46
rep cluster 2472	1	0	0.85	0.15
rep cluster 2475	4	0	3.39	0.61
rep cluster 2479	1	0	0.85	0.15
rep cluster 2483	6	0	5.09	0.91
rep cluster 2484	3	0	2.54	0.46
rep cluster 2486	9	0	7.63	1.37
rep cluster 25	1	0	0.85	0.15
rep cluster 251	5	0	4.24	0.76
rep cluster 252	23	4	22.89	4.11
rep cluster 252,rep cluster 735	49	0	41.54	7.46
rep cluster 256	13	0	11.02	1.98
rep cluster 260	1	0	0.85	0.15
rep cluster 261	3	0	2.54	0.46
rep cluster 263	5	0	4.24	0.76
rep cluster 267	1	0	0.85	0.15
rep cluster 268	10	0	8.48	1.52
rep cluster 271	2	0	1.7	0.3
rep cluster 273	11	0	9.33	1.67
rep cluster 275	1	0	0.85	0.15
rep cluster 28	10	0	8.48	1.52
rep cluster 280	1	0	0.85	0.15
rep cluster 281	1	0	0.85	0.15
rep cluster 282	8	0	6.78	1.22
rep cluster 282,rep cluster 312	1	0	0.85	0.15
rep cluster 283	135	0	114.44	20.56
rep cluster 284	1	0	0.85	0.15
rep cluster 287	45	4	41.54	7.46
rep cluster 287,rep cluster 359	5	0	4.24	0.76
rep cluster 287,rep cluster 512	1	0	0.85	0.15
rep cluster 287,rep cluster 946	1	0	0.85	0.15
rep cluster 288	56	0	47.47	8.53
rep cluster 288,rep cluster 377	1	0	0.85	0.15
rep cluster 289	6	0	5.09	0.91
rep cluster 29	1	0	0.85	0.15
rep cluster 290	1	0	0.85	0.15
rep cluster 292	4	0	3.39	0.61
rep cluster 293	1	0	0.85	0.15
rep cluster 294	7	0	5.93	1.07
rep cluster 296	3	0	2.54	0.46
rep cluster 3	17	0	14.41	2.59
rep cluster 30	8	0	6.78	1.22
rep cluster 300	3	0	2.54	0.46
rep cluster 304	1	0	0.85	0.15
rep cluster 305	3	0	2.54	0.46
rep cluster 305,rep cluster 392	1	0	0.85	0.15
rep cluster 305,rep cluster 579	1	0	0.85	0.15
rep cluster 306	2	0	1.7	0.3
rep cluster 308	29	0	24.58	4.42
rep cluster 309	1	0	0.85	0.15
rep cluster 310	2	0	1.7	0.3
rep cluster 312	17	2	16.11	2.89

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 313	5	0	4.24	0.76
rep cluster 313,rep cluster 906	1	0	0.85	0.15
rep cluster 314	1	0	0.85	0.15
rep cluster 317	5	0	4.24	0.76
rep cluster 32	1	0	0.85	0.15
rep cluster 321	4	0	3.39	0.61
rep cluster 322	5	0	4.24	0.76
rep cluster 323	8	0	6.78	1.22
rep cluster 324	3	0	2.54	0.46
rep cluster 327	4	2	5.09	0.91
rep cluster 328	7	0	5.93	1.07
rep cluster 329	3	0	2.54	0.46
rep cluster 33	3	0	2.54	0.46
rep cluster 332	6	0	5.09	0.91
rep cluster 334	1	0	0.85	0.15
rep cluster 335	1	0	0.85	0.15
rep cluster 337	6	0	5.09	0.91
rep cluster 339	5	0	4.24	0.76
rep cluster 341	1	0	0.85	0.15
rep cluster 342	9	0	7.63	1.37
rep cluster 342,rep cluster 554	1	0	0.85	0.15
rep cluster 342,rep cluster 554,rep cluster 799	1	0	0.85	0.15
rep cluster 342,rep cluster 799	1	0	0.85	0.15
rep cluster 345	1	0	0.85	0.15
rep cluster 346	1	0	0.85	0.15
rep cluster 349	2	0	1.7	0.3
rep cluster 349,rep cluster 672	1	0	0.85	0.15
rep cluster 352	4	7	9.33	1.67
rep cluster 352,rep cluster 490	1	0	0.85	0.15
rep cluster 355	4	0	3.39	0.61
rep cluster 356	7	0	5.93	1.07
rep cluster 359	1	0	0.85	0.15
rep cluster 360	1	0	0.85	0.15
rep cluster 361	50	0	42.39	7.61
rep cluster 364	5	0	4.24	0.76
rep cluster 366	1	0	0.85	0.15
rep cluster 369	8	2	8.48	1.52
rep cluster 37	0	2	1.7	0.3
rep cluster 371	2	0	1.7	0.3
rep cluster 373	1	0	0.85	0.15
rep cluster 377	53	0	44.93	8.07
rep cluster 378	14	0	11.87	2.13
rep cluster 378,rep cluster 746	1	0	0.85	0.15
rep cluster 38	3	0	2.54	0.46
rep cluster 380	5	0	4.24	0.76
rep cluster 380,rep cluster 674	2	0	1.7	0.3
rep cluster 382	15	0	12.72	2.28
rep cluster 383	1	0	0.85	0.15
rep cluster 385	142	8	127.16	22.84
rep cluster 386	6	0	5.09	0.91
rep cluster 386,rep cluster 734	1	0	0.85	0.15
rep cluster 389	1	0	0.85	0.15
rep cluster 390	1	0	0.85	0.15
rep cluster 392	11	0	9.33	1.67

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 392,rep cluster 517	7	2	7.63	1.37
rep cluster 392,rep cluster 610	2	0	1.7	0.3
rep cluster 398	7	0	5.93	1.07
rep cluster 4	1	0	0.85	0.15
rep cluster 402	5	0	4.24	0.76
rep cluster 405	5	0	4.24	0.76
rep cluster 408	1	0	0.85	0.15
rep cluster 409	2	0	1.7	0.3
rep cluster 41	4	0	3.39	0.61
rep cluster 412	1	0	0.85	0.15
rep cluster 413	2	0	1.7	0.3
rep cluster 414	19	0	16.11	2.89
rep cluster 414,rep cluster 516	2	0	1.7	0.3
rep cluster 414,rep cluster 520	2	0	1.7	0.3
rep cluster 414,rep cluster 554	4	2	5.09	0.91
rep cluster 414,rep cluster 554,rep cluster 656	1	0	0.85	0.15
rep cluster 414,rep cluster 556	2	0	1.7	0.3
rep cluster 414,rep cluster 656,rep cluster 667	1	0	0.85	0.15
rep cluster 414,rep cluster 707	2	0	1.7	0.3
rep cluster 414,rep cluster 799	1	0	0.85	0.15
rep cluster 416	2	0	1.7	0.3
rep cluster 418	2	0	1.7	0.3
rep cluster 419	1	0	0.85	0.15
rep cluster 420	2	3	4.24	0.76
rep cluster 421	1	0	0.85	0.15
rep cluster 427	24	0	20.35	3.65
rep cluster 428	4	2	5.09	0.91
rep cluster 429	34	3	31.37	5.63
rep cluster 429,rep cluster 461	34	5	33.06	5.94
rep cluster 43	2	0	1.7	0.3
rep cluster 433	2	0	1.7	0.3
rep cluster 435	16	2	15.26	2.74
rep cluster 439	27	0	22.89	4.11
rep cluster 440	10	0	8.48	1.52
rep cluster 441	5	0	4.24	0.76
rep cluster 443	3	4	5.93	1.07
rep cluster 446	1	2	2.54	0.46
rep cluster 446,rep cluster 512,rep cluster 592	1	0	0.85	0.15
rep cluster 453	1	0	0.85	0.15
rep cluster 455	3	0	2.54	0.46
rep cluster 457	4	0	3.39	0.61
rep cluster 459	8	0	6.78	1.22
rep cluster 459,rep cluster 920	16	0	13.56	2.44
rep cluster 46	1	0	0.85	0.15
rep cluster 460	1	0	0.85	0.15
rep cluster 461	6	0	5.09	0.91
rep cluster 463	1	0	0.85	0.15
rep cluster 465	1	0	0.85	0.15
rep cluster 466	3	0	2.54	0.46
rep cluster 471	29	0	24.58	4.42
rep cluster 471,rep cluster 512	1	0	0.85	0.15
rep cluster 471,rep cluster 946	1	0	0.85	0.15
rep cluster 473	6	0	5.09	0.91
rep cluster 473,rep cluster 546	1	0	0.85	0.15

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 475	78	28	89.86	16.14
rep cluster 477	1	0	0.85	0.15
rep cluster 478	17	0	14.41	2.59
rep cluster 480	1	0	0.85	0.15
rep cluster 481	17	0	14.41	2.59
rep cluster 482	1	0	0.85	0.15
rep cluster 484	12	0	10.17	1.83
rep cluster 485	5	2	5.93	1.07
rep cluster 485,rep cluster 617	4	0	3.39	0.61
rep cluster 488	112	2	96.64	17.36
rep cluster 49	2	0	1.7	0.3
rep cluster 490	44	44	74.6	13.4
rep cluster 490,rep cluster 553	12	3	12.72	2.28
rep cluster 491	3	0	2.54	0.46
rep cluster 491,rep cluster 685	1	0	0.85	0.15
rep cluster 492	7	0	5.93	1.07
rep cluster 492,rep cluster 946	1	0	0.85	0.15
rep cluster 493	3	0	2.54	0.46
rep cluster 494	2	0	1.7	0.3
rep cluster 496	3	0	2.54	0.46
rep cluster 497	7	0	5.93	1.07
rep cluster 5	4	0	3.39	0.61
rep cluster 50	1	0	0.85	0.15
rep cluster 500	3	0	2.54	0.46
rep cluster 502	1	0	0.85	0.15
rep cluster 505	6	0	5.09	0.91
rep cluster 506	13	0	11.02	1.98
rep cluster 507	5	0	4.24	0.76
rep cluster 508	21	0	17.8	3.2
rep cluster 509	1	0	0.85	0.15
rep cluster 512	5	0	4.24	0.76
rep cluster 513	4	0	3.39	0.61
rep cluster 514	1	0	0.85	0.15
rep cluster 515	1	0	0.85	0.15
rep cluster 516	8	0	6.78	1.22
rep cluster 516,rep cluster 554	1	0	0.85	0.15
rep cluster 516,rep cluster 656	1	0	0.85	0.15
rep cluster 516,rep cluster 660	1	0	0.85	0.15
rep cluster 517	3	0	2.54	0.46
rep cluster 519	16	0	13.56	2.44
rep cluster 520	11	0	9.33	1.67
rep cluster 521	1	0	0.85	0.15
rep cluster 522	9	0	7.63	1.37
rep cluster 525	1	0	0.85	0.15
rep cluster 526	3	0	2.54	0.46
rep cluster 528	2	0	1.7	0.3
rep cluster 529	1	0	0.85	0.15
rep cluster 530	7	0	5.93	1.07
rep cluster 533	1	0	0.85	0.15
rep cluster 538	1	2	2.54	0.46
rep cluster 541	2	0	1.7	0.3
rep cluster 542	2	0	1.7	0.3
rep cluster 543	3	0	2.54	0.46
rep cluster 545	2	0	1.7	0.3

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 545,rep cluster 596	1	0	0.85	0.15
rep cluster 546	7	0	5.93	1.07
rep cluster 546,rep cluster 91	1	0	0.85	0.15
rep cluster 547	2	0	1.7	0.3
rep cluster 548	1	0	0.85	0.15
rep cluster 551	1	0	0.85	0.15
rep cluster 553	12	0	10.17	1.83
rep cluster 554	58	0	49.17	8.83
rep cluster 554,rep cluster 556	3	0	2.54	0.46
rep cluster 554,rep cluster 570,rep cluster 667	1	0	0.85	0.15
rep cluster 554,rep cluster 571	1	0	0.85	0.15
rep cluster 554,rep cluster 656	3	0	2.54	0.46
rep cluster 554,rep cluster 660	1	0	0.85	0.15
rep cluster 554,rep cluster 667	14	6	16.95	3.05
rep cluster 554,rep cluster 667,rep cluster 799	1	0	0.85	0.15
rep cluster 554,rep cluster 796	1	0	0.85	0.15
rep cluster 554,rep cluster 980	2	0	1.7	0.3
rep cluster 554,rep cluster 992	4	2	5.09	0.91
rep cluster 555	2	0	1.7	0.3
rep cluster 556	67	4	60.19	10.81
rep cluster 556,rep cluster 570	1	0	0.85	0.15
rep cluster 556,rep cluster 619	2	0	1.7	0.3
rep cluster 556,rep cluster 707	1	0	0.85	0.15
rep cluster 556,rep cluster 893	3	0	2.54	0.46
rep cluster 556,rep cluster 976	2	0	1.7	0.3
rep cluster 557	15	0	12.72	2.28
rep cluster 558	62	13	63.58	11.42
rep cluster 559	6	0	5.09	0.91
rep cluster 56	1	0	0.85	0.15
rep cluster 560	1	0	0.85	0.15
rep cluster 562	1	0	0.85	0.15
rep cluster 564	7	0	5.93	1.07
rep cluster 565	8	0	6.78	1.22
rep cluster 567	6	0	5.09	0.91
rep cluster 570	24	11	29.67	5.33
rep cluster 570,rep cluster 571	1	0	0.85	0.15
rep cluster 570,rep cluster 633	0	3	2.54	0.46
rep cluster 570,rep cluster 656	4	0	3.39	0.61
rep cluster 570,rep cluster 980	2	0	1.7	0.3
rep cluster 571	2	0	1.7	0.3
rep cluster 571,rep cluster 800	1	0	0.85	0.15
rep cluster 573	19	4	19.5	3.5
rep cluster 577	2	0	1.7	0.3
rep cluster 578	23	9	27.13	4.87
rep cluster 579	66	0	55.95	10.05
rep cluster 579,rep cluster 610	2	0	1.7	0.3
rep cluster 580	1	0	0.85	0.15
rep cluster 581	1	0	0.85	0.15
rep cluster 583	1	0	0.85	0.15
rep cluster 586	1	0	0.85	0.15
rep cluster 589	78	0	66.12	11.88
rep cluster 590	8	0	6.78	1.22
rep cluster 591	4	0	3.39	0.61
rep cluster 592	13	3	13.56	2.44

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 593	13	0	11.02	1.98
rep cluster 597	2	0	1.7	0.3
rep cluster 599	6	0	5.09	0.91
rep cluster 60	1	0	0.85	0.15
rep cluster 600	4	0	3.39	0.61
rep cluster 601	1	0	0.85	0.15
rep cluster 602	6	0	5.09	0.91
rep cluster 604	1	0	0.85	0.15
rep cluster 605	7	0	5.93	1.07
rep cluster 607	1	0	0.85	0.15
rep cluster 608	2	0	1.7	0.3
rep cluster 61	1	0	0.85	0.15
rep cluster 610	11	0	9.33	1.67
rep cluster 611	1	0	0.85	0.15
rep cluster 613	9	0	7.63	1.37
rep cluster 617	35	17	44.08	7.92
rep cluster 618	23	0	19.5	3.5
rep cluster 619	7	0	5.93	1.07
rep cluster 622	7	0	5.93	1.07
rep cluster 625	1	0	0.85	0.15
rep cluster 626	4	0	3.39	0.61
rep cluster 628	5	0	4.24	0.76
rep cluster 633	11	0	9.33	1.67
rep cluster 634	2	0	1.7	0.3
rep cluster 635	2	0	1.7	0.3
rep cluster 636	18	0	15.26	2.74
rep cluster 637	2	0	1.7	0.3
rep cluster 639	53	8	51.71	9.29
rep cluster 639,rep cluster 718	1	0	0.85	0.15
rep cluster 64	10	0	8.48	1.52
rep cluster 640	3	0	2.54	0.46
rep cluster 641	3	0	2.54	0.46
rep cluster 647	13	0	11.02	1.98
rep cluster 648	2	0	1.7	0.3
rep cluster 650	23	0	19.5	3.5
rep cluster 655	2	0	1.7	0.3
rep cluster 656	4	0	3.39	0.61
rep cluster 657	1	0	0.85	0.15
rep cluster 658	4	4	6.78	1.22
rep cluster 660	5	0	4.24	0.76
rep cluster 661	2	0	1.7	0.3
rep cluster 662	1	0	0.85	0.15
rep cluster 663	20	9	24.58	4.42
rep cluster 664	8	0	6.78	1.22
rep cluster 667	29	0	24.58	4.42
rep cluster 669	2	0	1.7	0.3
rep cluster 67	1	0	0.85	0.15
rep cluster 673	3	0	2.54	0.46
rep cluster 674	1	0	0.85	0.15
rep cluster 675	1	0	0.85	0.15
rep cluster 678	3	0	2.54	0.46
rep cluster 680	2	0	1.7	0.3
rep cluster 681	5	0	4.24	0.76
rep cluster 684	2	0	1.7	0.3

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 685	3	0	2.54	0.46
rep cluster 687	2	0	1.7	0.3
rep cluster 688	18	0	15.26	2.74
rep cluster 689	4	3	5.93	1.07
rep cluster 694	1	0	0.85	0.15
rep cluster 698	1	0	0.85	0.15
rep cluster 703	1	0	0.85	0.15
rep cluster 704	6	0	5.09	0.91
rep cluster 707	217	2	185.65	33.35
rep cluster 707,rep cluster 731	4	0	3.39	0.61
rep cluster 707,rep cluster 816	1	0	0.85	0.15
rep cluster 71	9	0	7.63	1.37
rep cluster 711	5	0	4.24	0.76
rep cluster 715	17	0	14.41	2.59
rep cluster 718	104	0	88.16	15.84
rep cluster 719	1	0	0.85	0.15
rep cluster 721	1	0	0.85	0.15
rep cluster 722	1	0	0.85	0.15
rep cluster 723	3	0	2.54	0.46
rep cluster 724	7	0	5.93	1.07
rep cluster 725	2	0	1.7	0.3
rep cluster 728	5	0	4.24	0.76
rep cluster 731	104	0	88.16	15.84
rep cluster 734	155	18	146.66	26.34
rep cluster 735	14	0	11.87	2.13
rep cluster 736	13	0	11.02	1.98
rep cluster 736,rep cluster 754	1	0	0.85	0.15
rep cluster 737	38	0	32.21	5.79
rep cluster 741	1	0	0.85	0.15
rep cluster 746	5	0	4.24	0.76
rep cluster 748	2	0	1.7	0.3
rep cluster 749	1	0	0.85	0.15
rep cluster 750	1	0	0.85	0.15
rep cluster 754	4	0	3.39	0.61
rep cluster 755	3	0	2.54	0.46
rep cluster 757	13	0	11.02	1.98
rep cluster 759	5	0	4.24	0.76
rep cluster 762	8	0	6.78	1.22
rep cluster 763	10	0	8.48	1.52
rep cluster 765	46	0	39	7
rep cluster 766	2	0	1.7	0.3
rep cluster 768	7	0	5.93	1.07
rep cluster 77	2	0	1.7	0.3
rep cluster 771	6	0	5.09	0.91
rep cluster 776	4	0	3.39	0.61
rep cluster 778	36	0	30.52	5.48
rep cluster 779	2	0	1.7	0.3
rep cluster 780	1	0	0.85	0.15
rep cluster 782	1	0	0.85	0.15
rep cluster 783	1	0	0.85	0.15
rep cluster 784	4	0	3.39	0.61
rep cluster 787	3	0	2.54	0.46
rep cluster 792	4	0	3.39	0.61
rep cluster 793	6	0	5.09	0.91

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 795	5	6	9.33	1.67
rep cluster 796	5	0	4.24	0.76
rep cluster 799	14	0	11.87	2.13
rep cluster 800	4	0	3.39	0.61
rep cluster 802	0	2	1.7	0.3
rep cluster 803	9	0	7.63	1.37
rep cluster 806	5	2	5.93	1.07
rep cluster 807	36	0	30.52	5.48
rep cluster 81	3	0	2.54	0.46
rep cluster 814	8	0	6.78	1.22
rep cluster 816	5	0	4.24	0.76
rep cluster 822	1	0	0.85	0.15
rep cluster 824	4	0	3.39	0.61
rep cluster 827	10	0	8.48	1.52
rep cluster 83	1	0	0.85	0.15
rep cluster 831	3	0	2.54	0.46
rep cluster 832	1	0	0.85	0.15
rep cluster 833	1	0	0.85	0.15
rep cluster 835	1	0	0.85	0.15
rep cluster 837	1	0	0.85	0.15
rep cluster 84	1	0	0.85	0.15
rep cluster 840	1	0	0.85	0.15
rep cluster 841	73	2	63.58	11.42
rep cluster 842	2	0	1.7	0.3
rep cluster 843	2	0	1.7	0.3
rep cluster 845	2	0	1.7	0.3
rep cluster 848	8	0	6.78	1.22
rep cluster 849	1	0	0.85	0.15
rep cluster 850	13	0	11.02	1.98
rep cluster 853	1	0	0.85	0.15
rep cluster 855	1	0	0.85	0.15
rep cluster 856	1	0	0.85	0.15
rep cluster 858	7	2	7.63	1.37
rep cluster 862	1	0	0.85	0.15
rep cluster 865	3	0	2.54	0.46
rep cluster 866	1	0	0.85	0.15
rep cluster 867	8	0	6.78	1.22
rep cluster 870	11	12	19.5	3.5
rep cluster 871	2	0	1.7	0.3
rep cluster 873	3	0	2.54	0.46
rep cluster 881	2	0	1.7	0.3
rep cluster 887	1	0	0.85	0.15
rep cluster 889	185	0	156.83	28.17
rep cluster 889,rep cluster 893	5	0	4.24	0.76
rep cluster 889,rep cluster 893,rep cluster 943	8	0	6.78	1.22
rep cluster 889,rep cluster 935	3	0	2.54	0.46
rep cluster 889,rep cluster 943	1	0	0.85	0.15
rep cluster 891	1	0	0.85	0.15
rep cluster 892	5	0	4.24	0.76
rep cluster 893	242	0	205.15	36.85
rep cluster 893,rep cluster 935	2	0	1.7	0.3
rep cluster 893,rep cluster 943	4	0	3.39	0.61
rep cluster 896	8	0	6.78	1.22
rep cluster 899	5	0	4.24	0.76

Replicon combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
rep cluster 901	1	0	0.85	0.15
rep cluster 902	1	0	0.85	0.15
rep cluster 904	1	0	0.85	0.15
rep cluster 905	4	0	3.39	0.61
rep cluster 906	2	0	1.7	0.3
rep cluster 909	3	0	2.54	0.46
rep cluster 910	26	0	22.04	3.96
rep cluster 911	1	0	0.85	0.15
rep cluster 914	11	0	9.33	1.67
rep cluster 916	4	0	3.39	0.61
rep cluster 918	1	0	0.85	0.15
rep cluster 919	1	0	0.85	0.15
rep cluster 92	1	0	0.85	0.15
rep cluster 920	16	0	13.56	2.44
rep cluster 922	1	0	0.85	0.15
rep cluster 923	1	0	0.85	0.15
rep cluster 93	1	0	0.85	0.15
rep cluster 935	2	0	1.7	0.3
rep cluster 935,rep cluster 943	1	0	0.85	0.15
rep cluster 936	7	0	5.93	1.07
rep cluster 943	61	2	53.41	9.59
rep cluster 946	2	0	1.7	0.3
rep cluster 949	2	0	1.7	0.3
rep cluster 950	3	0	2.54	0.46
rep cluster 957	3	0	2.54	0.46
rep cluster 958	9	0	7.63	1.37
rep cluster 96	2	0	1.7	0.3
rep cluster 961	3	0	2.54	0.46
rep cluster 962	1	0	0.85	0.15
rep cluster 966	2	0	1.7	0.3
rep cluster 969	4	0	3.39	0.61
rep cluster 971	5	0	4.24	0.76
rep cluster 971,rep cluster 980	1	0	0.85	0.15
rep cluster 974	3	0	2.54	0.46
rep cluster 976	65	0	55.1	9.9
rep cluster 976,rep cluster 980	2	0	1.7	0.3
rep cluster 976,rep cluster 992	1	0	0.85	0.15
rep cluster 977	1	0	0.85	0.15
rep cluster 980	137	0	116.14	20.86
rep cluster 980,rep cluster 992	5	0	4.24	0.76
rep cluster 982	3	2	4.24	0.76
rep cluster 986	1	0	0.85	0.15
rep cluster 991	14	0	11.87	2.13
rep cluster 992	93	0	78.84	14.16
rep cluster 993	1	0	0.85	0.15
rep cluster 994	4	0	3.39	0.61
rep cluster 995	17	0	14.41	2.59
rep cluster 996	2	0	1.7	0.3
rep cluster 998	10	0	8.48	1.52

Table 9: Combinations of replicon types as detected by MOBtyper and the number of plasmids with said replicon types that were observed to be BHR on the level of species, observed to be related to a single host, and the number of plasmids that were expected to be BHR and SH under the assumption that host range classification is independent of the combined replicon types. When a replicon type was detected multiple times, e.g. IncX3 was found twice on a plasmid, only one occurrence was counted.

MPF type	Observed BHR	Observed SH	Expected BHR	Expected SH
No MPF	4151	26729	4702.06	26177.94
MPF F	1109	8678	1490.25	8296.75
MPF G	0	22	3.35	18.65
MPF I	717	2047	420.87	2343.13
MPF T	1737	5334	1076.69	5994.31
MPF UNKNOWN	1	142	21.77	121.23

Table 8: MPF types as detected by MOBtyper and the number of plasmids with said MPF type that were observed to be BHR on the level of species, observed to be related to a single host, and the number of plasmids that were expected to be BHR and SH under the assumption that host range classification is independent of MPF type.

Relaxase combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
No relaxase types	18959	1577	17409.01	3126.99
MOBB	43	14	48.32	8.68
MOBC	705	466	992.69	178.31
MOBC,MOBF	22	2	20.35	3.65
MOBC,MOBF,MOBP	1	0	0.85	0.15
MOBC,MOBH	4	0	3.39	0.61
MOBC,MOBH,MOBP	8	0	6.78	1.22
MOBC,MOBP	26	0	22.04	3.96
MOBC,MOBP,MOBT,MOBV	1	0	0.85	0.15
MOBC,MOBP,MOBV	1	0	0.85	0.15
MOBC,MOBQ	4	0	3.39	0.61
MOBC,MOBT	2	0	1.7	0.3
MOBC,MOBT,MOBV	1	0	0.85	0.15
MOBC,MOBV	17	0	14.41	2.59
MOBF	6516	879	6268.97	1126.03
MOBF,MOBH	85	2	73.75	13.25
MOBF,MOBH,MOBP	10	0	8.48	1.52
MOBF,MOBH,MOBQ	2	0	1.7	0.3
MOBF,MOBP	957	80	879.1	157.9
MOBF,MOBP,MOBQ	7	0	5.93	1.07
MOBF,MOBQ	40	7	39.84	7.16
MOBF,MOBV	2	0	1.7	0.3
MOBH	2059	495	2165.11	388.89
MOBH,MOBP	96	10	89.86	16.14
MOBH,MOBP,MOBQ	2	0	1.7	0.3
MOBH,MOBQ	4	0	3.39	0.61
MOBH,MOBV	1	0	0.85	0.15
MOBM	59	0	50.02	8.98
MOBM,MOBT	1	0	0.85	0.15
MOBP	8729	3370	10256.7	1842.3
MOBP,MOBQ	295	12	260.25	46.75
MOBP,MOBQ,MOBV	3	0	2.54	0.46
MOBP,MOBQ,MOB Unknown	1	0	0.85	0.15
MOBP,MOBT	8	0	6.78	1.22
MOBP,MOBT,MOBV	75	5	67.82	12.18
MOBP,MOBV	45	0	38.15	6.85
MOBP,MOB Unknown	15	0	12.72	2.28
MOBQ	2653	487	2661.88	478.12
MOBQ,MOBV	27	0	22.89	4.11
MOBQ,MOB Unknown	2	0	1.7	0.3
MOBT	72	4	64.43	11.57
MOBT,MOBV	17	0	14.41	2.59

Relaxase combinations	Observed BHR	Observed SH	Expected BHR	Expected SH
MOBT,MOB Unknown	2	0	1.7	0.3
MOBV	1262	298	1322.46	237.54
MOB Unknown	111	7	100.03	17.97

Table 10: Combinations of relaxase types as detected by MOBtyper and the number of plasmids with said relaxase types that were observed to be BHR on the level of species, observed to be related to a single host, and the number of plasmids that were expected to be BHR and SH under the assumption that host range classification is independent of the combined relaxase types. When a relaxase type was detected multiple times, e.g. MOBC was found twice on a plasmid, only one occurrence was counted.