



**INVESTIGATING THE ECOLOGY AND DIVERSITY OF ANTIBIOTIC
RESISTANCE PLASMIDS BELONGING TO DIFFERENT INCOMPATIBILITY
GROUPS USING A BIOINFORMATICS APPROACH**

by

MONAMODI KESAMANG

Reg. No: 14100719

MSc (Biology and Biotechnological Sciences) (BIUST)

Department of Biology and Biotechnological Sciences,

College of Science,

Botswana International University of Science and Technology

monamodi.kesamang@studentmail.biust.ac.bw, (+26774719359)

A Dissertation/Thesis Submitted to the College of Sciences in Partial Fulfilment of the
Requirements for the Award of the Degree of Master of Science in Biology and
Biotechnological Sciences of BIUST

Supervisor(s): Dr Teddie O Rahube

Department of Biology and Biotechnological Sciences,

College of Science, BIUST

rahubet@biust.ac.bw, (+267 76698812)

Signature:

Date: 27/03/2018

March, 2018

DECLARATION AND COPYRIGHT

I, Monamodi Kesamang, declare that this dissertation/thesis is my own original work and that it has not been presented and will not be presented to any other university for a similar or any other degree award.



Signature:

.....

This dissertation/thesis is copyright material protected under the Berne Convention, the Copyright Act of 1999 and other international and national enactments, in that behalf, on intellectual property. It must not be reproduced by any means, in full or in part, except for short extracts in fair dealing; for researcher private study, critical scholarly review or discourse with an acknowledgement, without the written permission of the office of the Provost, on behalf of both the author and the BIUST.

CERTIFICATION

The undersigned certifies that he/she has read and hereby recommends for acceptance by the College of Science a dissertation/thesis titled: INVESTIGATING THE ECOLOGY AND DIVERSITY OF ANTIBIOTIC RESISTANCE PLASMIDS BELONGING TO DIFFERENT INCOMPATIBILITY GROUPS USING A BIOINFORMATICS APPROACH, in fulfilment of the requirements for the degree of Master of Science in (Biology and Biotechnological Science) of BIUST.



.....
Dr. Teddie. O. Rahube

(Supervisor)

Date: 27/03/2018

ACKNOWLEDGEMENTS

I am sincerely thankful to my programme supervisor Dr. Teddie. O. Rahube for the opportunity to do this project under his supervision. Amongst your contributions I greatly appreciate the guidance you have given me in the approaches of simplifying the workload associated with the management of overwhelming DNA sequence data in this project. Without your supervision this project would not have been possible to complete or even manage; you are the best mentor I could have had for this project. I would also like to thank the Botswana International University of Science and Technology for the opportunity to conduct my study with their financial support during the entire course of the research study. This extends to the use of the campus facilities including the materials purchased solely for this project. Thank you.

DEDICATION

I wish to dedicate this project to my mother, **Mrs. Tshegofatso V. Kesamang** a woman I consider so dear to my life. You have countless times before always been there for me, encouraged and motivated me through tough times. I appreciate your unconditional love and countless prayers for me to reach my aspirations.

I also wish to extend my gratitude to my colleagues and friends as founding postgraduate students from the - Biology and Biotechnological Sciences who have writhed along with me in the journey to a successful completion of our postgraduate studies

*First use of the word **plasmid** by Joshua Lederberg 1952.*

*“..These discussions [i.e. those about cytoplasmic inheritance] have left a plethora of terms adrift: pangenesis, bioblasts, plasmogenes, plastogenes, cytogenes and proviruses, which have lost their original utility owing to the accretion of vague or contradictory connotations. At the risk of adding to this list, **I propose “plasmid” as a generic term for any extrachromosomal hereditary determinant.** The plasmid itself may be genetically simple or complex. (...). Thus it is possible to formulate a graded series of symbioses as genic interactions, from the co-habitants of a single chromosome, through heterokaryons and plasmids, to extracellular ecological associations of variable stability and specificity”.*

Lederberg, J., 1952.

TABLE OF CONTENTS

DECLARATION AND COPYRIGHT	i
CERTIFICATION	ii
ACKNOWLEDGEMENTS	iii
DEDICATION	iv
TABLE OF CONTENTS	v
LIST OF TABLES.....	ix
LIST OF FIGURES.....	xi
LIST OF APPENDICES.....	xiv
LIST OF ABBREVIATIONS	xx
ABSTRACT	xxi
CHAPTER 1: INTRODUCTION	1
1.1. INTRODUCTION	1
1.2. STATEMENT OF THE PROBLEM	2
1.3. SPECIFIC OBJECTIVES	3
RESEARCH QUESTION/ HYPOTHESES.....	3
1.4. SIGNIFICANCE OF THE STUDY	4
CHAPTER 2: LITERATURE REVIEW	5
2.1. PLASMID STRUCTURE, NATURAL OCCURRENCE AND MOBILITY	5
2.2. MOBILE GENETIC ELEMENTS AND PLASMID ACQUIRED ANTIBIOTIC RESISTANCE	6
2.3. PLASMID CLASSIFICATION, INCOMPATIBILITY AND REPLICON TYPING	8
2.4. BIOINFORMATICS ANALYSIS OF PLASMIDS	10
CHAPTER 3: METHODOLOGY	12
3.1. DATA COLLECTION AND SCREENING.	12
3.1.1. Source of DNA sequence data.....	12
3.1.2. Plasmids sorted by complete against incomplete sequences	12
3.1.3. Plasmid sorted by resistant against non-resistant sequences	13
3.1.4. Plasmid sorted by circular against linear sequences.....	13
3.1.5. Circular antibiotic resistance plasmids sorted by genera of host bacteria.....	13
3.1.6. Segregating resistance plasmids amongst the Gram stain reaction (positive or negative) of the host bacteria.....	13

3.1.7. Analysis of multidrug resistance in a population of all circular resistance plasmids.....	14
3.1.8. Analysis of the range of sizes (molecular weight) of circular resistance plasmids.....	14
3.2. PLASMID SEQUENCE ANALYSIS	15
3.2.1. Analysis of incompatibility in circular resistance plasmids using PBRT.	15
3.3. DETERMINING PLASMID ECOLOGY BY REFERENCE ENVIRONMENT	16
3.3.1. Description of the environment categories.....	16
3.3.2. Association of Incompatibility Groups with Isolation Source (Reference Environment).....	17
3.4. FREQUENCY ANALYSIS OF ANTIBIOTIC RESISTANCE GENES IN PLASMIDS	17
3.4.1. Antibiotic Resistance Gene Frequency in Incompatibility groups.....	17
3.5. COMPARATIVE GENOMIC ANALYSIS OF ANTIBIOTIC RESISTANCE PLASMIDS	18
3.5.1. Phylogenetic Analysis of Plasmid Incompatibility	18
3.6. ANALYSIS OF REPLICON TYPING MARKERS.....	19
3.6.1. Development of new Inc/rep group replicon typing markers.....	19
CHAPTER 4: RESULTS AND DISCUSSION.....	22
4.1. Plasmids sorted by domain of host organism.	22
4.2. Plasmids sorted by complete against incomplete sequences.	22
4.3. Plasmids sorted by resistant against non-resistant sequences.....	23
4.4. Plasmids sorted by circular against linear sequences.	24
4.5. Plasmids sorted by genera of host bacteria.	25
4.6. Segregating resistant plasmids by the Gram stain reaction (positive or negative) of the host bacteria.....	26
4.7. Analysis of multidrug resistance in a population of all circular resistance plasmids	27
4.8. Analysis of the range of sizes (molecular weight) of circular resistance plasmids.	29
4.9. Analysis of ARGs frequency within the range of sizes (Kbp) of circular resistance plasmids.....	31
5.0. Analysis of the different classes of antibiotic resistance genes found in all resistance plasmids	32
5.1. Plasmid replicon typing	34
5.2. Characterization of multiple replicon AR plasmids.....	35

5.3.	Incompatibility/replicon sorting of typed plasmids	37
5.4.	The frequency of specific ARGs as determined in different plasmid incompatibility groups.....	39
5.4.1.	Inc W (AR profile)	39
5.4.2.	Inc F's (AR profile)	40
5.4.3.	Inc HI1 (AR profile)	41
5.4.4.	Inc HI2 (AR profile)	42
5.4.5.	Inc I1 (AR profile).....	43
5.4.6.	Inc X (AR profile)	44
5.4.7.	Inc X1&X2 (AR profile)	45
5.4.8.	Inc X3 (AR profile)	46
5.4.9.	Inc X4 (AR profile)	47
5.4.10.	Inc L & M (AR profile)	48
5.4.11.	Inc N (AR profile).....	49
5.4.12.	Inc Y (AR profile).....	50
5.4.13.	Inc P (AR profile)	50
5.4.14.	Inc A & C (AR profile).....	51
5.4.15.	Inc T (AR profile)	52
5.4.16.	Inc K & B (AR profile).....	53
5.4.17.	Inc B & O (AR profile).....	54
5.4.18.	Inc R (AR profile)	55
5.4.19.	Inc U (AR profile).....	57
5.4.20.	ColE1 (AR profile)	58
5.4.21.	Comprehensive AR profile of all typeable plasmids.	59
5.5.	Plasmid analysis by reference environment.....	62
5.6.	Distribution of Plasmid Incompatibility Groups in different environments	63
5.7.	Comparative genomic analysis of completely sequenced antibiotic resistance plasmid groups.....	69
5.8	New primer pair sequences for detection of plasmid groups.....	71
CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS		77
CONCLUSIONS AND RECOMMENDATIONS		77
REFERENCES		79
CHAPTER 6: SUPPLEMENTARY MATERIALS		84
6.1.	APPENDICES	84

6.1.1.	APPENDIX A: Phylogenetic trees (Analysis of Plasmid Incompatibility) ..	84
6.1.2.	APPENDIX B: Primer MSA against target sequences (Primer diagnosis).	103
6.1.3.	APPENDIX C: Additional/Supplementary Figures	123

LIST OF TABLES

Table 3.1: Plasmid incompatibility replicon typing primers, reference sequences and Amplicon size.....	19
Table 5.4a: Inc W (AR frequency).....	39
Table 5.4b: Inc F's (AR frequency).....	41
Table 5.4c: IncHI1 (AR frequency).....	42
Table 5.4d: Inc HI2 (AR frequency).....	42
Table 5.4e: IncI1 (AR frequency).....	44
Table 5.4f: Inc X (AR frequency).....	44
Table 5.4g: Inc X1 & X2 (AR frequency).....	45
Table 5.4h: Inc X3 (AR frequency).....	46
Table 5.4i: Inc X4 (AR frequency).....	47
Table 5.4j: Inc L & M (AR frequency).....	48
Table 5.4k: Inc N (AR frequency).....	50
Table 5.4l: Inc Y (AR frequency).....	50
Table 5.4m: Inc P (AR frequency).....	51
Table 5.4n: Inc A & C (AR frequency).....	53
Table 5.4o: Inc T (AR frequency).....	53
Table 5.4p: Inc K & B (AR frequency).....	54
Table 5.4q: Inc B & O (AR frequency).....	56
Table 5.4r: Inc R (AR frequency).....	57

Table 5.4s: Inc U (AR frequency).....	58
Table 5.4t: ColE1 (AR frequency).....	60
Table 5.4u: Comprehensive AR profile of all typeable plasmids.....	61
Table 5.5: Allocation of plasmids by incompatibility against their isolation source (reference environment).....	63
Table 5.8: New primer sequences developed to complement generic primer sequences...	72
Table 5.9: IUPAC system for nucleotide nomenclature.....	77

LIST OF FIGURES

Figure 1: Generic structure of a conventional circular plasmid (basic features).....	5
Figure 2: Beta-lactamase enzymes documented from 1965 to 2010.....	7
Figure 4.1: Plasmids sorted by domain of host organism.....	22
Figure 4.2: Plasmids sorted by complete against incomplete sequences.....	23
Figure 4.3: Plasmids sorted by resistant against non-resistant.....	24
Figure 4.4: Plasmids sorted by circular against linear sequences.....	25
Figure 4.5: Circular resistant plasmids sorted by bacterial species.....	26
Figure 4.6: Circular resistant plasmids sorted by Gram stain reaction of host bacteria.....	27
Figure 4.7: Plasmids sorted by multidrug resistance in a population of circular resistance plasmids.....	28
Figure 4.8: Distribution of resistance plasmids (%) according to the size in molecular weight (Kbp)	29
Figure 4.9: Analysis of resistance genes within the range of sizes (Kbp) of circular resistance.....	31
Figure 5.0a: Classes and relative quantity of antibiotic resistant genes found in all resistant plasmids.....	32
Figure 5.0b: Classes and relative quantity of antibiotic resistant genes found in circular resistant plasmids.....	33
Figure 5.1a: Antibiotic resistant plasmids sorted by success in typing using PBRT.....	34
Figure 5.2a: Antibiotic resistant plasmids sorted by single against multiple replicons.....	35
Figure 5.2b: Antibiotic resistant plasmids sorted by number or typeable replicons.....	36

Figure 5.3c: Antibiotic resistant plasmids (typeable using PBRT) sorted by Inc/rep region.....	37
Figure 5.3d: Quantifying Inc F resistant plasmids into subgroups.....	38
Figure 5.4a: Classes of antibiotic resistance genes in Inc W.....	39
Figure 5.4b: Classes of antibiotic resistance genes in Inc F's.....	40
Figure 5.4c: Classes of antibiotic resistance genes in Inc HI1.....	41
Figure 5.4d: Classes of antibiotic resistance genes in Inc HI2.....	42
Figure 5.4e: Classes of antibiotic resistance genes in Inc II.....	43
Figure 5.4f: Classes of antibiotic resistance genes in Inc X.....	44
Figure 5.4g: Classes of antibiotic resistance genes in Inc X1&X2.....	45
Figure 5.4h: Classes of antibiotic resistance genes in Inc X3.....	46
Figure 5.4i: Classes of antibiotic resistance genes in Inc X4.....	47
Figure 5.4j: Classes of antibiotic resistance genes in Inc L&M.....	48
Figure 5.4k: Classes of antibiotic resistance genes in Inc N.....	49
Figure 5.4m: Classes of antibiotic resistance genes in Inc P.....	51
Figure 5.4n: Classes of antibiotic resistance genes in Inc A&C.....	52
Figure 5.4o: Classes of antibiotic resistance genes in Inc T.....	53
Figure 5.4p: Classes of antibiotic resistance genes in Inc K&B.....	54
Figure 5.4q: Classes of antibiotic resistance genes in Inc B&O.....	55
Figure 5.4r: Classes of antibiotic resistance genes in Inc R.....	57
Figure 5.4s: Classes of antibiotic resistance genes in Inc U.....	58
Figure 5.4t: Classes of antibiotic resistance genes in ColE1.....	59
Figure 5.4u: Classes of antibiotic resistance genes in all typeable plasmids.....	61

Figure 5.6a. Chord diagram of Inc/rep groups against environment categories (Inc W, Inc F's, Inc HI1 &Inc HI2). The environment categories are segregated into the right portion of the diagrams, denoted as; Natural (dark blue-DB), Host (light green-LG), Managed (Dark Green-DG), Unknown (Pink-P).....	65
Figure 5.6b. Chord diagram of Inc/rep groups against environment categories (Inc I1, IncX, IncX1&X2 and IncX3). The environment categories are segregated into the right portion of the diagrams, denoted as; Natural (dark blue-DB), Host (light green-LG), Managed (Dark Green-DG), Unknown (Pink-P).....	66
Figure 5.6c. Chord diagram of Inc/rep groups against environment categories (Inc X4, Inc L&M, Inc N &Inc Y).The environment categories are segregated into the right portion of the diagrams, denoted as; Natural (dark blue-DB), Host (light green-LG), Managed (Dark Green-DG), Unknown (Pink-P).....	67
Figure 5.6d. Chord diagram of Inc/rep groups against environment categories (Inc P, Inc A & C, Inc T and Inc K & B). The environment categories are segregated into the right portion of the diagrams, denoted as; Natural (dark blue-DB), Host (light green-LG), Managed (Dark Green-DG), Unknown (Pink-P).....	68
Figure 5.6e. Chord diagram of Inc/rep groups against environment categories (Inc B&O, Inc R, Inc U & ColE1). The environment categories are segregated into the right portion of the diagrams, denoted as; Natural (dark blue-DB), Host (light green-LG), Managed (Dark Green-DG), Unknown (Pink-P).....	69
Figure 5.7. Phylogenetic tree of all Inc/rep groups, two representative plasmids from each group; one representing the most recent terminal node plasmid (a) and another representing the root or reference outgroup plasmid (b). Emphasis is on representing cross relationships of Inc/rep groups of one to another where applicable.....	71

LIST OF APPENDICES

Appendix A1. Inc W phylogenetic tree.....	85
Appendix A2. Inc F(rep B) phylogenetic tree.....	86
Appendix A3. Inc FIA-R phylogenetic tree.....	87
Appendix A4. Inc FIB-R phylogenetic tree.....	87
Appendix A5. Inc FIC-R phylogenetic tree.....	87
Appendix A6. Inc FII-R phylogenetic tree.....	87
Appendix A7. Inc FIA2-R phylogenetic tree.....	88
Appendix A8. Inc FIB2-R phylogenetic tree.....	88
Appendix A9. Inc FIBs-R phylogenetic tree.....	88
Appendix A10. Inc FIC2-R phylogenetic tree.....	89
Appendix A11. Inc FII2-R phylogenetic tree.....	89
Appendix A12. Inc FII(repA) phylogenetic tree.....	90
Appendix A13. Inc FII[Fd] phylogenetic tree.....	90
Appendix A14. Inc FIIk phylogenetic tree.....	91
Appendix A15. Inc FIIs phylogenetic tree.....	91
Appendix A16. Inc FIly phylogenetic tree.....	91
Appendix A17. Inc HI1 phylogenetic tree.....	92
Appendix A18. Inc HI2 phylogenetic tree.....	92
Appendix A19. Inc I1 phylogenetic tree.....	93
Appendix A20. Inc X phylogenetic tree.....	93
Appendix A21. Inc X1 phylogenetic tree.....	94

Appendix A22. Inc X2 phylogenetic tree.....	94
Appendix A23. Inc X3 phylogenetic tree.....	95
Appendix A24. Inc X4 phylogenetic tree.....	95
Appendix A25. Inc L &M phylogenetic tree.....	96
Appendix A26. Inc N phylogenetic tree.....	97
Appendix A27. Inc Y phylogenetic tree.....	98
Appendix A28. Inc P phylogenetic tree.....	98
Appendix A29. Inc A & C phylogenetic tree.....	98
Appendix A30. Inc T phylogenetic tree.....	99
Appendix A31. Inc K & B phylogenetic tree.....	99
Appendix A32. Inc B & O phylogenetic tree.....	99
Appendix A33. Inc R phylogenetic tree.....	100
Appendix A34. Inc U phylogenetic tree.....	100
Appendix A35. ColE1 phylogenetic tree.....	101
Appendix B1. Inc W ^a Forward primer MSA against 3 amplicons.....	102
Appendix B2. Inc W ^a Reverse primer MSA against 3 amplicons.....	102
Appendix B3. Inc F(repB) ^a Forward primer MSA against 74 amplicons.....	102
Appendix B4. Inc F(repB) ^a Reverse primer MSA against 74 amplicons.....	103
Appendix B5. Inc FIA ^a Forward primer MSA against 7 amplicons.....	103
Appendix B6. Inc FIA ^a Reverse primer MSA against 7 amplicons.....	103
Appendix B7. Inc FIB ^a Forward primer MSA against 29 amplicons.....	104
Appendix B8. Inc FIB ^a Reverse primer MSA against 29 amplicons.....	104
Appendix B9. Inc FIC ^a Forward primer MSA against 8 amplicons.....	104

Appendix B10. Inc FIC ^a Reverse primer MSA against 8 amplicons.....	104
Appendix B11. Inc FII ^a Forward primer MSA against 13 amplicons.....	105
Appendix B12. Inc FII ^a Reverse primer MSA against 13 amplicons.....	105
Appendix B13. Inc FIA ^b Forward primer MSA against 17 amplicons.....	105
Appendix B14. Inc FIA ^b Reverse primer MSA against 17 amplicons.....	105
Appendix B15. Inc FIB ^b Forward primer MSA against 26 amplicons.....	106
Appendix B16. Inc FIB ^b Reverse primer MSA against 26 amplicons.....	106
Appendix B17. Inc FIBs ^b Forward primer MSA against 27 amplicons.....	106
Appendix B18. Inc FIBs ^b Reverse primer MSA against 27 amplicons.....	107
Appendix B19. Inc FIC ^b Forward primer MSA against 8 amplicons.....	107
Appendix B20. Inc FIC ^b Reverse primer MSA against 8 amplicons.....	107
Appendix B21. Inc FII ^b Forward primer MSA against 42 amplicons.....	107
Appendix B22. Inc FII ^b Reverse primer MSA against 42 amplicons.....	108
Appendix B23. Inc FII(repA) ^b Forward primer MSA against 50 amplicons.....	108
Appendix B24. Inc FII(repA) ^b Reverse primer MSA against 50 amplicons.....	108
Appendix B25. Inc FII[Fd] ^b Forward primer MSA against 1 amplicon.....	109
Appendix B26. Inc FII[Fd] ^b Reverse primer MSA against 1 amplicon.....	109
Appendix B27. Inc FIIk ^b Forward primer MSA against 23 amplicons.....	109
Appendix B28. Inc FIIk ^b Reverse primer MSA against 23 amplicons.....	109
Appendix B29. Inc FIIIs ^b Forward primer MSA against 9 amplicons.....	110
Appendix B30. Inc FIIIs ^b Reverse primer MSA against 9 amplicons.....	110
Appendix B31. Inc FIIy ^b Forward primer MSA against 6 amplicons.....	110
Appendix B32. Inc FIIy ^b Reverse primer MSA against 6 amplicons.....	110

Appendix B33. Inc HI1 ^a Forward primer MSA against 15 amplicons.....	110
Appendix B34. Inc HI1 ^a Reverse primer MSA against 15 amplicons.....	111
Appendix B35. Inc HI2 ^a Forward primer MSA against 10 amplicons.....	111
Appendix B36. Inc HI2 ^a Reverse primer MSA against 10 amplicons.....	111
Appendix B37. Inc I1 ^a Forward primer MSA against 40 amplicons.....	111
Appendix B38. Inc I1 ^a Reverse primer MSA against 40 amplicons.....	112
Appendix B39. Inc X ^a Forward primer MSA against amplicons.....	112
Appendix B40. Inc X ^a Reverse primer MSA against amplicons.....	112
Appendix B41. Inc X1 ^d Forward primer MSA against 19 amplicons.....	112
Appendix B42. Inc X1 ^d Reverse primer MSA against 19 amplicons.....	113
Appendix B43. Inc X2 ^d Forward primer MSA against 19 amplicons.....	113
Appendix B44. Inc X2 ^d Reverse primer MSA against 19 amplicons.....	113
Appendix B45. Inc X3 ^d Forward primer MSA against 22 amplicons.....	114
Appendix B46. Inc X3 ^d Reverse primer MSA against 22 amplicons.....	114
Appendix B47. Inc X4 ^d Forward primer MSA against 2 amplicons.....	114
Appendix B48. Inc X4 ^d Reverse primer MSA against 2 amplicons.....	114
Appendix B49. Inc L & M ^a Forward primer MSA against 21 amplicons.....	115
Appendix B50. Inc L & M ^a Reverse primer MSA against 21 amplicons.....	115
Appendix B51. Inc N ^a Forward primer MSA against 44 amplicons.....	115
Appendix B52. Inc N ^a Reverse primer MSA against 44 amplicons.....	116
Appendix B53. Inc Y ^a Forward primer MSA against 1 amplicons.....	116
Appendix B54. Inc Y ^a Reverse primer MSA against 1 amplicons.....	116
Appendix B55. Inc P ^a Forward primer MSA against 18 amplicons.....	116

Appendix B56. Inc P ^a Reverse primer MSA against 18 amplicons.....	117
Appendix B57. Inc A & C ^a Forward primer MSA against 57 amplicons.....	117
Appendix B58. Inc A & C ^a Reverse primer MSA against 57 amplicons.....	117
Appendix B59. Inc T ^a Forward primer MSA against 2 amplicons.....	118
Appendix B60. Inc T ^a Reverse primer MSA against 2 amplicons.....	118
Appendix B61. Inc K & B ^a Forward primer MSA against 12 amplicons.....	118
Appendix B62. Inc K & B ^a Reverse primer MSA against 12 amplicons.....	118
Appendix B63. Inc B & O ^a Forward primer MSA against 22 amplicons.....	119
Appendix B64. Inc B & O ^a Reverse primer MSA against 22 amplicons.....	119
Appendix B65. Inc R ^c Forward primer MSA against 23 amplicons.....	119
Appendix B66. Inc R ^c Reverse primer MSA against 23 amplicons.....	120
Appendix B67. Inc U ^c Forward primer MSA against 11 amplicons.....	120
Appendix B68. IncU ^c Reverse primer MSA against 11 amplicons.....	120
Appendix B69. ColE ^c Forward primer MSA against 83 amplicons.....	121
Appendix B70. ColE ^c Reverse primer MSA against 83 amplicons.....	121
Appendix C1. Plasmids sorted by multidrug resistance in a population of circular resistance plasmid.....	122
Appendix C2. Distribution of resistance plasmids according to the size in molecular weight (Kbp).....	123
Appendix C3. Classes and relative quantity of antibiotic resistant genes found in all resistant plasmids.....	124
Appendix C4. Classes and relative quantity of antibiotic resistant genes found in circular resistant plasmids.....	125

Appendix C5. Antibiotic resistant plasmids (typeable using PBRT) sorted by Inc/rep region.....	126
Appendix C6. Quantification of ARG's identified in all resistance plasmids into ARG Classes.....	127
Appendix C7. Quantification of ARG's identified in circular resistance plasmids into ARG Classes.....	128
Appendix C8. Comprehensive AR frequency table of all typeable plasmids in various Inc/rep groups.....	129

LIST OF ABBREVIATIONS

AR	Antibiotic Resistance
ARGs	Antibiotic Resistant Genes
ARDB	Antibiotic Resistance Genes Database
BLAST	Basic Local Alignment Search Tool
CARD	Comprehensive Antibiotic Resistance Database
CDC	Centre for Disease Control
ESBLs	Extended Spectrum β -lactamases
HGT	Horizontal Gene Transfer
Inc	Incompatibility
ICEs	Integrative and Conjugative elements
MDR	Multi Drug Resistant
MGEs	Mobile Genetic Elements
MSA	Multiple Sequence Alignment
NCBI	National Centre For Biotechnology Information
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PBRT	PCR Based Replicon Typing
PMLST	Plasmid Multi Locus Sequence Typing
Rep	Replicon
WGS	Whole Genome Sequence
WHO	World Health Organisation

ABSTRACT

Plasmids are extra-chromosomal mobile DNA elements found in bacteria, and often serve as vectors for accumulation and transfer of antibiotic resistance genes (ARGs) within bacterial populations in different environments. Plasmid mediated transfer of ARGs has accounted for the spread of antibiotic resistance within clinically relevant pathogenic bacteria; this has escalated into a global health problem. Plasmids can be classified into several groups based on their genetic incompatibility; meaning that plasmid with similar replication (rep) genes cannot co-exist in the same bacterial cell. Even though ARGs have been detected in samples from clinical and non-clinical environments, their occurrence and distribution within plasmid incompatibility groups is not well understood. The study sought to determine the extent of diversity and distribution of different plasmid incompatibility (Inc/rep) groups that are harbouring ARGs in three different environments broadly classified as natural, host-associated and managed. The hypothesis is stated that antibiotic resistance plasmids of a certain Inc/rep group occur in certain environments with higher frequency than other Inc/rep groups. In this research, DNA sequence data of nearly all known natural plasmids from NCBI databases was used to determine the correlation between plasmid Inc/rep group encoding antibiotic resistance and their source environment. The majority of antibiotic resistance plasmids were found in host associated environments particularly in human samples from clinical settings. Inc P was the only group clearly associated with all the 3 broad categories of environments. Results suggest no clear or notable variability on the correlation proportions of plasmids of other incompatibility groups being associated with specific environmental sources. The major limitation within this bioinformatics study was the plasmid sequences that were not completely sequenced and those not clearly described in terms of source or isolation. It is recommended that sequence data submission to public sequence databases be improved for future bioinformatics studies that aim to look at evolution of plasmids encoding antibiotic resistance.

Keywords: Plasmids, antibiotic resistance genes, incompatibility, bioinformatics

CHAPTER 1: INTRODUCTION

1.1. INTRODUCTION

The global community has been struggling with deterring the continuous emergence of bacteria that have acquired resistance to different varieties of antibiotics referred to as Multi Drug Resistant (MDR) bacteria (Alanis, 2005). With this problem, from a compromised health perspective, the annual comprehensive death reports have reached more than 100 thousand and over 160 million cases reported worldwide in 2014 (CDC, 2013; WHO, 2014). The prime cause of this problem is inadequate health-care in undeveloped countries and as for developed countries, it is the reality that MDR bacteria are undisputedly difficult to treat even with adequate health-care.

Antibiotics are naturally occurring chemical compounds or drugs that kill bacteria or prevent bacterial growth, they occur as a natural defence of microorganisms against one another. Antibiotic resistant genes (ARGs) have been around along with the prevalence of antibiotics in nature. The first antibiotic was discovered in 1929 by Alexander Fleming and afterwards resistance genes in *Staphylococcus aureus* were documented in 1940 by Edward Abraham and colleagues in a publication titled "An enzyme from bacteria able to destroy penicillin". The enzymes were termed penicillinases (Abraham & Chain, 1940). Subsequent research has since then witnessed the documentation of a collection of ARGs from community of bacteria in various environments that are now referred to as the resistome (D'Costa *et al.*, 2006; Wright, 2007). The information of DNA sequences is now managed in online databases. Such databases are a source of un-deciphered data of which is potentially useful as aid to the management of ARGs as a health concern or other applications (Liu & Pop, 2009) (Seshadri *et al.*, 2007).

ARGs are commonly transmitted between bacteria by extra chromosomal DNA elements called plasmids through a mechanism of horizontal (also known as lateral) gene transfer called conjugation. There is currently a large number of plasmids that harbour antibiotic resistance in bacteria. A single plasmid can acquire ARGs in succession and this leads to bacteria that have this plasmid expressing resistance against multiple antimicrobial agents including antibiotics, quaternary ammonium compounds and heavy metals (Rahube *et al.*, 2016). This phenomenon is called multidrug resistance (MDR), and has serious implications

in the treatment of pathogenic bacteria. Plasmids can be organised into groups depending on their genetic characteristics. In 1971, Hedges and Datta came up with a plasmid classification scheme based on the stability of plasmids during conjugation; this was referred to as incompatibility. The incompatibility of plasmids can be described based on the type of replication gene (replicon) on a particular plasmid. The type of replicon that a plasmid has is correlated to its evolutionary path and as such it is somehow a measure of the genetic diversity within a population of natural plasmids (Garcillán-Barcia *et al.*, 2009).

From the scientific study of how MDR bacteria occur in the various environments, it is important to recognise how genes of resistance are transferred from one environment to the next. This is done so as to track or predict means by which they aggregate in MDR bacteria using the data that already exists about ARGs from private and public databases such as NCBI (Martinez, 2012). The documented data from these databases may reveal the occurrence, diversity, abundance and fate of ARGs at each environment (Martinez, 2012). Metagenomic data of unculturable majority of microorganisms offers a richer source of DNA sequence data in addition to genomic data obtained through conventional culture based methods, which have limitations in revealing all of the genomic profile of an environment (Riesenfeld *et al.*, 2004). There has not been a systematic study of the resistome showing exactly how large and diverse it actually is in environmental bacterial communities (Hugenholtz & Tyson, 2008). Therefore some of the documented metagenomic data is yet to be implemented in bioinformatics studies.

1.2. STATEMENT OF THE PROBLEM

The premise of this study is based on the conjecture that, as more antibiotics are being overly used or misused and getting developed against pathogenic bacteria, a Darwinian adaptation mechanism known as natural selection takes effect. There is a progressive emergence of new ARGs in such bacteria against the selective pressure. To ensure their survival, bacteria develop the resistance gene through mutations or acquire resistance genes by various horizontal gene transfer mechanisms. The ARGs are then incorporated in plasmids and subsequently get transferred across bacterial communities in different environments to provide a selective advantage. However, little is known about occurrence and distribution of the different classes of plasmids encoding ARGs in different ecosystems, such knowledge

can be extrapolated from genomic and metagenomic data using a Bioinformatics approach
The study sought to determine the extent of diversity, distribution and evolution of different
classes of plasmids that harbour ARGs in three different environments broadly classified as
natural, host-associated and managed.

1.3. SPECIFIC OBJECTIVES

I. To extract plasmid sequences, sort and classify plasmids encoding ARGs into
different incompatibility groups and the environments which they were isolated
using a bioinformatics approach.

II. Determine the frequency of specific ARGs in different plasmid incompatibility
groups.

III. Infer evolutionary relationships by comparative genomic analysis of completely
sequenced plasmid groups encoding antibiotic resistance

IV. Design new DNA sequence markers for tracking specific antibiotic resistance
plasmids in the environment.

RESEARCH QUESTION/ HYPOTHESES

ARGs are widely distributed in natural (pristine) and human impacted (polluted)
environments (clinical and non-clinical). The occurrence, diversity and distribution of
specific plasmid incompatibility groups carrying ARGs is limited to environments where they
are maintained in bacterial communities. In light of this, the postulation of this study is that
only plasmid groups capable of replicating in a wide range of bacterial hosts (broad host
range) will occur in environments with higher diversity and activity of bacteria such as
polluted/impacted environments. Conversely, narrow host range plasmid groups will occur in
micro-environments associated with lower microbial activity and little antibiotic selective
pressure such as pristine environments. The hypothesis is then stated that antibiotic resistance

plasmids of a certain incompatibility group can be found to occur in certain environments with higher frequency (abundance) than other groups plasmids

1.4. SIGNIFICANCE OF THE STUDY

This research has expended and expanded the value of the availability of plasmid genomic and metagenomic data for Bioinformatics research purposes. It has identified gaps that will encourage further research into DNA sequencing of more plasmids in under-represented environmental sources. The study has also provided insights into plasmid evolution, persistence and spread across different ecological niches, and subsequently assist by providing alternative molecular tools for detecting and tracking antibiotic resistance plasmids in clinical and non-clinical environmental samples.

CHAPTER 2: LITERATURE REVIEW

2.1. PLASMID STRUCTURE, NATURAL OCCURRENCE AND MOBILITY

Plasmids are a set of nucleic acid molecules that exist as extra chromosomal genetic elements and have the characteristic ability of initiating independent replication (Lederberg, 1952; Marmur *et al.*, 1961). These genetic elements have a variety of sizes ranging from under 1kbp (1000bp) to nearly 3000kbp (Shintani *et al.*, 2015). They occur in abundance in almost all genera of bacteria and some in archaea (Funnell and Phillips, 2004). A few plasmids are known to exist in plant cells (Hanza, 2008). In nature these molecules have ordinarily been known not to carry any essential genes for the cell, for this reason they initially served as selfish genetic elements with the only role of self-replication using the cell's machinery. In addition, plasmids also played a role of parasites since they impose a metabolic burden for the cell in order to maintain their own survival (Werren, 2011). Kado (1998) has attributed the earliest existence of plasmids to the uncontrollable nature of the stability of DNA and related molecular processes. He proposed that variant forms of plasmids came into being when short strands of nucleic acids developed into nascent replicons through evolution (Kado, 1998). Today, most plasmids carry accessory genes that enhance adaptation of the host cell. Plasmids are commonly found as covalently bound circular DNA molecules while a few are made of linear double stranded DNA molecules (Hinnebusch and Tilly, 1993). A typical plasmid structure comprises of replication genes, transfer genes and accessory genes such as those that help bacteria survive in the presence of antibiotics (Figure 1).

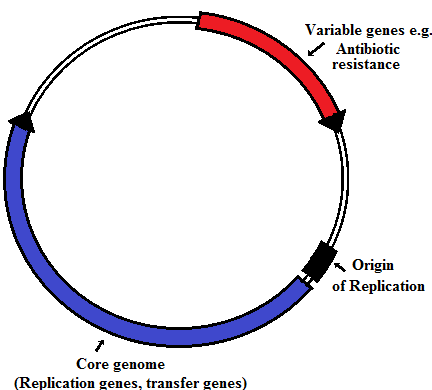


Figure 1. Generic structure of a conventional circular plasmid (basic features) (Garcillán-Barcia *et al.*, 2011).

One of the important properties of plasmids is their association with movement from one location to another. Plasmids can have mobility in one of two general ways; intra-cellular movement when they either integrate or excise themselves into the host cell chromosome. Another mechanism of plasmid movement is extra-cellular; this is when plasmids transfer themselves from one cell to another (Smillie *et al.*, 2010). Plasmid mobility is an essential component of their survival through generating genetic diversity which helps them spread to new hosts. The movement of plasmid from one cell to another is called conjugation, one of the mechanisms described as horizontal gene transfer (HGT). Plasmids can be classified into three categories with regards to mobility. These are mobile (conjugative), mobilizable, and non-mobilizable plasmids (Smillie *et al.*, 2010). Almost half of all known plasmids are in one way or the other capable of initiating their transfer from one host to another (Shintani *et al.*, 2015), (Smillie *et al.*, 2010). Mobile plasmids also easily acquire new genes and this essentially helps them evolve and better adapt to their host and environment.

2.2. MOBILE GENETIC ELEMENTS AND PLASMID ACQUIRED ANTIBIOTIC RESISTANCE

Mobile genetic elements (MGEs) are DNA molecules that are capable of moving from one genetic locus to another (Bennett, 2008). Mobile plasmids can be considered to be mobile genetic elements that are capable of stable independent replication (Rankin *et al.*, 2011). However, the majority of MGEs are transposons (transposable elements). Through their inherent action of integrating themselves from one genetic locus to the other, transposons are a major source of genetic diversity, (Dimitriu *et al.*, 2015; Bennett, 1999). Transposons can insert and excise different types of genes from chromosomes and into plasmids or between different plasmids co-existing in the same cell. This characteristic action has made plasmids essential extra chromosomal elements that have adapted a role of functionality for their host. An adaptive function like the acquisition of an essential gene can be readily integrated into plasmids as an account of transposons and other mobile genetic elements (Hall, 1997; Bennett, 2004). Antibiotic resistance genes constitute some of the well-studied essential genes that are a major component of the adaptive role of plasmids that have transpired into a global health problem (Davies and Davies, 2010). Initially, resistance genes occur as part of the chromosome of a naturally resistant microbe, they are then get carried from a resistant cell to a non-resistant cell through mobile genetic elements like transposons and other gene

capture systems called integrons and gene cassettes (Hall, 1997). These eventually get integrated into conjugative and mobilizable plasmids which eventually get spread through a range of bacterial species via conjugation (Zankari *et al.*, 2012; Perry and Wright, 2014; Martínez *et al.*, 2015).

The first antibiotic resistance gene isolated in a bacterial plasmid was documented in 1961 by Watanabe and Fukasawa (1961). This has since led to the aggregation of resistance genes against a range of antibiotics that have been developed since then. Antibiotic resistance has been shown to develop due to the ever increasing development of new antibiotics which serve as selective pressure that induce genetic mutations (Davies and Davies, 2010; WHO, 2014). Some ARGs have been segregated into families and named after the structurally related group of similar antibiotics they act against, e.g. Beta-lactamases are enzymes that act against a class of antibiotics called beta-lactams. Davies and Davies have correlated and quantified the aggregation of such resistance genes with their corresponding antibiotics for nearly 40 years since their first introduction; (Figure 2) (Davies and Davies, 2010). This has indicated that more and more ARGs are aggregating in different environments at an exponential rate, most of which are prone to spread efficiently via plasmids.

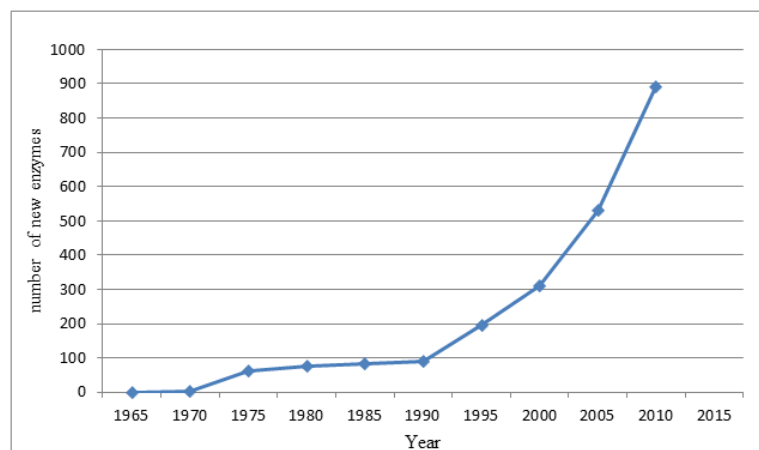


Figure 2. Number of new enzymes documented from 1965 to 2010. (Adapted from Davies and Davies, 2010).

As antibiotic resistance genes increase in quantity so does the chance of their dissemination to different environmental niches. Rahube and colleagues have indicated the propagation of resistance genes from wastewater treatment plants (as a major reservoir) to natural water sources after treatment (Rahube *et al.*, 2014). In the same study it was also indicated that plasmid acquired resistance is not easily managed from one environment to the next. The

aggregation of resistance genes in the environment increases the chances of their co-occurrence in one host cell and most frequently captured in one plasmid (Carattoli, 2013). This co-occurrence of resistance genes in a single plasmid results in the expression of multi-drug resistance by the host organism. Multi-drug resistance is the most devastating adversity posed by the spread of antibiotic resistant genes. Multidrug resistance manifests in two common ways, by the accumulation of multiple genes that encode resistance to different antibiotics in one cell. It also occurs when membrane proteins called efflux pumps acquire an adaptive function through mutation to extrude a wide range of drugs, these are now called multidrug efflux pumps (Nikaido, 2009). In the former case, the ARGs are captured and accumulated via MGEs and are stably and effectively expressed in plasmids (Nikaido, 2009). The general types of bacterial efflux pumps associated with antibiotic resistance are membrane proteins that fall into five classes; the resistance nodulation division family (RND); only found in Gram negative bacteria as it spans the inner and outer membrane, the small multidrug resistance family (SMR), the multidrug and toxic compound extrusion family (MATE), the drug metabolite transporter (DMT) and the Major facilitator superfamily (MFS). The latter mechanisms can occur in both Gram positive and negative bacteria (Zgurskaya and Nikaido, 2000). This fact highlights the role of the plasmids in the development and spread of multidrug resistance genes in different environments. It has recently been documented in a case study by Pal and colleagues that heavy metal resistance genes can have a co-occurrence with ARGs in plasmids (Pal *et al.*, 2015). This is another plasmid mediated adaptive role in response to extreme environments (such as heavy metal polluted). This co-existence means ARGs can still be maintained in bacterial population in the absence of antibiotic selection. In light of how plasmids acquire multidrug resistance and how they spread it between different environments, it is important to know the class of plasmids affiliated with different hosts and their ecological niche.

2.3. PLASMID CLASSIFICATION, INCOMPATIBILITY AND REPLICON TYPING

Plasmid classification has always been considered an important aspect to study ever since the early discoveries that were made in plasmid biology, just as it is important to classify living things. Initially, natural plasmids were classified as either F⁺ or F⁻. During the 1960s, the F plasmid was determined to have mutant phenotypes that were non-transmissible, it was later discovered that the reason for this was because the fertility functions (conjugative genes) of

the F factor were inhibited and this prevented the plasmids from becoming conjugative (Couturier *et al.*, 1988). These plasmids were referred to as F- plasmids, which are distinguishable from F+ plasmids. In the early 1970s plasmids were then classified according to a phenomenon known as incompatibility (Novick *et al.*, 1976). This refers to the inability of two plasmids to cohabit the same host cell (Datta and Hedges, 1972; Novick, 1987). Incompatibility is a measure of the evolutionary relatedness of two plasmids, this is because the plasmids utilize the same replication machinery and consequently they have the similar replication genes (replicon) (Novick, 1987). The reason for the incompatibility in this case was because similar replicons were in competition of the same replication machinery of the host (Novick, 1987). Eventually, as time progressed it was noted that incompatibility was an inefficient way to classify all plasmids because some plasmids, despite having the same replicon could sometimes be compatible. More-so, some plasmids could also have more than one replicon. Along with all these issues, incompatibility tests were laborious and practically challenging to successfully implement (Couturier *et al.*, 1988; Bergquist *et al.*, 1986).

As an alternative to the shortcomings related to incompatibility classification, Couturier and colleagues developed a set of molecular probes associated with the replicon regions of 31 incompatibility groups (Couturier *et al.*, 1988). 25 of these were found in plasmids that were in *Enterobacteriaceae* and 7 of them were found in *Staphylococcus* species. These replication probes were referred to as Inc/rep probes and had been in use for nearly 15 years until they became inefficient due to their non-specificity. In 2005, Carattoli and colleagues re-developed the Inc/rep probes and created PCR primers based on the re-developed probe sequences. This established a practical plasmid typing technique using PCR called Plasmid based replicon typing (PBRT) (Carattoli *et al.*, 2005). This technique was readily adapted and practiced for classifying most natural plasmids. Carattoli and colleagues designed PBRT primers for the incompatibility (Inc) groups FIA, FIB, FIC, HI1, HI2, I1-Ig, L/M, N, P, W, T, A/C, K, B/O, X, Y, F, and FIIA replicons (Carattoli *et al.*, 2005). The PBRT technique was further developed by denoting unique replicon sequences exclusively for Inc F plasmids (Villa *et al.*, 2010), then for Inc R, Inc U and ColE (García-Fernández *et al.*, 2009) and then for Inc X1 to X4 (Johnson *et al.*, 2012). However efficient the PBRT method is, it has had its limitations when considering the dynamic nature of DNA, which is subject to mutations. This means that new primers have to be redesigned to renew their specificity. Along with that problem is that this method only looks at one genetic locus (the replication gene) as an identification site for the replicon being typed. In addition to these issues, sequencing

technologies have been developed to a point that sometimes it is impractical to use PCR-oriented approaches for typing of many plasmids simultaneously. This aspect been addressed by the use of computer tools for plasmid analysis. The use of computer tools (programmes and software) makes it possible and easy to quickly analyse lots of sequence data that is constantly generated from high throughput DNA sequencing projects.

During early 2014, Carattoli and colleagues re-developed the replicon typing technique by identifying more loci. The additional loci were to be included for plasmid typing along with the replicon region for each incompatibility group. They called this scheme Plasmid multi-locus sequence typing (PMLST) (Carattoli *et al.*, 2014). This technique uses multiple genetic loci associated with a given plasmid incompatibility group for characterization, the technique has been specially made for *in-silico* application so that it is compatible with the analysis of virtual plasmid DNA sequence data such as with metagenomic sequence data.

2.4. BIOINFORMATICS ANALYSIS OF PLASMIDS

Bioinformatics is the study and analysis of biological data using computer based tools. It has been a major breakthrough discipline with various technologies that were formally developed to organise information on a large scale. These technologies gained popularity during the period of the human genome project (Luscombe *et al.*, 2001). It has great practical importance particularly when applied to DNA sequence analysis, thousands and even millions to billions of base pairs of sequence data cannot be arranged or deciphered without the relevant software. As applied to the study of plasmids, bioinformatics methods have been used to evaluate multiple DNA sequence data comprehensively by computational analysis (McArthur and Wright 2015). There is ever more increasing plasmid sequence data that gets deposited into online databases from different researches around the world. This data is a source of great un-deciphered information about the genomics of natural plasmids. More to this, sequencing technology is getting more efficient and techniques of environmental sampling of plasmids have taken a major stride with the development of metagenomics.

Metagenomics is the study of analysing a community of genomes directly from environmental samples without the need to culture. This approach has allowed the rudimentary practice of environmental DNA profiling without limitations, an important leap from the use of conventional microbiological techniques. The DNA data extracted using a

metagenomic technique is massive and it has greatly increased the scope of computational data analysis. It has revealed that 99% of the microbiome found in the environment has not been cultured by standard culture methods (Riesenfeld *et al.*, 2004). In metagenomics, molecular markers assist a gene centred approach to the identification of various antibiotic resistant gene loci from random sequences of the microbial genomes including extra chromosomal DNA elements like plasmids (Hugenholtz and Tyson, 2008). Plasmid identification in metagenomic data include recognition of contigs making up the different replicon regions associated with each plasmid under study.

Due to the immense increase in plasmid sequence data being deposited in online databases, *in-silico* analysis platforms have been developed to manage specific research applications for this data. Such research applications include the *in-silico* analysis of antibiotic resistance in plasmids as described in the Antibiotic Resistance Genes Database (ARDB) (Liu and Pop, 2009) and later in the Comprehensive Antibiotic Resistance Database (CARD) (McArthur *et al.*, 2013). The CARD has since been the popularised *in-silico* platform for AR analysis of both DNA sequence data and proteins due to the immense demand for its applications. This tool has currently been updated and expanded by new visualisation tools and powerful analysis models (Jia *et al.*, 2017).

In other applications, Carattoli and colleagues described the PlasmidFinder, an application for online typing of Inc/rep F, HI1, HI2, I1, N plasmids using the previously described PMLST technique (Carattoli *et al.*, 2014). Such plasmid databases although efficient as they may be, are disposed to not being consistently updated so they have limited data. However the National Centre for Biotechnology Information (NCBI) harbours almost all of the sequence data relevant for a comprehensive study on plasmids (Pruitt *et al.*, 2007).

CHAPTER 3: METHODOLOGY

3.1. DATA COLLECTION AND SCREENING.

The DNA sequence data collection and screening approach used in this project was intended to collect all the known plasmid sequences and to allow only the most relevant sequences from those collected to be used to answer the project objectives.

3.1.1. Source of DNA sequence data.

The plasmid DNA sequences are publicly available as part of the NCBI database resources. They were downloaded as a compressed tape-archive gunzip (tar.gz) file from the NCBI ftp directory, index of; (<ftp://ftp.ncbi.nlm.nih.gov/genomes/Plasmids/>). From this link the plasmids were contained in the file named “plasmids.all.fna.tar.gz”. During the course of this research, the file was retrieved in November 15 2015. It comprised a total of 6,090 plasmid sequences in fna (fasta nucleic acid) format and a sequence identification file “Plasmid.ids”. The identification file contained a list of 6,197 plasmids, this was 107 more than the total number of fna plasmid sequences in the compressed file. The list specified the plasmid host bacteria and corresponsive NCBI accession and gi numbers.

3.1.2. Plasmids sorted by complete against incomplete sequences

Plasmid sequences that have been submitted to the NCBI databases can either be complete or incomplete depending on the purpose for which the research was done. For the purposes of this research, only complete sequences were used for analysis. This is to confirm that the sequences have not been altered in any way, that is; the plasmid sequences are as complete as they exist in the environment that they were isolated from. As part of the annotation of names of plasmid sequences in the NCBI, complete plasmid sequences are distinguished from incomplete sequences by naming them with the keywords; “complete sequence”, “complete genome”, “complete plasmid” or “whole genome shotgun sequence”. On the other hand, all plasmid sequences denoted “partial sequence” are incomplete. Those which are not specified are just left without denoting them with either keyword. Screening for complete sequences required selecting plasmid sequences with the relevant keywords.

3.1.3. *Plasmid sorted by resistant against non-resistant sequences*

The comprehensive antibiotic resistance database (CARD) was used to select for antibiotic resistance genes from all complete bacterial plasmid sequences. For this analysis, all sequences were uploaded subsequently in groups of 25 to 100 (depending on the intensity of the resistance genes per group) into an analysis function in the CARD called the Resistance gene identifier (VERSION 2) online, available at; (<http://arpcard.mcmaster.ca/?q=CARD/tools/RGI>). The resistance gene identifier (RGI) function determines the open reading frames (ORF) of all genes in the plasmid sequences, after which it selects and classifies for those associated with antibiotic resistance in the plasmids from the NCBI database.

3.1.4. *Plasmid sorted by circular against linear sequences*

From the previous analysis, only plasmids with resistance genes were considered for further downstream analysis. Circular bacterial plasmids are selected for this research because they constitute the majority of the plasmid population and for that reason they are the prime candidates for spreading antibiotic resistance. This was done by cross referencing each of the plasmids in the NCBI, to determine if it was linear or circular.

3.1.5. *Circular antibiotic resistance plasmids sorted by genera of host bacteria*

These plasmids were sorted by the genera of bacteria, inclusive of which were a community of those plasmids which had been isolated from bacteria that was not cultured or obtained by non-culture based methods. These were denoted uncultured bacterium. The plasmids were sorted by host bacteria, by cross-referencing each plasmid's title name (of which each has the scientific name of the host) and placing each into a folder denoted by the bacteria genus then quantify how many belong to each genus of bacteria.

3.1.6. *Segregating resistance plasmids amongst the Gram stain reaction (positive or negative) of the host bacteria*

It is important to note the relative prevalence of resistance plasmids between Gram negative (G/-) and Gram positive (G/+) bacteria. Coliforms are renowned (G/-) bacteria that tend to be associated with pathogenicity and are considered indicator organisms. Their association with the increasing diversity of resistance plasmids remains a health concern. There are some (G/+) bacteria that are pathogenic and it is also important to note their prevalence in relation to resistance plasmids. Plasmids from uncultured bacteria were exempted from this analysis because the features of their host are unknown.

3.1.7. Analysis of multidrug resistance in a population of all circular resistance plasmids

This was done by manually interrogating the plasmids for resistance to more than one plasmid in the CARD analysis results. Multidrug resistance is the occurrence of more than one open reading frame coding for a resistance gene. This analysis is a measure of the abundance and diversity of resistant genes that are determined in each plasmid sequence.

3.1.8. Analysis of the range of sizes (molecular weight) of circular resistance plasmids.

Natural plasmids are available in sizes ranging from less than 1(Kbp) to over 2 000 (Kbp). It is important to know which plasmids in particular within this range of sizes are prone to confer resistance much more than others. In this analysis, the plasmids were grouped into clusters of the same sizes (+/- 100 Kbp) and from each cluster of plasmids, the distribution of resistant genes was determined. To determine the range of each individual plasmid, they were each displayed on the DNASTAR software MegAlign pro function (DNASTAR ® Lasergene 12, Madison, USA). After this, the last base in the sequence was determined on the molecular weight scale corresponding to the particular plasmid.

402

403 3.2. PLASMID SEQUENCE ANALYSIS

404

405 3.2.1. Analysis of incompatibility in circular resistance plasmids using PBRT.

406

407 Plasmids have a sequence specific replicon that describes their incompatibility, (Carattoli *et*
408 *al.*, 2005; García-Fernández *et al.*, 2009; Villa *et al.*, 2010). A molecular biology typing
409 technique has been developed that applies the use of these incompatibility group sequence
410 specific replicons *in-vitro* using a set of primers, this method is called plasmid PCR Based
411 Replicon Typing (PBRT).

412 This method can be applied *in-silico* using multi sequence alignment software from a given
413 population of plasmids. For this research a set of amplicons derived *in-silico* from reference
414 plasmids associated with each incompatibility group were developed. This was done by
415 determining the primer annealing site on each reference DNA sequence and extracting the
416 amplicon. Each of these amplicons was individually aligned against the set of circular
417 resistant plasmids selected for downstream analysis (described in clause 3.1.4.). The DNASTAR
418 software genomics suite was used for alignments. Amongst the alignment functions available,
419 the Mauve alignment function was selected from the Meg-Align pro utility. All positive hit
420 alignments with a consensus sequence determined by default settings of the mauve alignment
421 tool were considered plasmids belonging to the particular incompatibility group's reference
422 amplicon. These were then selected from the population of circular resistant plasmids and
423 sorted amongst their respective incompatibility groups.

3.3. DETERMINING PLASMID ECOLOGY BY REFERENCE ENVIRONMENT

Natural plasmids are found in various environments that can be generally classified into three different categories; natural environments, host associated environments and managed environments. For *in-silico* data, analysing this cannot be practically done on site from the isolation environment of such a plasmid, so this information has to be inferred from sequence annotations and documented information of plasmids published in NCBI and journal articles. In this analysis, the antibiotic resistance plasmids associated with each incompatibility group (typeable by PBRT) were classified into their different categories of environments. In this procedure, each individual plasmid was cross referenced with its NCBI nucleotide database (Refseq & Genbank) accession submission details. Some plasmids are designated an isolation source/host to make it easier for the user to determine where that plasmid sequence originated. In some cases some plasmid sequences are not assigned an isolation source or host upon entry into the database by a user. In such cases, the environments were cross referenced with corresponding literature. In instances where no specific source could be determined for a given plasmid in both interrogation procedures the plasmid was considered to have no documented isolation source for this study. These plasmids were excluded from the plasmids for which environments could be determined.

3.3.1. Description of the environment categories.

Natural environments: all open pristine environments that are not affected by anthropogenic activities e.g. natural soils and sediments, natural aquatic sources etc.

Host associated environments: all environments through which the isolated plasmids were associated with a bacteria or sample from a host organism, e.g. human and animal gastrointestinal tracts, skin, clinical samples etc.

Managed environments: all inanimate environments in which human activities are practiced e.g. waste water treatment plants (WWTP), acid mine drainage, households interior surfaces etc.

3.3.2. Association of Incompatibility Groups with Isolation Source (Reference Environment)

The plasmids belonging to each of the incompatibility groups were each interrogated and segregated into the environment categories depending on the reference isolation source. Chord diagrams were used to represent how plasmids of each incompatibility group are distributed in the environments described above. Chord diagram designs were based on an excel template freely available online: <https://sites.google.com/site/e90e50fx/home/talent-traffic-chart-with-chord-diagram-in-excel>. The corresponding Inc/rep groups have variable colour coding depending on each diagram. The size of the interconnecting lines (hierarchical edge bundles) is representative of the number of plasmids and the colour of the bundles indicates the environmental category colour code.

3.4. FREQUENCY ANALYSIS OF ANTIBIOTIC RESISTANCE GENES IN PLASMIDS

3.4.1. Antibiotic Resistance Gene Frequency in Incompatibility groups

In this analysis, the typeable plasmids associated with each incompatibility group/replicon (determined in clause 3.2.1) were segregated into their respective resistance profiles that were determined earlier using the CARD (clause 3.1.3). The antibiotic resistance gene frequency in each incompatibility group was initially quantified by determining the average resistance gene open reading frame (ORF) per plasmid per replicon/incompatibility group using the formula below.

$$\text{Average number of ORF per plasmid(Inc/rep group)} = \frac{\text{Total number of gene ORFs}}{\text{Total number of plasmids}}$$

This value would generally deduce the prevalence of multidrug resistance associated with each incompatibility group; (the greater the average factor, the greater the prevalence of multidrug resistance associated with the incompatibility group). The classes of antibiotic resistance genes were then determined within each replicon.

3.5. COMPARATIVE GENOMIC ANALYSIS OF ANTIBIOTIC RESISTANCE PLASMIDS

3.5.1. *Phylogenetic Analysis of Plasmid Incompatibility*

Phylogenetic trees of the plasmids per Inc/rep group were determined to elucidate the evolutionary history of the plasmids particularly the relationship of the Inc/rep groups amongst themselves. The outcome from this exercise would later be used to link demographic information with ecology or reference environmental isolation sources of the individual plasmids. Using the MegAlign function of the DNASTar software (DNASTar® Lasergene12), multiple sequence alignments (MSA) were generated from ClustalW with parameters set in slow and accurate alignment. All the plasmids within each incompatibility group were aligned from a common locus (the Amplicon target site). After MSA had been created amongst plasmids within each incompatibility group, they were then created for plasmids amongst all the incompatibility groups. In the alignment for all incompatibility groups, two plasmids from each incompatibility group that are most divergent from each other were selected. These plasmids represent the extremities of which all other plasmids within that incompatibility group are nested in between. Trees were generated and interrogated for relevant information by genealogical inference.

3.6. ANALYSIS OF REPLICON TYPING MARKERS.

3.6.1. Development of new Inc/rep group replicon typing markers

The primers developed using the PBRT for determining the replicons associated with each respective incompatibility group are not as sensitive as expected. This is prominent when analysing a large number of plasmids for a given incompatibility group. This is because the primer annealing regions of the plasmid tend to acquire mutations over time. The larger the plasmid number discovered per incompatibility group, the greater the likelihood of mutations that accumulate. As a result of this, the primer will not anneal properly to the target template region to an extent that the application for *in vitro* PCR method becomes limited. For these reasons, a primer diagnostics exercise was implemented in this project to develop or revise the sensitivity and efficiency of PCR primers developed initially. In this exercise, using the DNA star software, primers described in Table 3.1 were aligned against their respective plasmid sequences to highlight and extract the target sequence associated with each plasmid. Amongst the alignment functions available, the Mauve alignment function was selected from the Meg-Align pro utility. The target sequences were then aligned against each other and the primer annealing regions were denoted and inspected for mutations in the consensus sequence against the expected sequence. From the analysis of this data, degenerate primers were developed from the generic primer sequences in the case of substitution or indel mutations. In case of truncated amplicons, totally new primer sequences were created.

Table 3.1. Plasmid incompatibility replicon typing primers, reference sequences and Amplicon size.

	Incompatibility / Replicon	Replicon typing primers (5'- 3' direction)	Accession no.	Amplicon size (bp)
1	INC W ^a	FW: CCTAAGAACAACAAAGCCCCCG RV: GGTGCGCGGCATAGAACCGT	U12441	242
2	INC F(repB) ^a	FW: GATCGTTTAAGGAATTTTG RV: GAAGATCAGTCACACCATCC	AY234375	270
3	INC FIA ^a	FW: CCATGCTGGTTCTAGAGAAGGTG RV: GTATATCCTTACTGGCTTCCGCAG	J01724	462

4	INC FIB ^a	FW: GGAGTTCTGACACACGATTTTCTG RV: CTCCCGTCGCTTCAGGGCATT	M26308	702
5	INC FIC ^a	FW: GTGAACTGGCAGATGAGGAAGG RV: TTCTCCTCGTCGCCAAACTAGAT	AH003523	262
6	INC FII ^a	FW: CTGTCGTAAGCTGATGGC RV: CTCTGCCACAAACTTCAGC	AE006471	270
7	INC FIA ^b	FW: CCATGCTGGTTCTAGAGAAGGTG RV: GTATATCCTTACTGGCTTCCGCAG	J01724	462
8	INC FIB ^b	FW: TCTGTTTATTCTTTTACTGTCCAC RV: CTCCCGTCGCTTCAGGGCATT	M26308	683
9	INC FIBs ^b	FW: TGCTTTTATTCTTAACTATCCAC RV: CTCCCGGCGCTTCAGGGCATT	FN432031	683
10	INC FIC ^b	FW: GTGAACTGGCAGATGAGGAAGG RV: TTCTCCTCGTCGCCAAACTAGAT	AH003523	262
11	INC FII ^b	FW: CTGATCGTTTAAGGAATTTT RV: CACACCATCCTGCACTTA	AP000342	258–262
12	INC FII(repA) ^b	FW: CCAAACGTATTACCGCCAGGTAA RV: CGGTTACGTGACAGAATCATGCGC	AP000342	809
13	INC FII[Fd] ^b	FW: CTGATCGTTTTAAGGAATTTT RV: GGAATGTAGCACCCGAT	AM886293	700
14	INC FIIk ^b	FW: TCTTCTTCAATCTTGGCGGA RV: GCTTATGTTGCACRGAAGGA	CP000648	142–148
15	INC FIIs ^b	FW: CTAAAGAATTTTGATGGCTGGC RV: CAGTCACTTCTGCCTGCAC	AE006471	259–260
16	INC FIIy ^b	FW: TGGYAGGGAAGTGGTTCTG RV: GTRAGTCACACCTTCCCGC	CP001590	227
17	INC HI 1 ^a	FW: GGAGCGATGGATTACTTCAGTAC RV: TGCCGTTTCACCTCGTGAGTA	AF250878	471
18	INC HI 2 ^a	FW: TTTCTCCTGAGTCACCTGTTAACAC RV: GGCTCACTACCGTTGTCATCCT	BX664015	644
19	INC I 1 ^a	FW: CGAAAGCCGGACGGCAGAA RV: TCGTCGTTCCGCCAAGTTCGT	M20413	139
20	INC X ^a	FW: AACCTTAGAGGCTATTTAAGTTGCTGAT RV: TGAGAGTCAATTTTTATCTCATGTTTTAGC	Y00768	376
21	INC X1 ^d	FW: GCTTAGACTTTGTTTTATCGTT RV: TAATGATCCTCAGCATGTGAT	JN935897	461
22	INC X2 ^d	FW: GCGAAGAAATCAAAGAAGCTA RV: TGTTGAATGCCGTTCTTGTCCAG	JQ269335	679
23	INC X3 ^d	FW: GTTTTCTCCACGCCCTTGTTCA RV: CTTTGTGCTTGGCTATCATAA	JN935899	350

24	INC X4 ^d	FW:AGCAAACAGGGAAAGGAGAAGACT RV:TACCCCAAATCGTAACCTG	FN543504	570
25	INC L & M ^a	FW: GGATGAAAACATATCAGCATCTGAAG RV: CTGCAGGGGCGATTCTTTAGG	U27345	785
26	INC N ^a	FW: GTCTAACGAGCTTACCGAAG RV: GTTTCAACTCTGCCAAGTTC	NC_00329 2	559
27	INC Y ^a	FW: AATTCAAACAACACTGTGCAGCCTG RV: GCGAGAATGGACGATTACAAAACCTT	K02380	765
28	INC P ^a	FW: CTATGGCCCTGCAAACGCGCCAGAAA RV: TCACGCGCCAGGGCGCAGCC	M20134	534
29	INC A & C ^a	FW: GAGAACCAGAAAGACAAAGACCTGGA RV: ACGACAAACCTGAATTGCCTCCTT	X73674	465
30	INC T ^a	FW: TTGGCCTGTTTGTGCCTAAACCAT RV: CGTTGATTACACTTAGCTTTGGAC	K00053	750
31	INC K & B ^a	FW: GCGGTCCGAAAGCCAGAAAAC RV: TCTTTCACGAGCCCGCCAAA	AY234375	270
32	INC B & O ^a	FW: GCGGTCCGAAAGCCAGAAAAC RV: TCTGCGTTCCGCAAGTTCGA	M28718	159
33	INC R ^c	FW: TCGCTTCATTCCTGCTTCAGC RV: GTGTGCTGTGGTTATGCCTCA	DQ449578	251
34	INC U ^c	FW: TCACGACACAAGCGCAAGGG RV: TCATGGTACATCTGGGCGC	DQ401103	843
35	ColE1 ^c	FW: GTTCGTGCATACAGTCCA RV: GGCGAAACCCGACAGGACT	AM746977	187

524 ^a = Replicon primers developed by Carattoli *et al.*, 2005, ^b = Replicon primers developed by
525 Villa *et al.*, 2010, ^c = Replicon primers developed by García-Fernández *et al.*, 2009, ^d =
526 Replicon primers developed by Johnson *et al.*, 2012.

CHAPTER 4: RESULTS AND DISCUSSION

4.1. Plasmids sorted by domain of host organism.

From sorting the data, the compressed file was found to contain 6,090 plasmids that belonged to 429 genera of organisms from all the 3 domains of life. An additional 56 plasmids were included from uncultured bacteria, these plasmids were not part of the compressed file and they were downloaded separately from NCBI. The total number of plasmids was 6,146. These plasmids were distributed throughout the domains of life as follows; 45 (1%) of the plasmids were isolated directly from organisms of domain eukaryota, these were mainly fungi, amoeba and some plants; most of them are usually found in the plastids of eukaryotic organisms such as chloroplasts and mitochondria. 150 (2%) of the plasmids were isolated from the domain archaea. The remaining majority of the plasmids were from domain bacteria and these constituted 5951 (97%), within the compressed file (Figure 4.1).

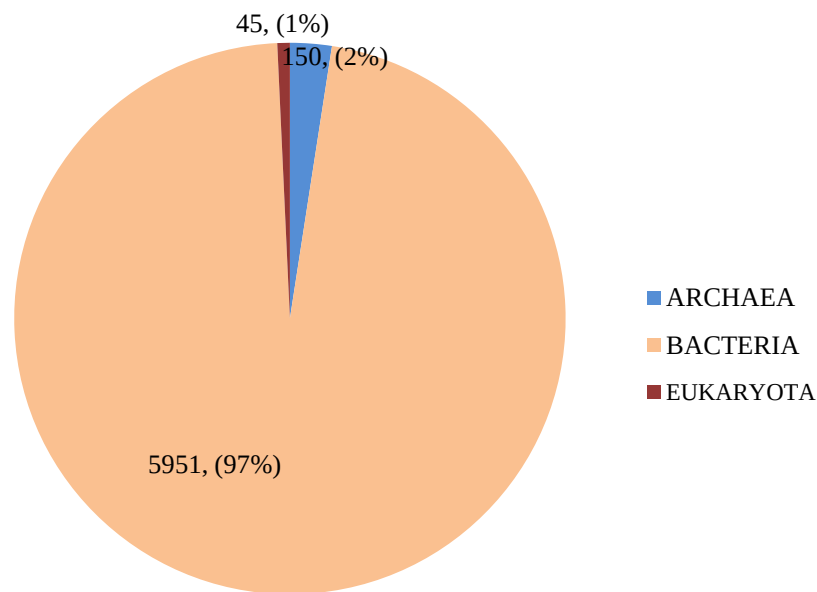


Figure 4.1. Plasmids sorted by domain of host organism.

4.2. Plasmids sorted by complete against incomplete sequences.

After the previous sorting there were a total of 5897 bacterial plasmids in the compressed file, 27 (0.4%) of the plasmids were incomplete sequence, 5598 (95%) were complete sequences and 272 (5%) were not specified, which were denoted “others” in the representative pie chart; (Figure 4.2).

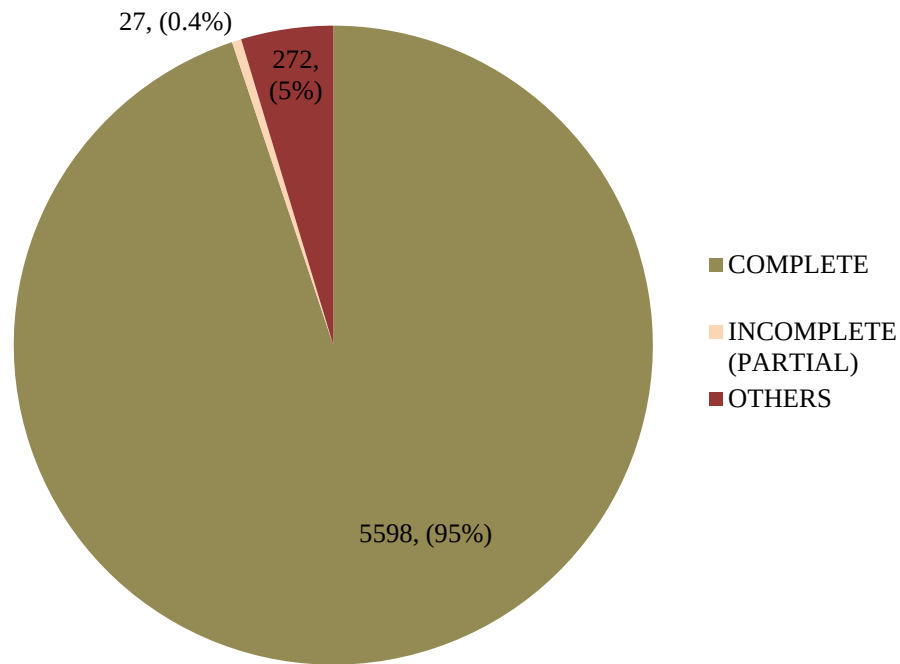


Figure 4.2. Plasmids sorted by complete against incomplete sequences.

4.3. Plasmids sorted by resistant against non-resistant sequences

A total of 5598 plasmids from the previous analysis were increased to a total of 5682 plasmids with 84 more manually downloaded plasmid sequences. The plasmid sequences were not part of the original compressed file but they appeared in the sequence identification file denoting all known bacterial plasmids that was part of the compressed file. All of the manually downloaded plasmids were from uncultured bacteria and they were isolated via metagenomic (non-culture based) techniques. From this new total (5682), the sequences were screened for antibiotic resistance. The analysis showed that 4276 (75%) plasmids were found to be non-resistant and 1404 (25%) were resistant plasmids (Figure 4.3).

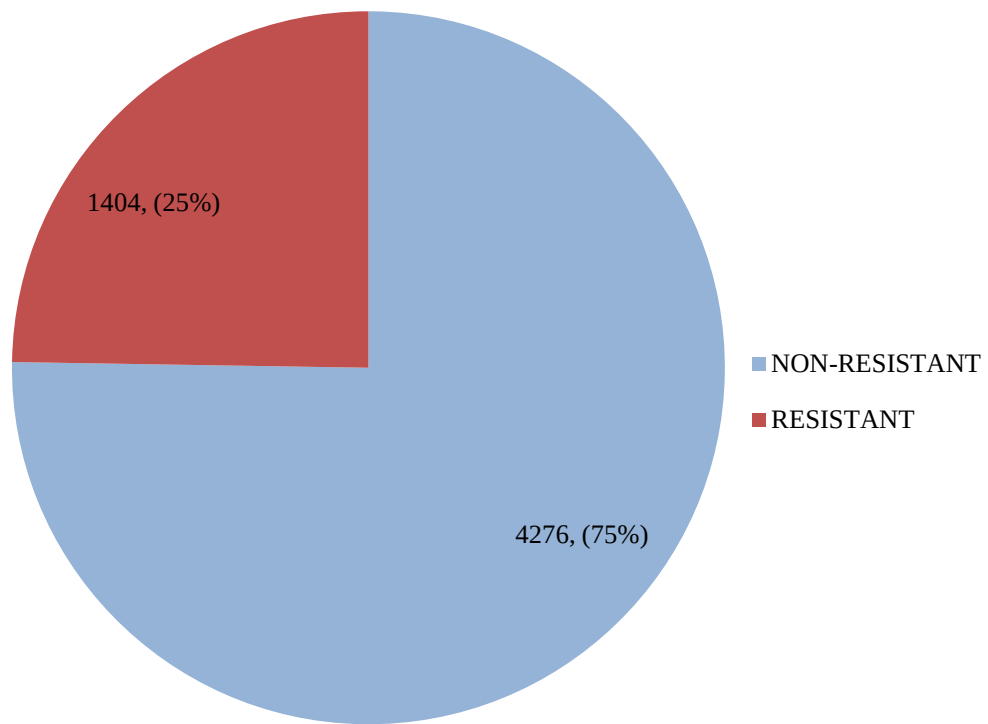


Figure 4.3. Plasmids sorted by resistant against non-resistant

4.4. Plasmids sorted by circular against linear sequences.

The 1404 resistant plasmids were sorted by circular resistant plasmids against linear resistant plasmids. From this screening, 59 (4%) of the plasmids were found to be linear and 1345 (96%) were found to be circular; (Figure 4.4). Only the circular resistant plasmids were considered for downstream analysis.

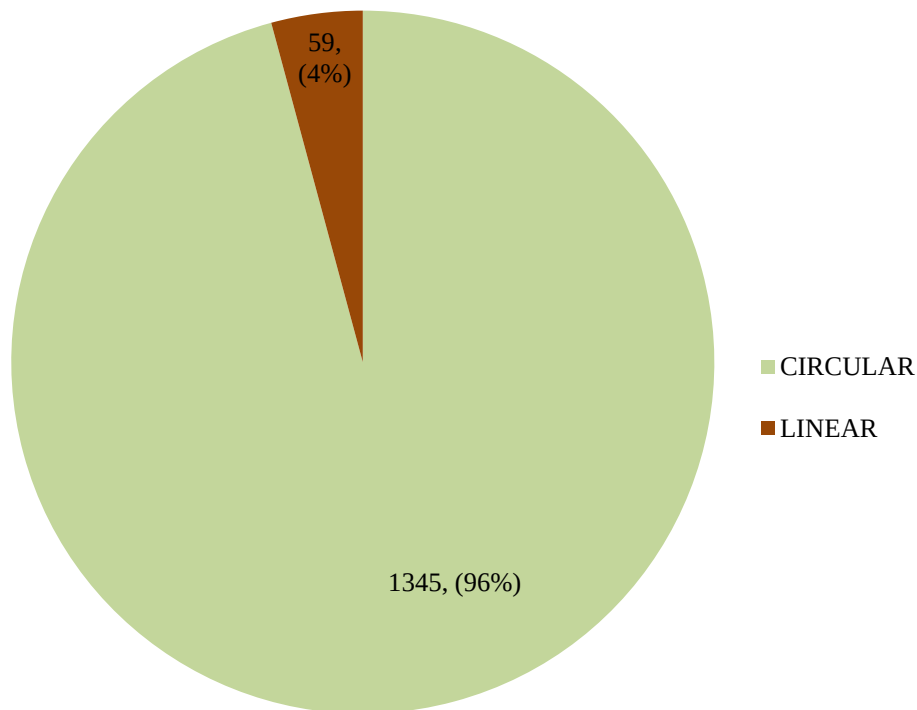


Figure 4.4. Plasmids sorted by circular against linear sequences

4.5. Plasmids sorted by genera of host bacteria.

The total number of circular resistance plasmids was 1345, these plasmids were sorted by genera of host bacteria. From characterizing plasmids into host bacteria, it was found that the majority of plasmids occur in *Escherichia spp* at 198 (15%) followed by *Klebsiella spp* at 186 (14%), *Staphylococcus spp* at 156 (12%) and *Salmonella spp* at 7% (90). The remaining 713 (52%) plasmids are distributed in genera for which each genus constitutes less than 5%, (Figure 4.5).

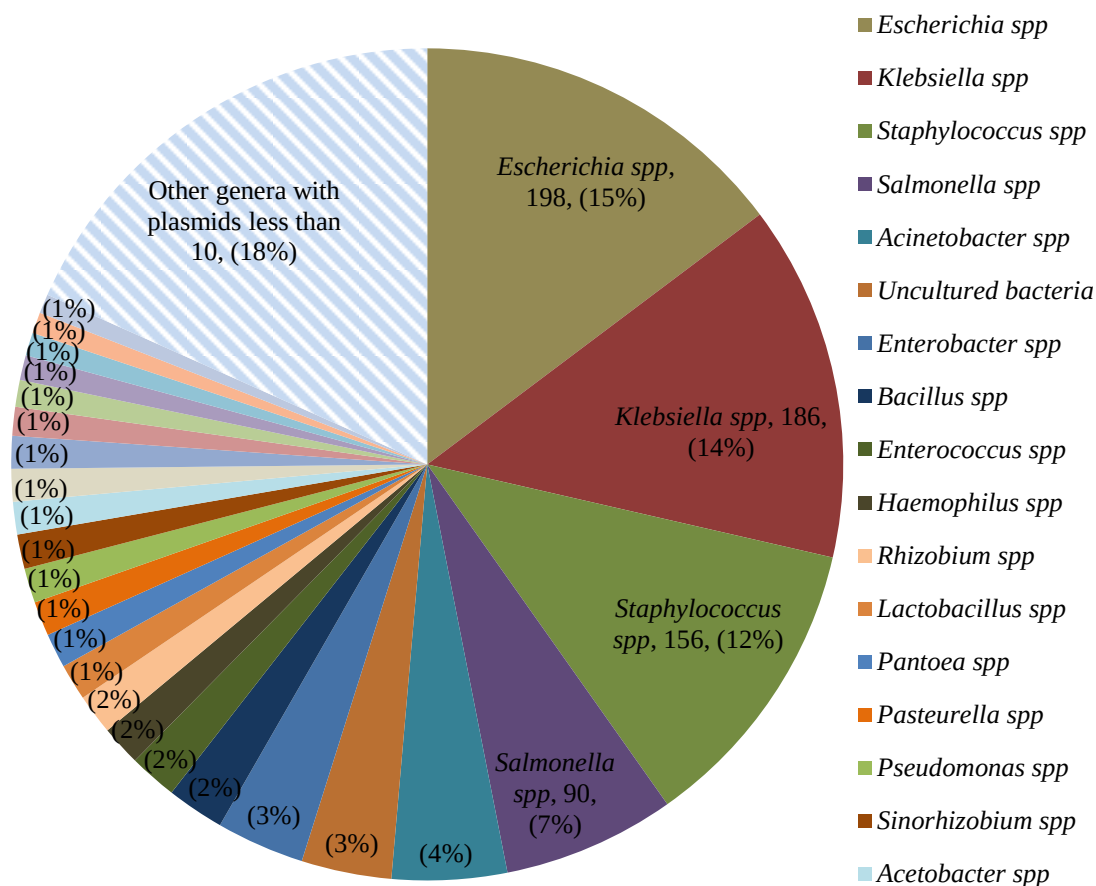


Figure 4.5. Circular resistant plasmids sorted by bacterial species

4.6. Segregating resistant plasmids by the Gram stain reaction (positive or negative) of the host bacteria

Plasmids from uncultured bacteria were exempt of this analysis because the features of their host are unknown. These plasmids were 47 in total, the new total for this analysis was reduced to 1298. Of the 1298, after the analysis; 991 (76%) were resistance plasmids from Gram negative bacteria and 307 (24%) were resistant plasmids from Gram positive bacteria; (Figure 4.6)

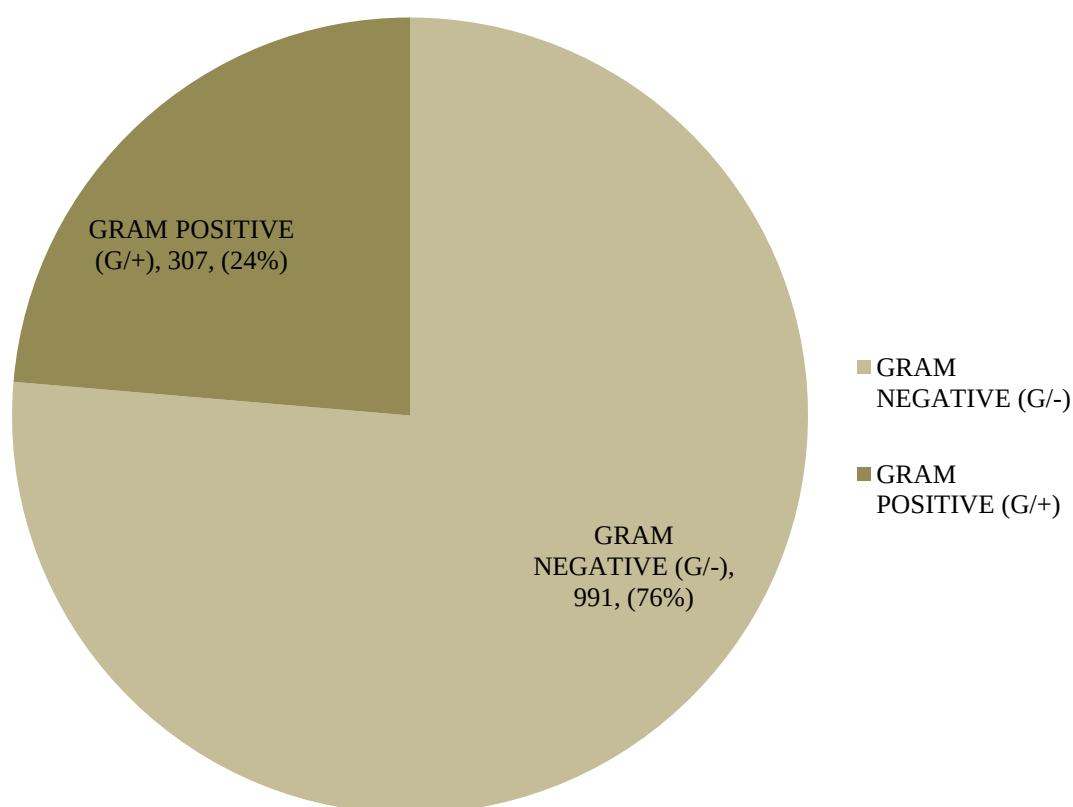


Figure 4.6. Circular resistant plasmids sorted by Gram stain reaction of host bacteria

4.7. Analysis of multidrug resistance in a population of all circular resistance plasmids

The 1347 circular resistant plasmids were sorted by their degree of multidrug resistance by further interrogating the Comprehensive Antibiotic Resistance Database results. This is a measure of the number of resistance genes per resistance plasmid, indicative of multidrug resistance. The results are represented in the pie chart in Figure 4.7. The chart represents the percentage of antibiotic resistance plasmids with varying numbers of antibiotic resistance gene(s) i.e. (1 ARG, 2 ARG etc.) in the entire population of resistance plasmids. From this analysis, plasmids with 1 ARG per plasmid constituted 617 (46%), followed by 2 ARG per plasmid at 227 (17%), 3 ARG per plasmid at 127 (9%) and 4 ARG per plasmid at 78 (6%), etc., (Figure 4.7).

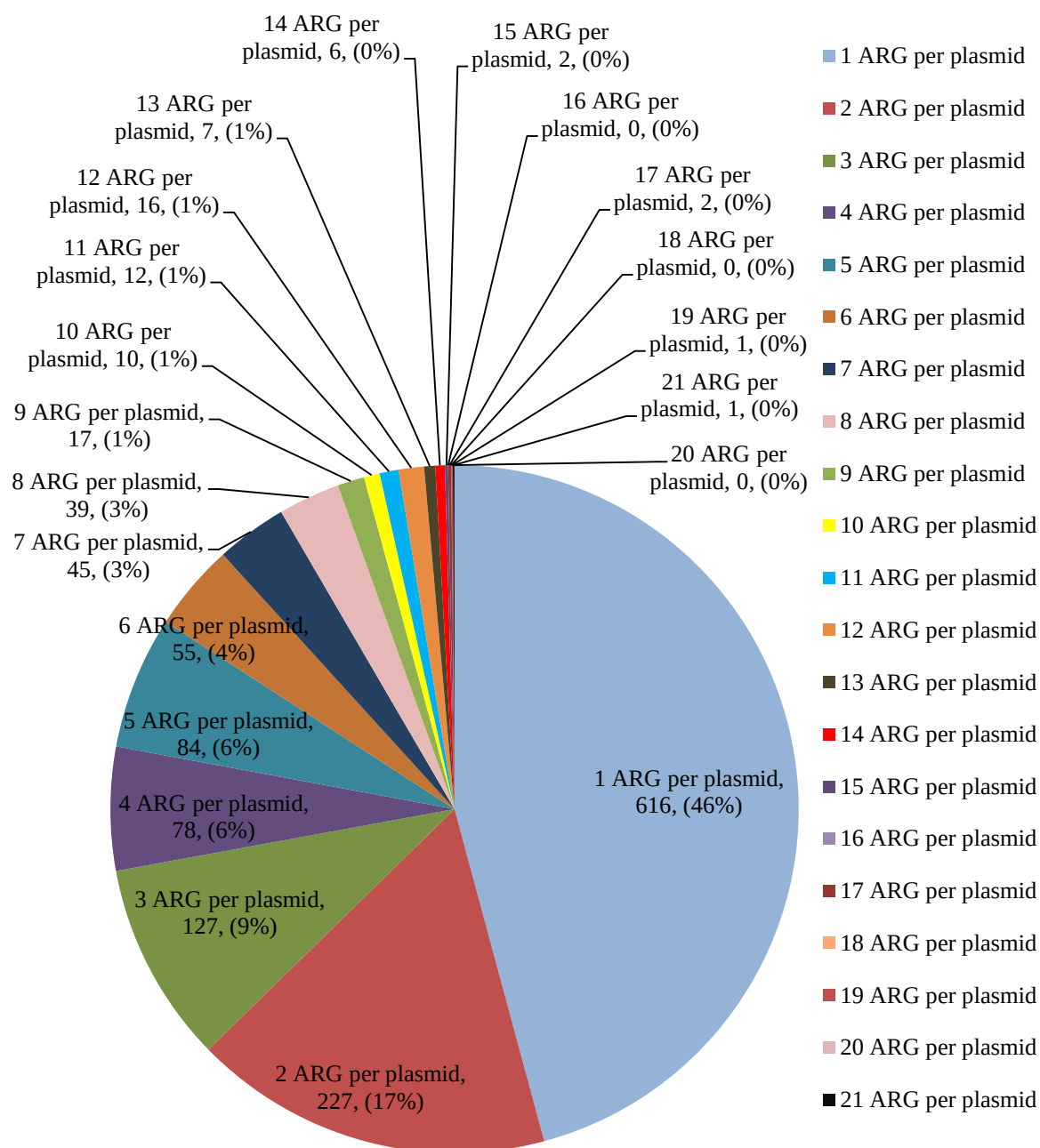


Figure 4.7. Plasmids sorted by multidrug resistance in a population of 1345 circular resistance plasmids (Data in graph form at Appendix C1)

4.8. Analysis of the range of sizes (molecular weight) of circular resistance plasmids.

With regards to genome size, the retrieved plasmid dataset had plasmids ranging from 537 bp (*Xanthomonas campestris* plasmid unnamed-p7) to 2,580,080 bp (*Cupriavidus metallidurans* CH34 megaplasmid). The majority of plasmids 871 (65 %) range between 1 – 100 Kbp, this was followed by plasmids ranging between 100 – 200 Kbp at 260 (19%), then 200 – 300 Kbp at 74 (5%) and 300 – 400 Kbp at 39 (3%), etc., (Figure 4.8).

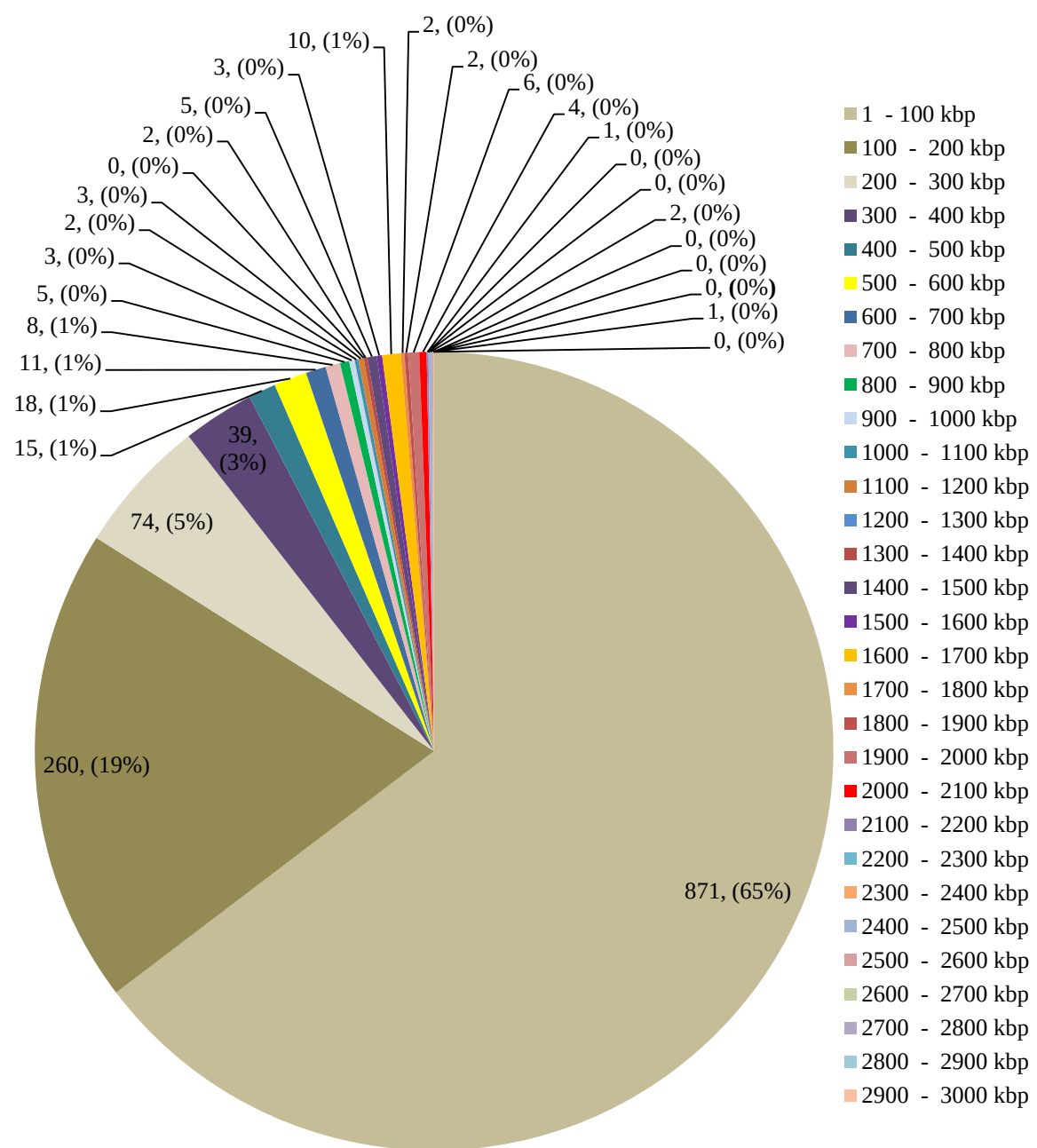


Figure 4.8. Distribution of resistance plasmids (%) according to the size in molecular weight (Kbp) (Data in graph form at Appendix C2).

629

630 Natural plasmids have been found to have two distinctive densely populated size ranges,
631 when represented in a graph they represent two peaks. One peak is observed at the 1-10 Kbp
632 region with 279 plasmids and the other at 11-100 Kbp region with 592 plasmids. These two
633 peaks had been earlier described by Smillie and colleagues (2010) and affirmed later by
634 Shintani and colleagues (2015) when the population of plasmids available online NCBI had
635 increased from 1730 to 4602 plasmids.

636 Bacterial plasmids seem to be more abundant in Gram-negative bacteria as opposed to Gram-
637 positive bacteria, The majority of these are reported to be from the phylum proteobacteria
638 (Shintani *et al.*, 2015) particularly from bacteria *Escherichia coli* and *Klebsiella spp.* Alone,
639 these two species make up 29% of the plasmid population, (Figure 4.5). These findings do
640 not necessarily mean that natural plasmids are mostly abundant in these two bacterial species
641 over the rest. The reason there is such extensive plasmid sequence data found in these two
642 species in particular is because they are the most studied in part due to their implications on
643 public health.

4.9. Analysis of ARGs frequency within the range of sizes (Kbp) of circular resistance plasmids.

A graphical representation of the number of quantifiable ARGs in circular resistance plasmids show that the quantity of resistance genes decreases steadily at a logarithmic rate between 1 and 1000 Kbp, suggesting that as the plasmid size increases the number of resistance genes decreases. A possible explanation for this observation is that smaller plasmids are more mobile than larger plasmids as such they are likely to acquire resistance genes as a consequent adaptive function for the host. However, in much larger plasmids between 1100 Kbp and 2600 Kbp there is variation in the number or ARGs relative to increase in the size, (Figure 4.9).

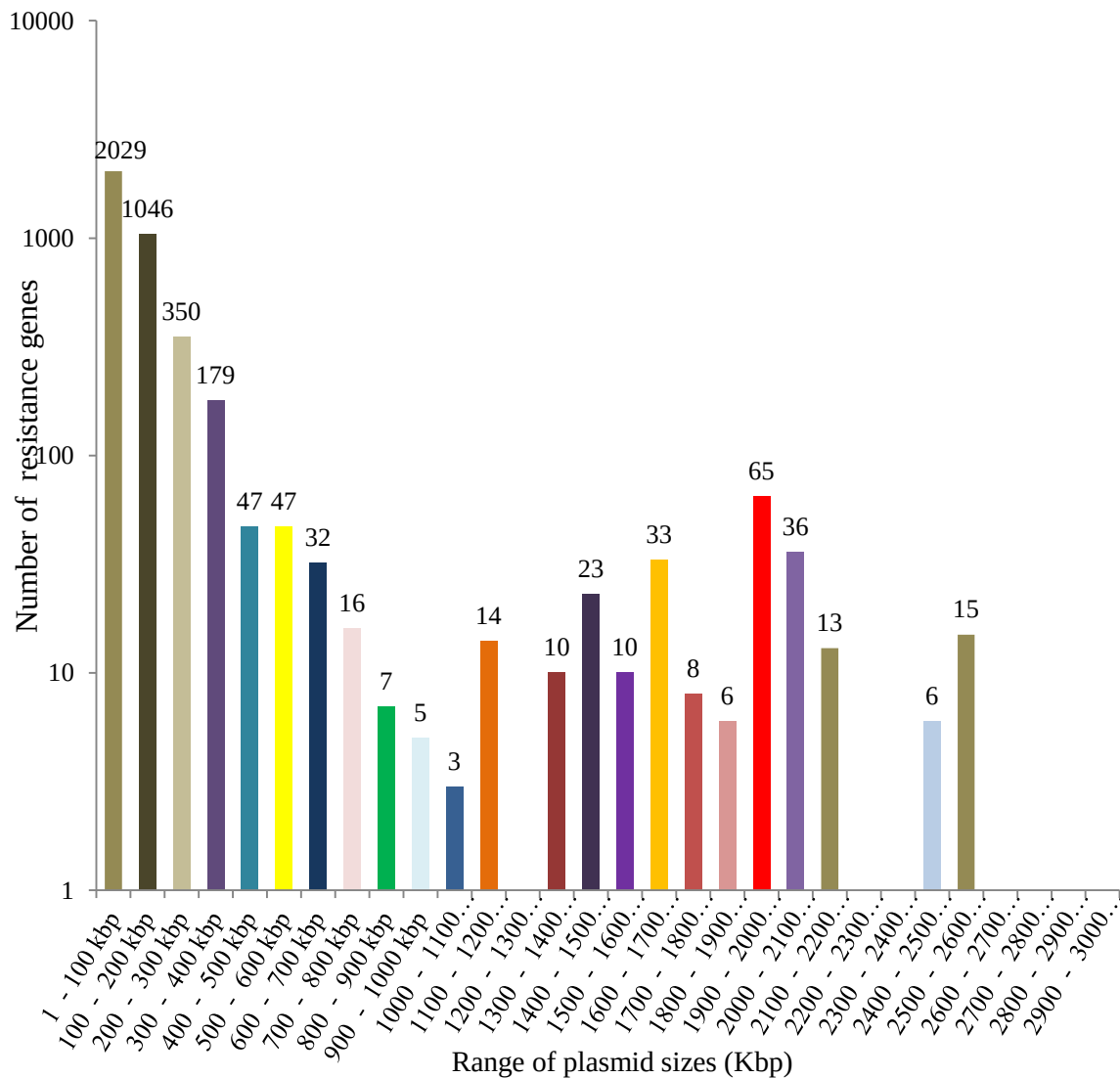


Figure 4.9. Analysis of resistance genes within the range of sizes (Kbp) of circular resistance

5.0. Analysis of the different classes of antibiotic resistance genes found in all resistance plasmids

The distribution of the gene classes attributed to resistance in all resistance plasmids was determined. The results reveal that majority of the plasmids carry the antibiotic efflux mechanism (30%), and resistance to Beta lactam (27%) and Aminoglycosides (21%) classes of antibiotics, (Figure 5.0a).

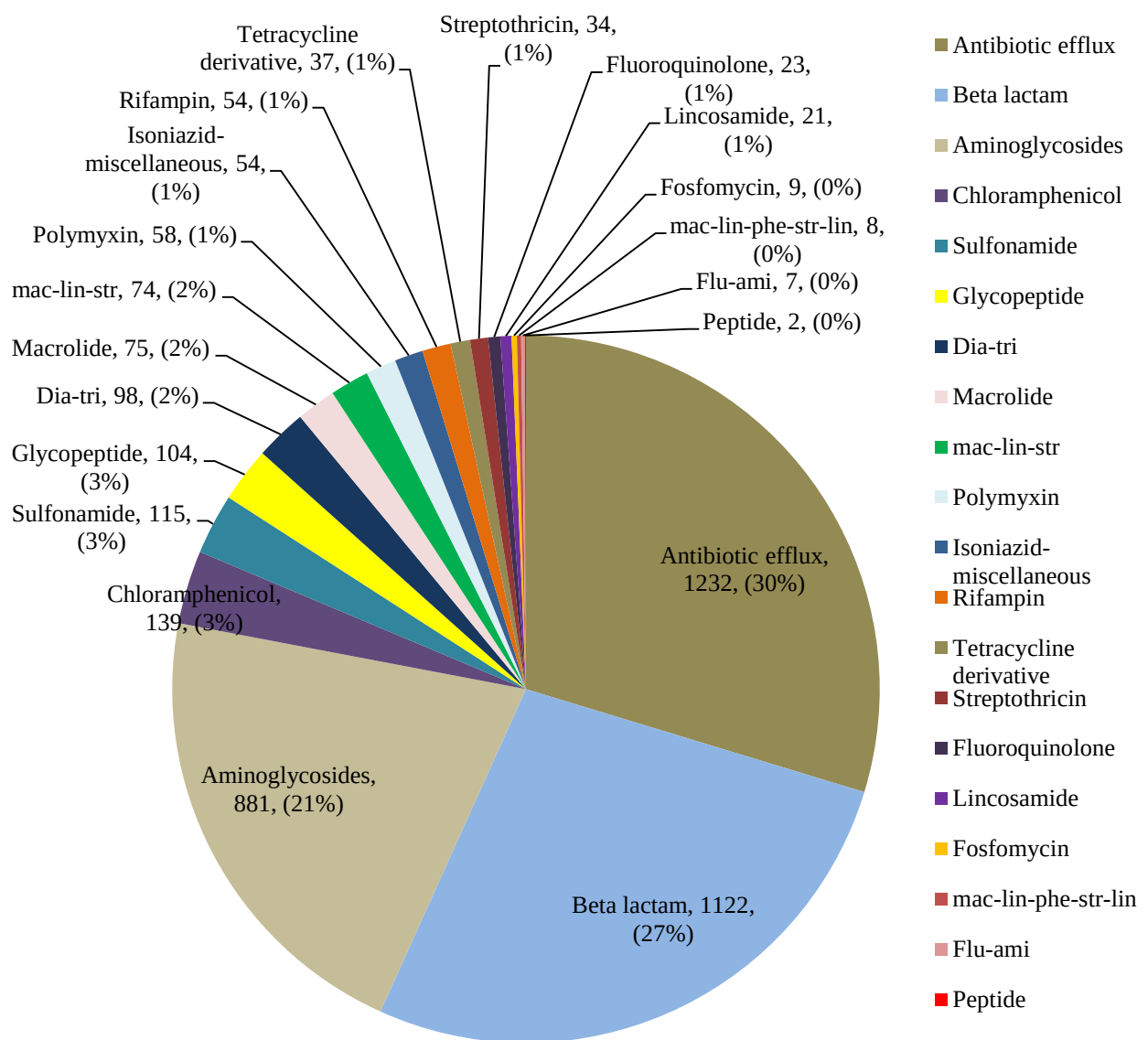
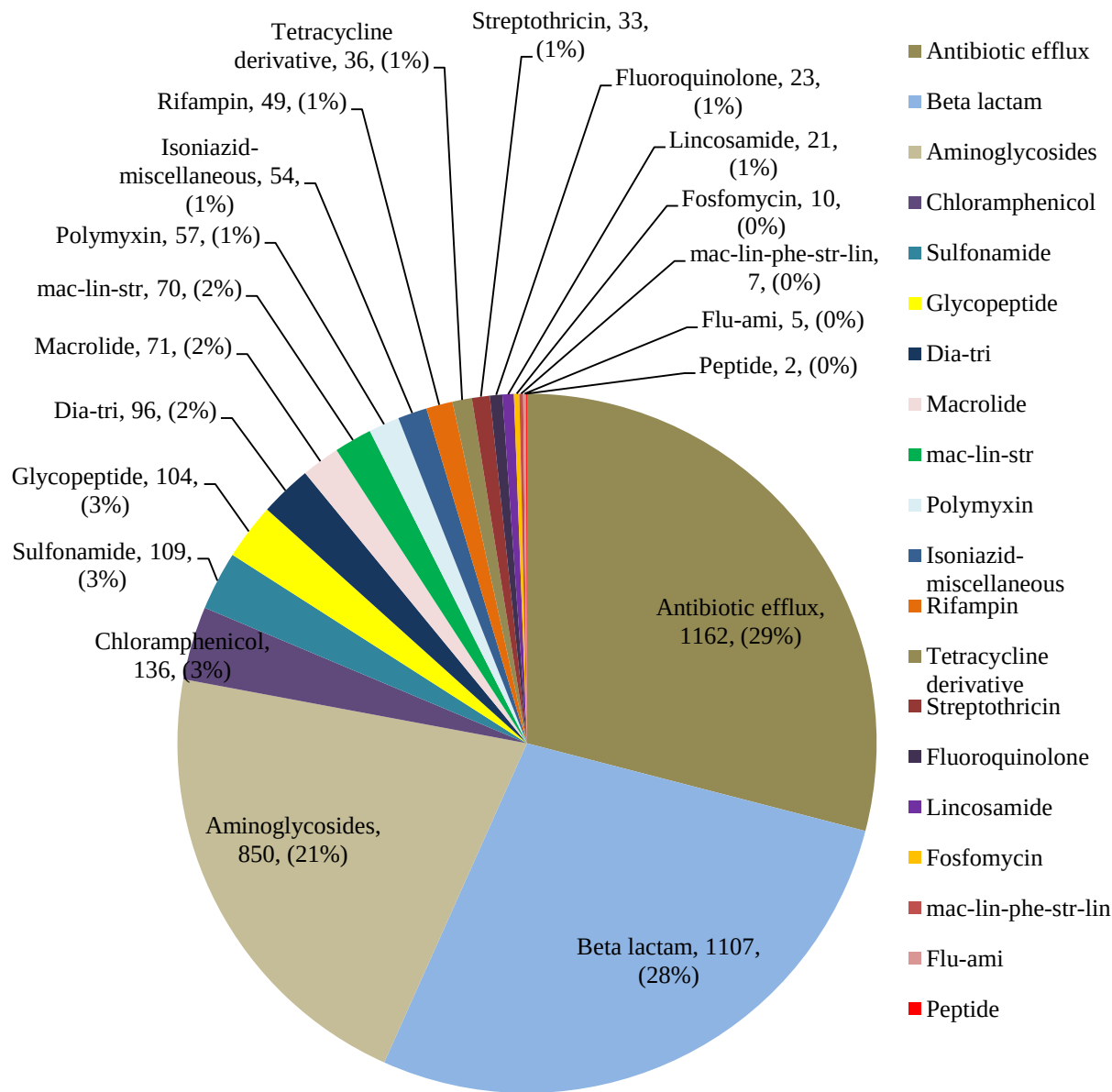


Figure 5.0a. Classes and relative quantity of antibiotic resistant genes found in all resistant plasmids (Data in graph form; Appendix C3)

669 In addition, considering only circular plasmids, a similar pattern is observed with the
 670 antibiotic efflux mechanism (29%), together with resistance mechanisms for Beta-lactams
 671 (28%) and Aminoglycosides (21%) dominating over others (Figure 5.0b).



672
 673 Figure 5.0b. Classes and relative quantity of antibiotic resistant genes found in circular
 674 resistant plasmids

5.1. Plasmid replicon typing

The PBRT approach for plasmid typing proved to work efficiently in extracting plasmids intended for use in this study. However the caveat of this method is that there are limited replicon reference sequences to type all known plasmids with resistance. In this study, the technique managed to identify only 460 (35%) of resistance plasmids belonging to *Enterobacteriaceae*, the remaining 885 (65%) could not be typed, (Figure 5.1a).

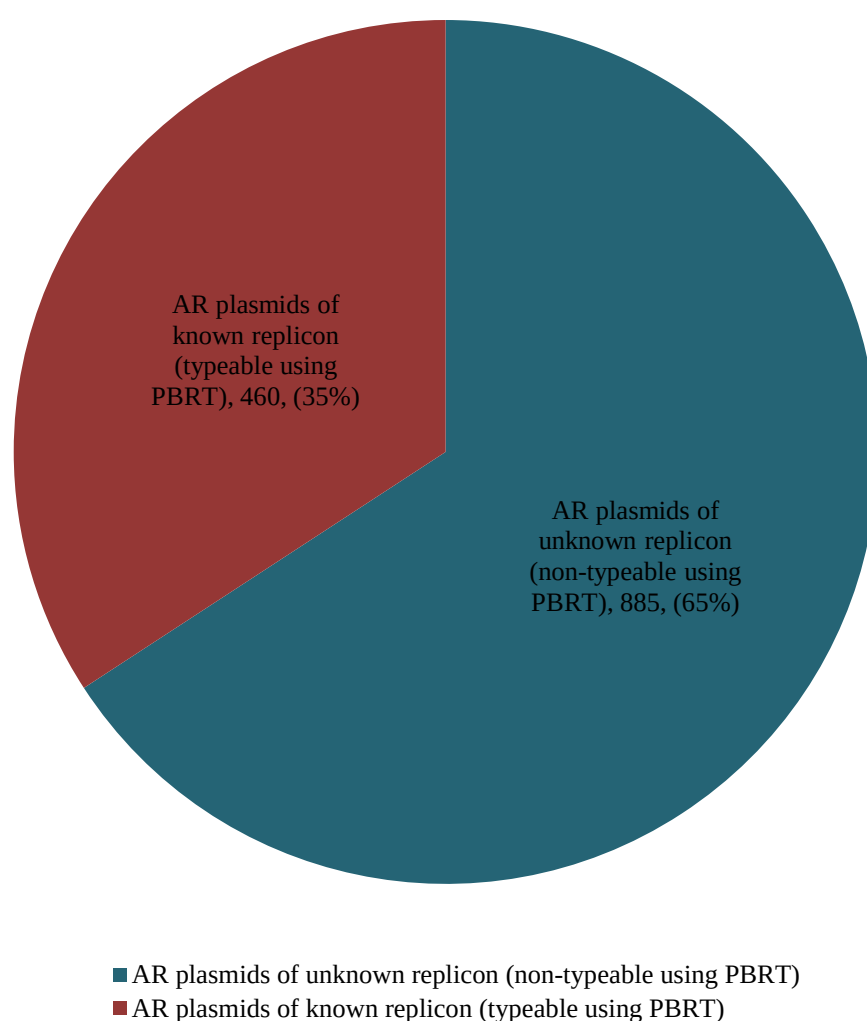


Figure 5.1a. Antibiotic resistant plasmids sorted by success in typing using PBRT

The *in-silico* detection of plasmids belonging to *enterobacteriaceae* has been undertaken by Caratolli and colleagues in 2014 using an online typing tool they developed known as PlasmidFinder. Their results were from a dataset of 559 plasmids corresponding exclusively

to *Enterobacteriaceae*. They showed that 263 (47%) out of 559 were readily typeable using the PBRT technique, the remaining 296 were not detected by the PBRT probes (Caratolli *et al* 2014). These results are somehow in agreement with the limitation of the efficiency of the PBRT probes themselves in identifying Inc/rep regions readily, despite this, Caratolli's group resorted to other means for identifying Inc/rep regions namely pMLST. This is because the pMLST technique characterises more than one locus in a plasmid to allocate it as belonging to a particular replicon hence it is more efficient in characterising replicons than conventional PBRT for replicon typing. From their findings they noted 77 novel replicon sequences.

5.2. Characterization of multiple replicon AR plasmids

Upon characterizing the distribution of multiple replication genes (replicons) in plasmids, It was determined that 124 (27%) plasmids have more than one Inc/rep region that could be determined by the PBRT probes whilst 336 (73%) of plasmids had only 1 Inc/rep region (Figure 5.2a). A further characterization of the exact number of replicons per plasmid within the distribution was determined. From this, 33 (7%) plasmids had 2 replicons, 55 (12%) plasmids had 3 replicons, 18 (4%) plasmids had 4 replicons and 4 (1%) plasmids had 5 replicons etc., (Figure 5.2b).

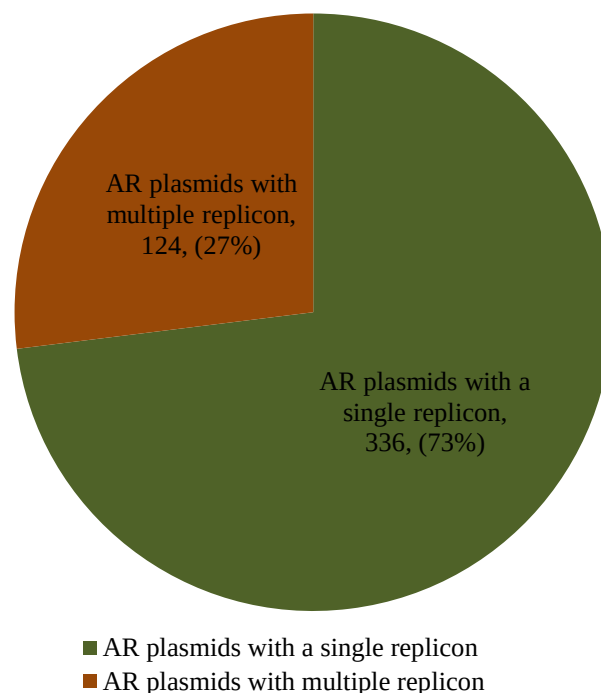


Figure 5.2a. Antibiotic resistant plasmids sorted by single against multiple replicons

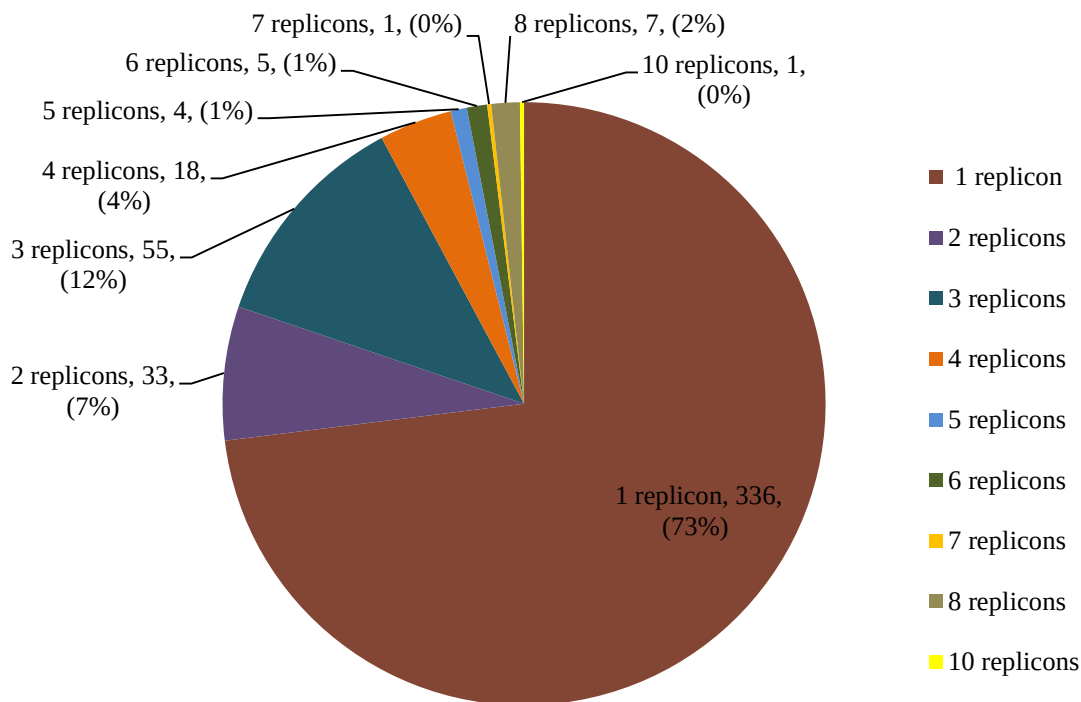


Figure 5.2b. Antibiotic resistant plasmids sorted by number or typeable replicons

Natural plasmids are known to have multiple replicons (Couturier *et al.*, 1988). This research also showed that some plasmids that were typeable by the PBRT technique have multiple replicons. The most abundant multi-replicon Inc/rep group are Inc F's, most Inc F plasmids had replicons of other Inc F subgroups. A majority of these multi-replicon plasmids have the FIA and FIB replicons, which are reported in literature to represent the binding site for replication protein A (rep A), (Villa *et al.*, 2010;Kline 1985). The second most abundant multi replicon group of plasmids are Inc X's and group subtypes (X1 to X4). The typing of these plasmids is based on different markers (other than the replicon) which were developed by Johnson and colleagues based on the *taxC* gene locus. This gene was demonstrated to have 92 - 99% sequence homology within the X subtypes (X1-X4) (Johnson *et al.*, 2012). As a result typing the replicon with the probes in the dataset identified similar plasmids repeatedly for IncX1 through X4. Furthermore, Inc X1 and X2 were identical plasmids and had to be grouped together because they were probing a single locus at the same position. Inc K&B, Inc B&O and Inc I1 had plasmids which displayed multi-replicons of the three Inc/rep groups

or were closely related. These plasmids also had close association in phylogeny indicating that they share a common ancestral plasmid lineage.

5.3. Incompatibility/replicon sorting of typed plasmids

There were 465 plasmids that could be typed as already established, from this population 115 (24%) was represented by the Inc F incompatibility group. The other groups were the ColE1 at 17% (83), Inc A&C 58 (12%), Inc N 44 (9%) and Inc I1 40 (8%); (Figure 5.3a). The dominant group, Inc F plasmids were further divided into subgroups, the Inc F(rep B) replicon occurred in 74 plasmids, followed by Inc FII(rep A) in 51 plasmids and Inc FII in 50 plasmids; (Figure 5.3b).

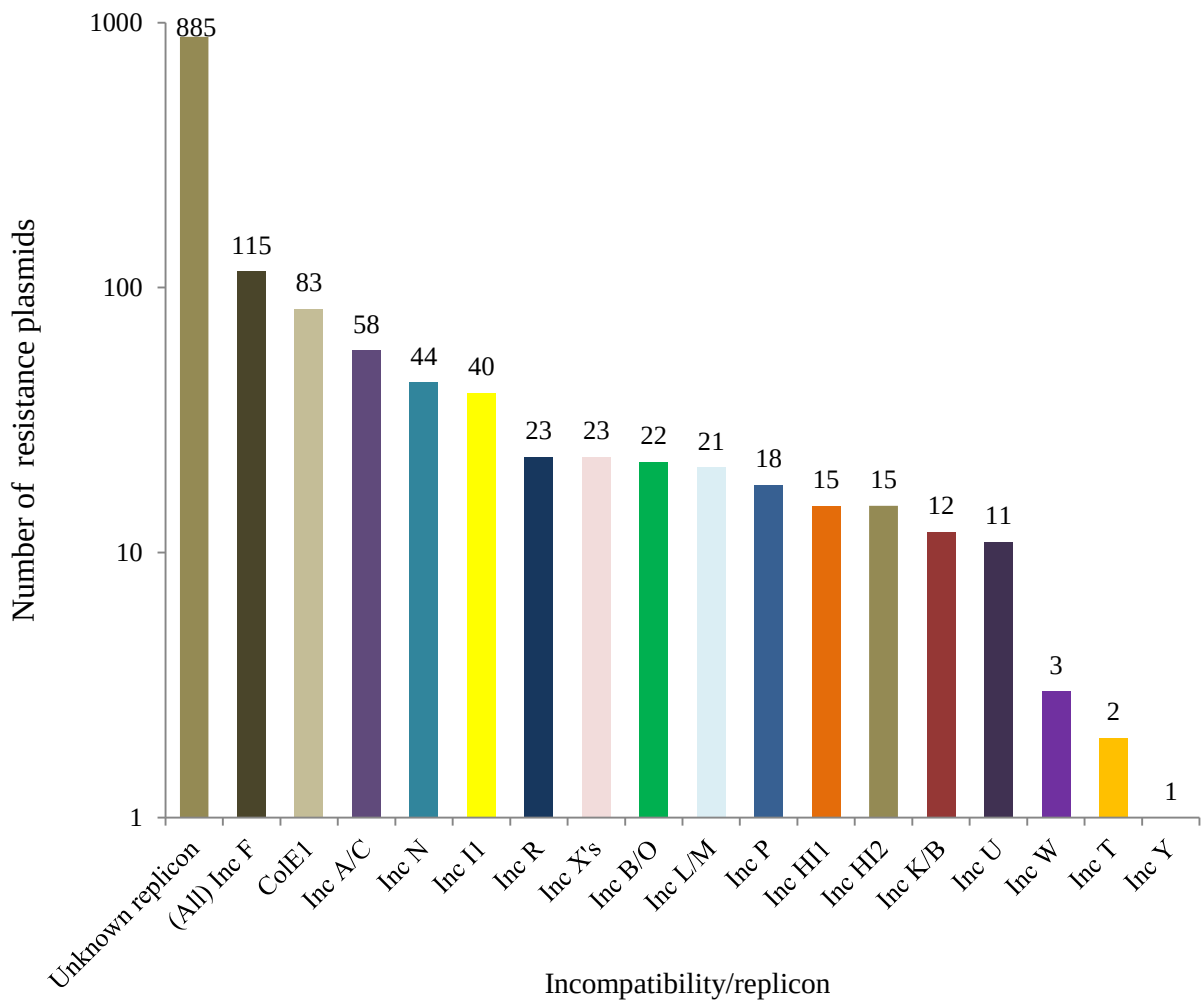


Figure 5.3c. Antibiotic resistant plasmids (typeable using PBRT) sorted by Inc/rep region (Data represented in pie chart; Appendix C5)

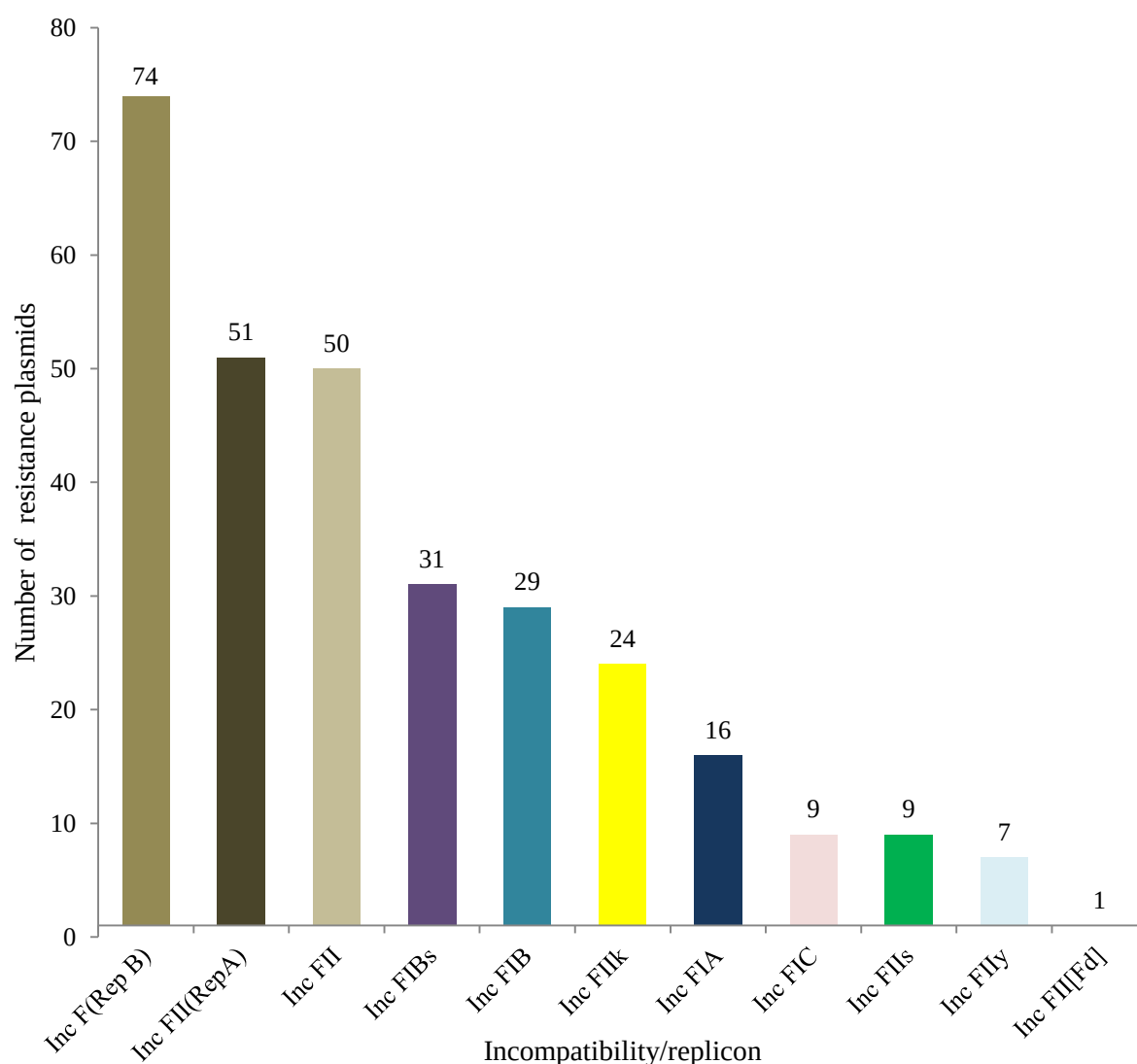


Figure 5.3d. Quantifying Inc F resistant plasmids into subgroups

The above results are not surprising because Inc F plasmids are the most characterised and studied and as such represent the bulk of known plasmids. The second most abundant were ColE1 like plasmids, which are small narrow host range (NHR) group of plasmids that tend to have very high copy numbers in a bacterium. Because of their size, they tend to be ubiquitous in the nature. Caratolli and colleagues produced somewhat similar findings, Inc F was most abundant, compared to Inc X, Inc N, Inc I1 and Inc A/C (Caratolli *et al* 2014). However, the only contrast from the findings of this research compared to that of Caratolli is that they did not type ColE1 and from their statistics. Inc X group and subgroups were second most abundant to Inc F's.

5.4. The frequency of specific ARGs as determined in different plasmid incompatibility groups.

5.4.1. Inc W (AR profile)

The Inc W plasmids were most characterized by high numbers of ARGs to aminoglycosides (56%), followed by sulfonamides (22%), beta-lactams and efflux pumps (11%) (Figure 5.4a). The corresponding Table 5.4a shows the Inc W ARG frequency determined from the population size of 3 Inc W plasmids, the total number of ARG ORF determined by CARD is 9, and the classes of resistance genes found in Inc W is 4. The calculated ARG frequency within the Inc W plasmids was found to be 3.

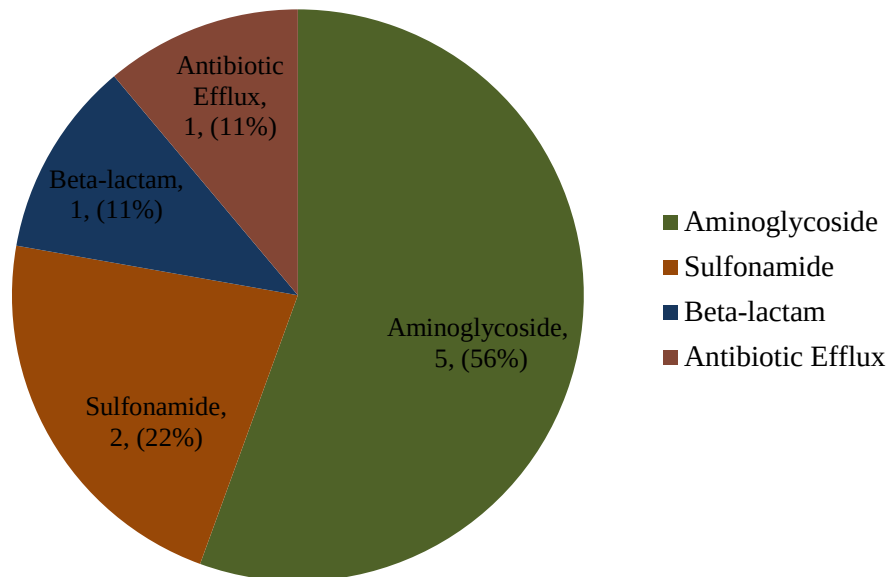


Figure 5.4a. Classes of Antibiotic resistance genes in Inc W

Table 5.4a. Inc W (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
3	9	4	9/3 = (3)

*; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.2. *Inc F's (AR profile)*

When examining *Inc F's* plasmids, beta-lactams had highest number of ARGs (36%), followed by aminoglycosides (21%), efflux pumps (18%) and sulfonamides (8%); (Figure 5.4b). The antibiotic resistance frequency of *Inc F* plasmids determined from a population size of 144 *Inc F* plasmids has 451 resistance genes sequences (ARG ORF determined by CARD) which. The classes of resistance genes found in *Inc F's* is 14. This information supplements the *Inc F's* pie chart and is used to calculate the average number of resistance genes per plasmid within *Inc F's* which is determined to be 3.95.

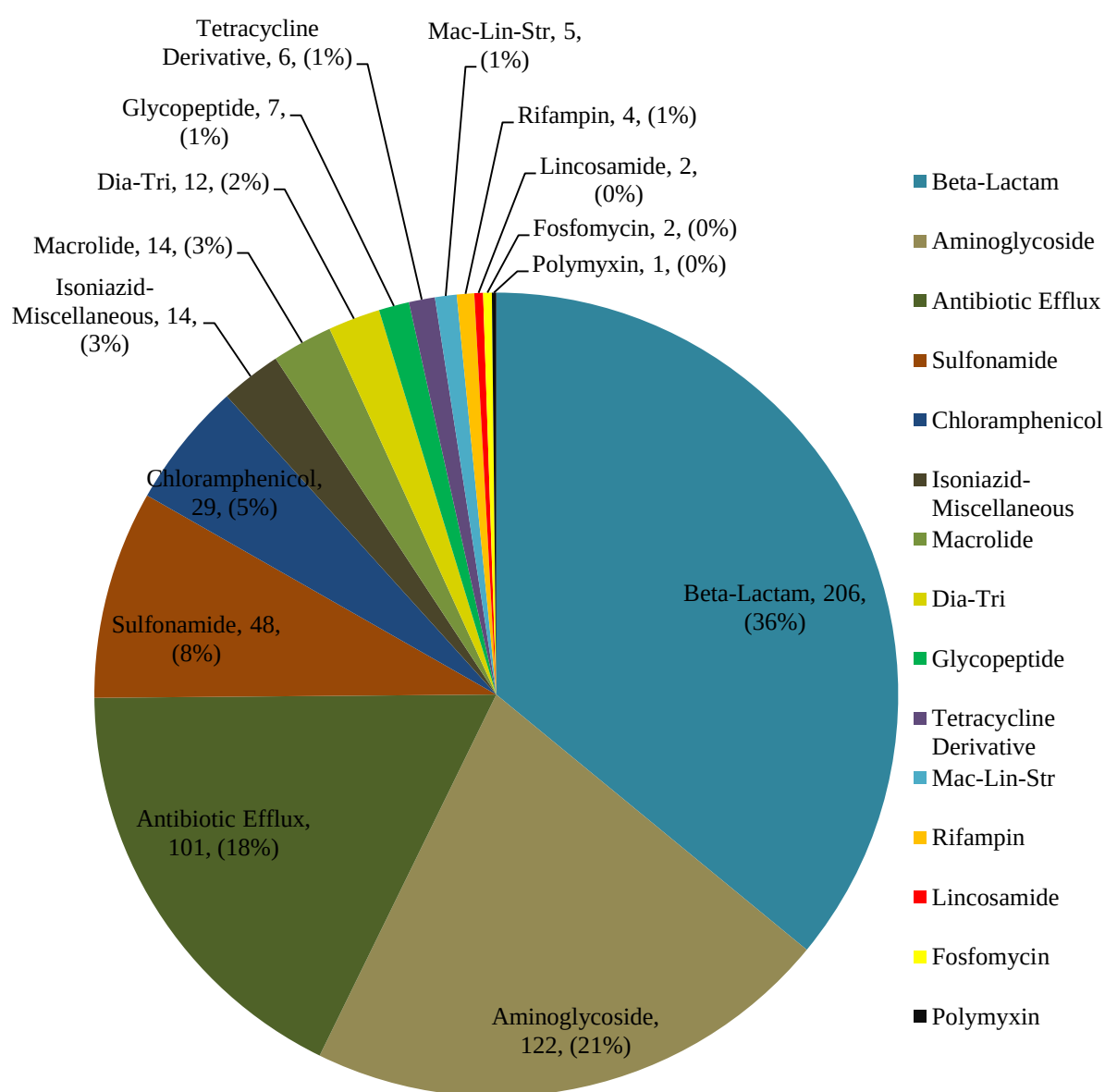


Figure 5.4b. Classes of Antibiotic resistance genes in *Inc F's*

Table 5.4b. IncF's (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
114	451	14	451/114 = (3.95)

*; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.3. Inc HI1 (AR profile)

The Inc HI1 AR profile also shows higher diversity and frequency of the classes of ARGs conferring resistance to Beta-lactams (26%), followed by aminoglycosides (24%), efflux pumps (22%) and sulfonamides (10%) (Figure 5.4c). The corresponding table, Table 5.4c shows the Inc HI1 antibiotic resistance frequency determined from the population size of 15 Inc HI1 plasmids with 93 resistance gene sequences (ARG ORF determined by CARD) and 10 classes of resistance genes found in Inc HI1. This information supplements the Inc HI1 pie chart and is used to calculate the average number of resistance genes per plasmid within Inc HI1 which is determined to be 6.2

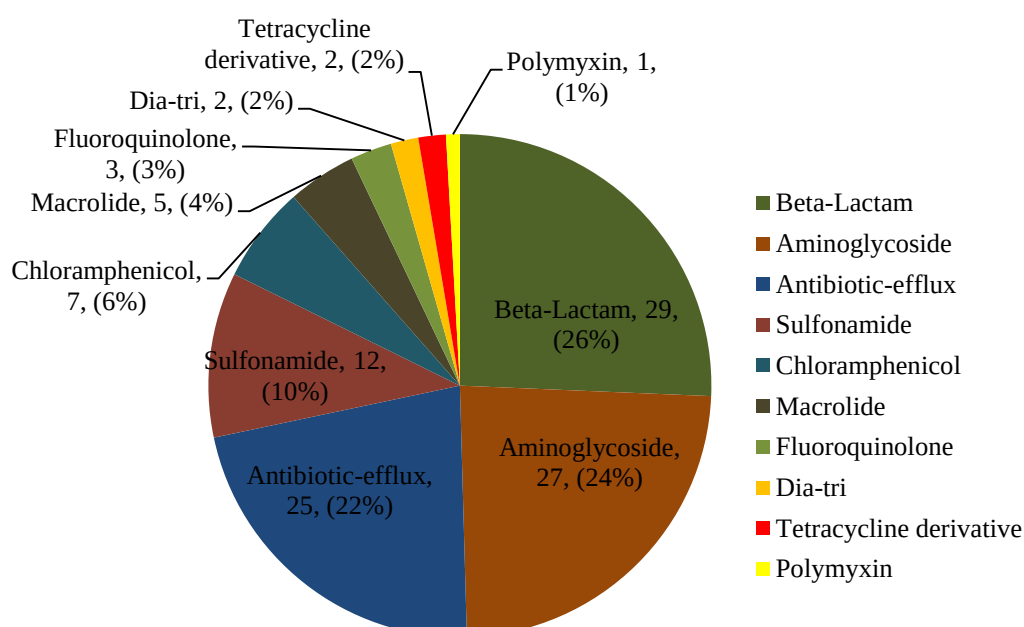


Figure 5.4c. Classes of Antibiotic resistance genes in Inc HI1

Table 5.4c. IncHI1 (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
15	93	10	93/15 = (6.2)

*; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.4. Inc HI2 (AR profile)

Resistance to aminoglycosides (27%) were frequently found to dominate in Inc HI2 (AR profile), others included Beta-lactams (20%), efflux pumps (17%) and sulfonamides (15%) (Figure 5.4d). The corresponding table, Table 5.4d shows the Inc HI2 antibiotic resistance frequency determined from the population size of 10 Inc HI2 plasmids with 102 resistance genes sequences (ARG ORF determined by CARD) and 10 classes of resistance genes. This information supplements the Inc HI2 pie chart and is used to calculate the average number of resistance genes per plasmid within Inc HI2 which is determined to be 10.2

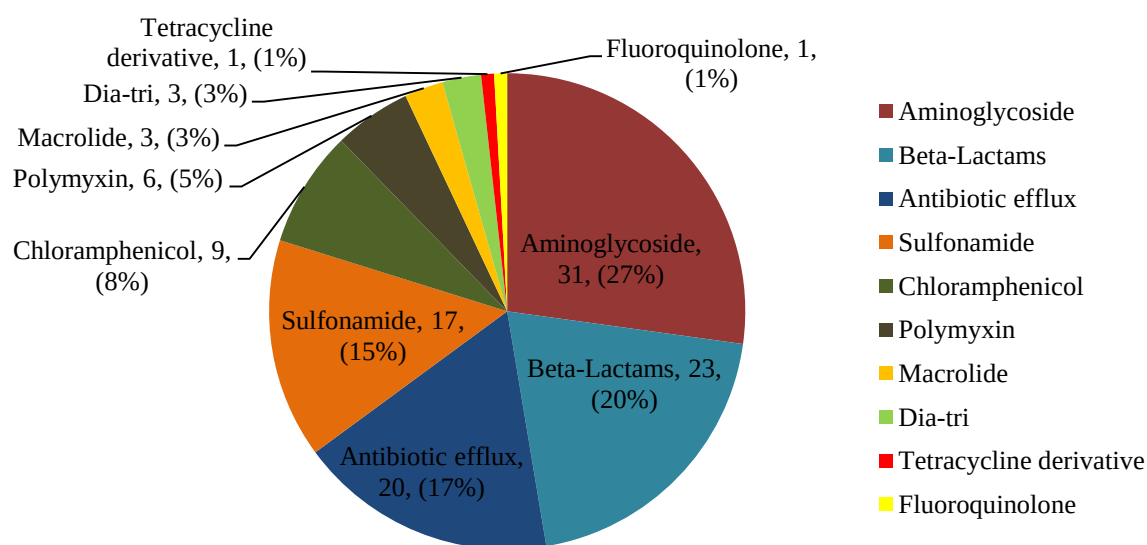


Figure 5.4d. Classes of Antibiotic resistance genes in Inc HI2

Table 5.4d. Inc HI2 (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
10	102	10	102/10 = (10.2)

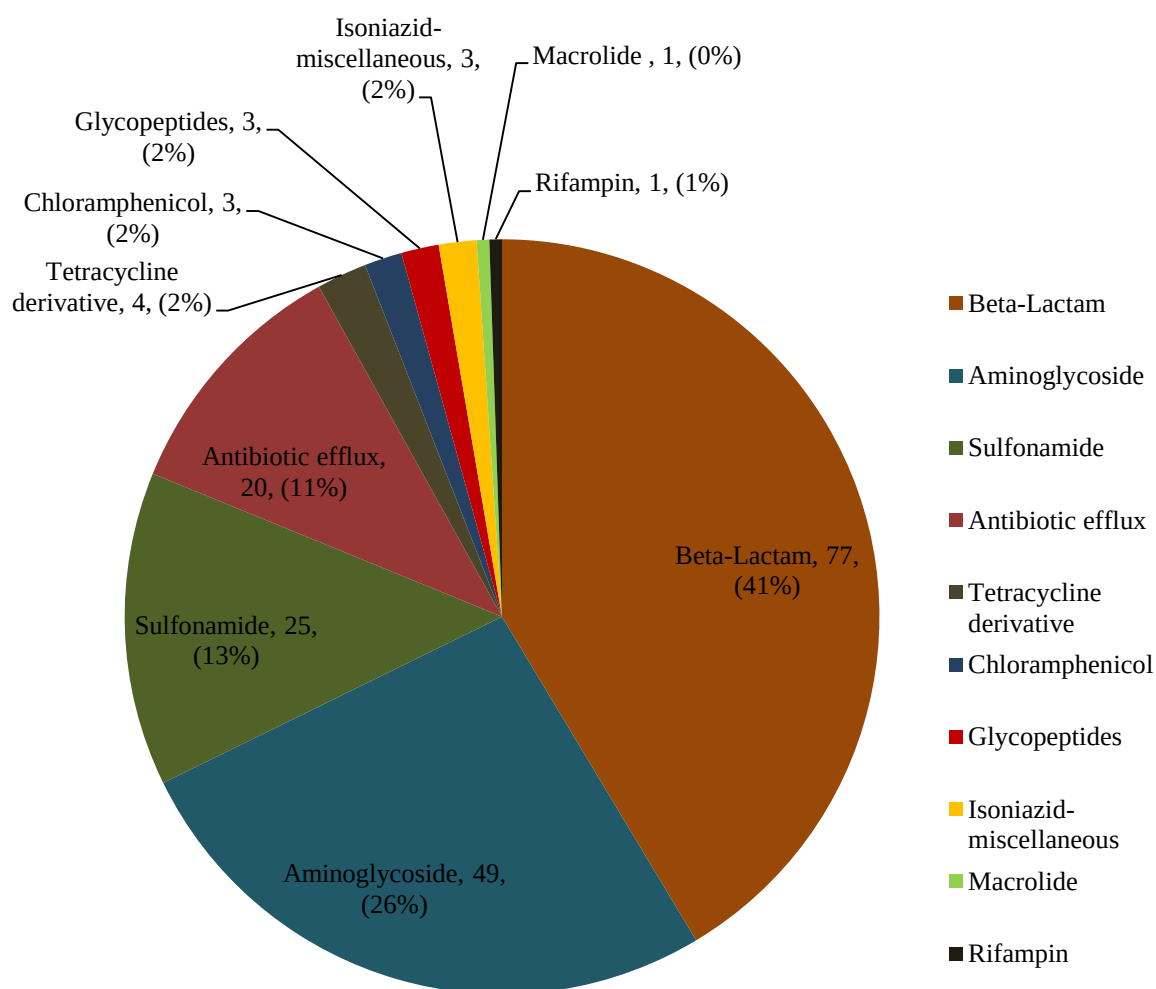
*; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

818

819 5.4.5. *Inc I1* (AR profile)

820

821 The *Inc I1* plasmids were characterized by high numbers of ARGs to beta-lactams (41%),
 822 followed by aminoglycosides (26%), sulfonamides (13%) and efflux pumps (11%) (Figure
 823 5.4e). The corresponding table, Table 5.4e shows the *Inc I1* antibiotic resistance frequency
 824 determined from a population size of 40 *Inc I1* plasmids with 121 resistance gene sequences
 825 (ARG ORF determined by CARD) and 10 classes of resistance genes found in *Inc I1*. This
 826 information supplements the *Inc I1* pie chart and is used to calculate the average number of
 827 resistance genes per plasmid within *Inc I1* which is determined to be 3.025



828

829 Figure 5.4e. Classes of Antibiotic resistance genes in *Inc I1*

830

831

Table 5.4e. IncI1 (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
40	121	10	121/40 = (3.025)

*; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.6. Inc X (AR profile)

The Inc X plasmids were characterized by high numbers of ARGs to beta-lactams (62%), followed by aminoglycosides (25%) and efflux pumps (13%) (Figure 5.4f). The corresponding table, Table 5.4f shows the Inc X antibiotic resistance frequency determined from the population size of 5 Inc X plasmids with 5 resistance genes sequences (ARG ORF determined by CARD) and 3 the classes of resistance genes. This information supplements the Inc X pie chart and is used to calculate the average number of resistance genes per plasmid within Inc X which is determined to be 1.25

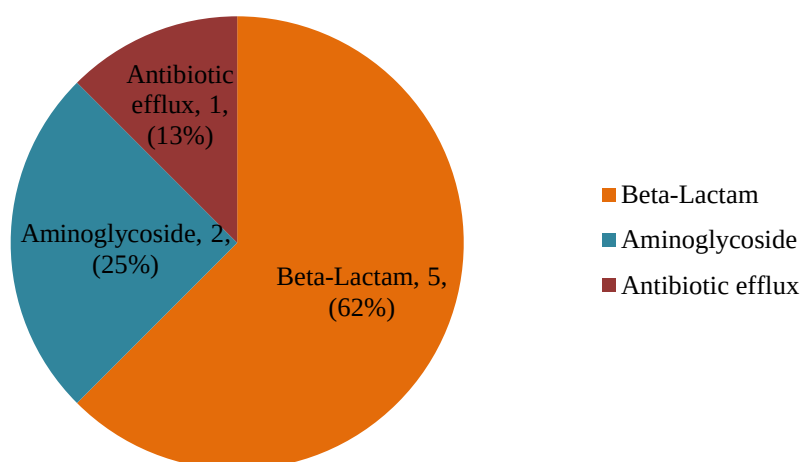


Figure 5.4f. Classes of Antibiotic resistance genes in Inc X

Table 5.4f. Inc X (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
4	5	3	5/4 = 1.25

*; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.7. Inc X1&X2 (AR profile)

The Inc X1&X2 plasmids were characterized by high numbers of ARGs to beta-lactams (67%), followed by aminoglycosides (12%), efflux pumps (6%) and sulfonamides (6%); (Figure 5.4g). The corresponding table, Table 5.4g shows the Inc X1&X2 antibiotic resistance frequency determined from the population size of 19 Inc X1&X2 plasmids; the resistance genes sequences (ARG ORF determined by CARD) is 27 and the classes of resistance genes found in Inc X1&X2 is 7. This information supplements the Inc X1&X2 pie chart and is used to calculate the average number of resistance genes per plasmid within Inc X1&X2 which is determined to be 1.42

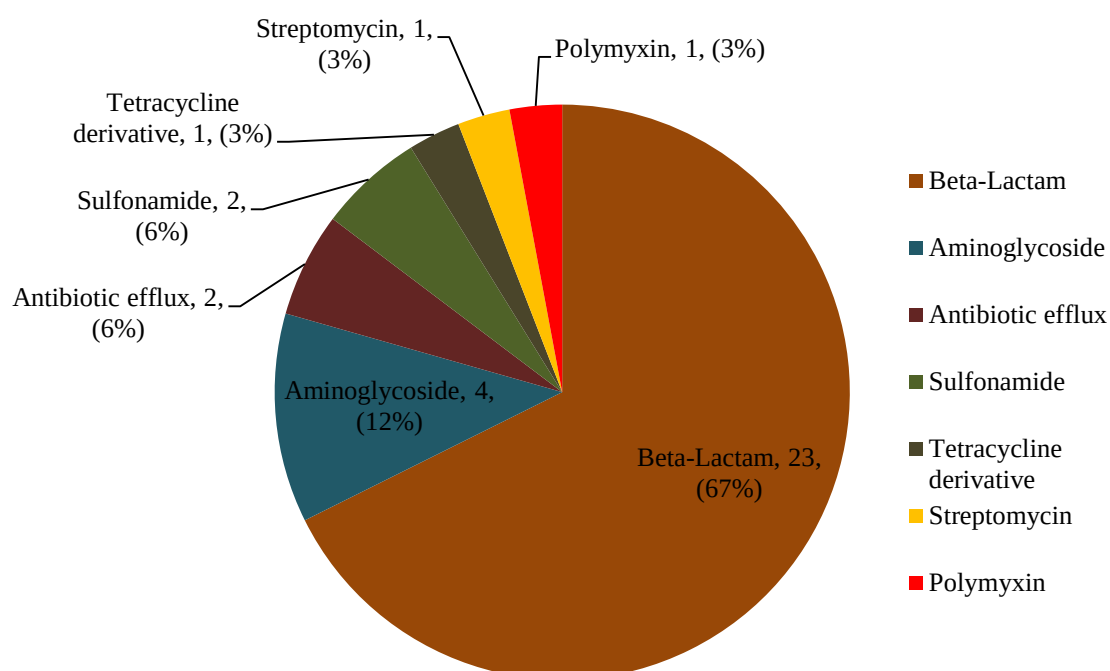


Figure 5.4g. Classes of Antibiotic resistance genes in Inc X1 & X2

Table 5.4g. Inc X1 & X2 (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
19	27	7	27/19 = (1.42)

*; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.8. Inc X3 (AR profile)

The Inc X3 plasmids were characterized by high numbers of ARGs to Beta-lactams (57%), followed by aminoglycosides (15%), efflux pumps (13%) and sulfonamides (7%); (Figure 5.4h). The corresponding table, Table 5.4h shows the Inc X3 antibiotic resistance frequency determined from the population size of 22 Inc X3 plasmids with 37 resistance gene sequences (ARG ORF determined by CARD) and 8 classes of resistance genes. This information supplements the Inc X3 pie chart and is used to calculate the average number of resistance genes per plasmid within Inc HI1 which is determined to be 1.68

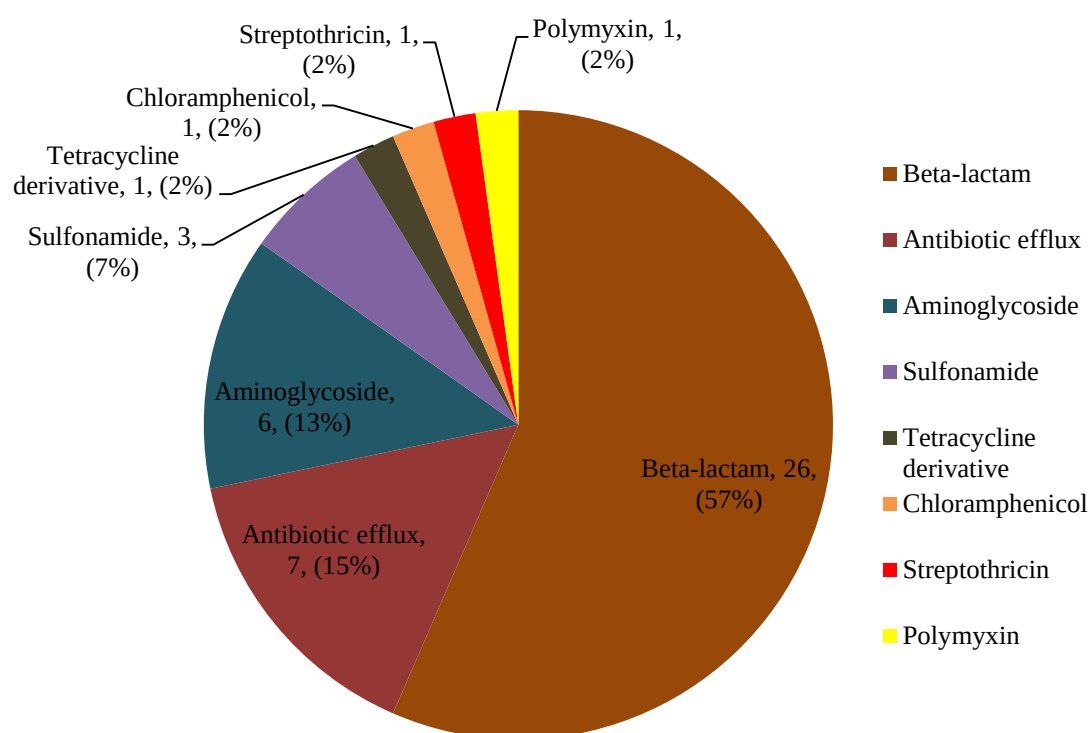


Figure 5.4h. Classes of Antibiotic resistance genes in Inc X3

Table 5.4h. Inc X3 (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
22	37	8	37/22 = (1.68)

*; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.9. Inc X4 (AR profile)

The Inc X4 plasmids were characterized by high numbers of ARGs to Beta-lactams (72%), followed by aminoglycosides (14%) and efflux pumps (14%) (Figure 5.4i). The corresponding table, Table 5.4i shows the Inc X4 antibiotic resistance frequency determined from the population size of 2 Inc X4 plasmids with 5 resistance genes sequences (ARG ORF determined by CARD) and 3 classes of resistance genes. This information supplements the Inc X4 pie chart and is used to calculate the average number of resistance genes per plasmid within Inc X4 which is determined to be 2.5

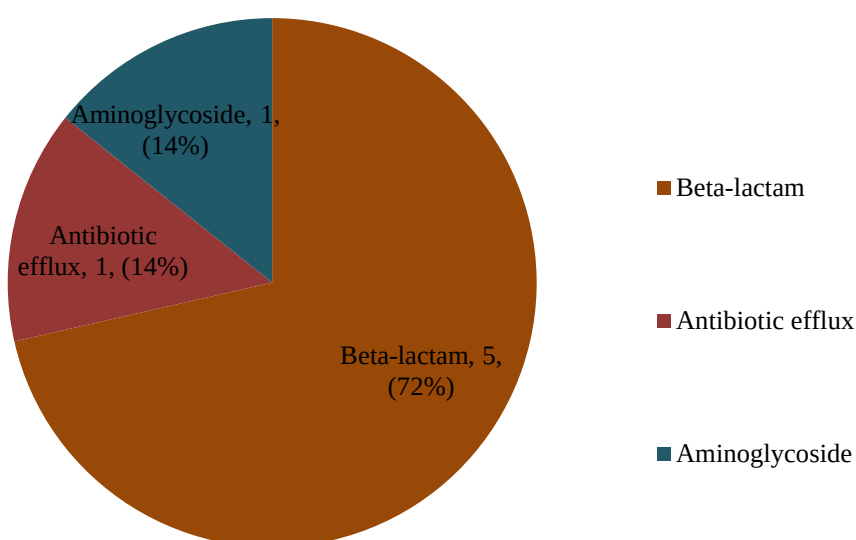


Figure 5.4i. Classes of Antibiotic resistance genes in Inc X4

Table 5.4i. Inc X4 (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
2	5	3	5/2 = (2.5)

*, number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.10. Inc L & M (AR profile)

The Inc L & M plasmids were characterized by high numbers of ARGs to Beta-lactams (45%), followed by aminoglycosides (24%), efflux pumps (11%) and sulfonamides (7%); (Figure 5.4j). The corresponding table, Table 5.4j shows the Inc L & M antibiotic resistance frequency determined from the population size of 21 Inc L & M plasmids; the resistance genes sequences (ARG ORF determined by CARD) is 79 and the classes of resistance genes found in Inc L & M is 11. This information supplements the Inc L & M pie chart and is used to calculate the average number of resistance genes per plasmid within Inc L & M which is determined to be 3.76

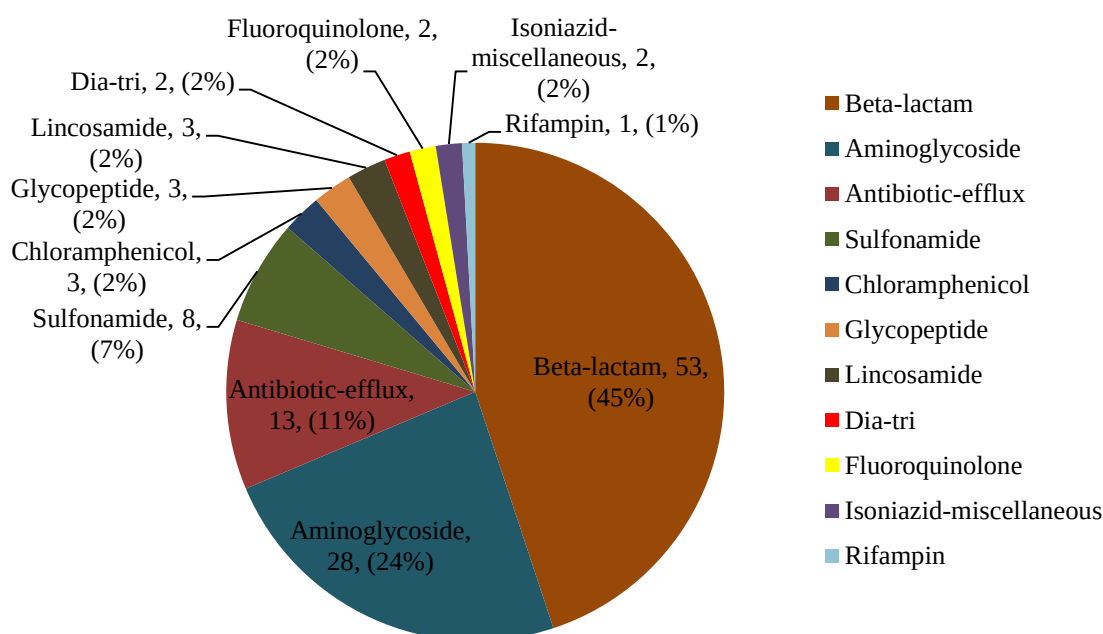


Figure 5.4j. Classes of Antibiotic resistance genes in Inc L & M

Table 5.4j. Inc L & M (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
21	79	11	79/21 = (3.76)

*; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.11. Inc N (AR profile)

The Inc N plasmids were characterized by high numbers of ARGs to Beta-lactams (42%), followed by aminoglycosides (20%), sulfonamides (10%) and efflux pumps (10%); (Figure 5.4k). The corresponding table, Table 5.4k shows the Inc N antibiotic resistance frequency determined from the population size of 44 Inc N plasmids. The resistance genes sequences (ARG ORF determined by CARD) is 189 and the classes of resistance genes found in Inc N is 14. This information supplements the Inc N pie chart and is used to calculate the average number of resistance genes per plasmid within Inc N which is determined to be 4.3

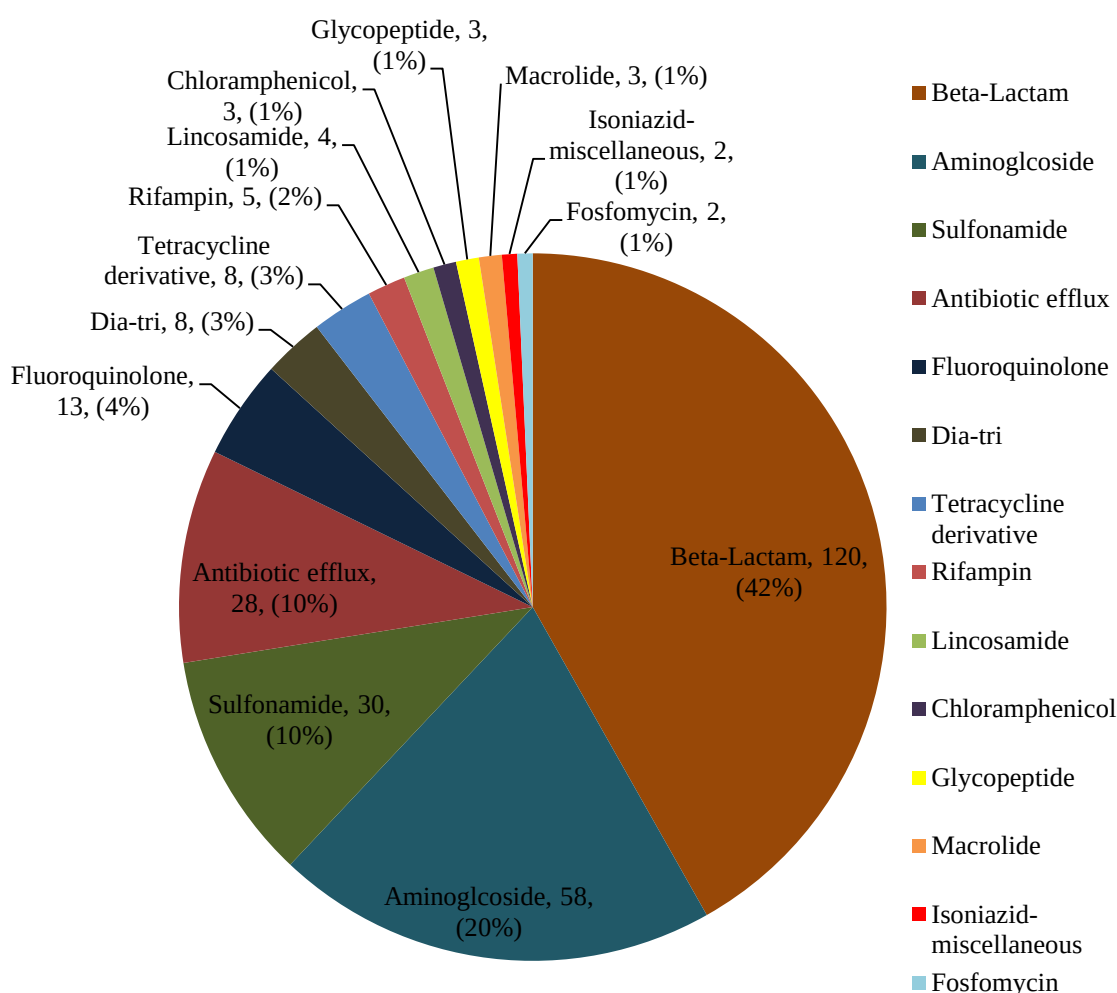


Figure 5.4k. Classes of Antibiotic resistance genes in Inc N

929 Table 5.4k. Inc N (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
44	189	14	189/44 = (4.3)

930 *; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid
931 population

932

933 *5.4.12. Inc Y (AR profile)*

934

935 There was only 1 Inc Y plasmid characterized by only 1 ARG which is an efflux pump. The
936 corresponding table, Table 5.4l shows the Inc Y antibiotic resistance frequency which
937 consists of 1 plasmid typed for Inc Y having 1 resistance gene sequence (ARG ORF
938 determined by CARD) and only 1 class of resistance genes found in Inc Y.

939

940

941 Table 5.4l. Inc Y (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
1	1	1	1/1 = (1)

942 *; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid
943 population

944

945 *5.4.13. Inc P (AR profile)*

946

947 The Inc P plasmids were characterized by high numbers of ARGs to Beta-lactams (35%),
948 followed by aminoglycosides (32%), sulfonamides (18%) and efflux pumps (3%); (Figure
949 5.4m). The corresponding table, Table 5.4m shows the Inc P antibiotic resistance frequency
950 determined from the population size of 18 Inc P plasmids; the resistance genes sequences
951 (ARG ORF determined by CARD) is 56 and the classes of resistance genes found in Inc P is
952 10. This information supplements the Inc P pie chart and is used to calculate the average
953 number of resistance genes per plasmid within Inc P which is determined to be 3.1.

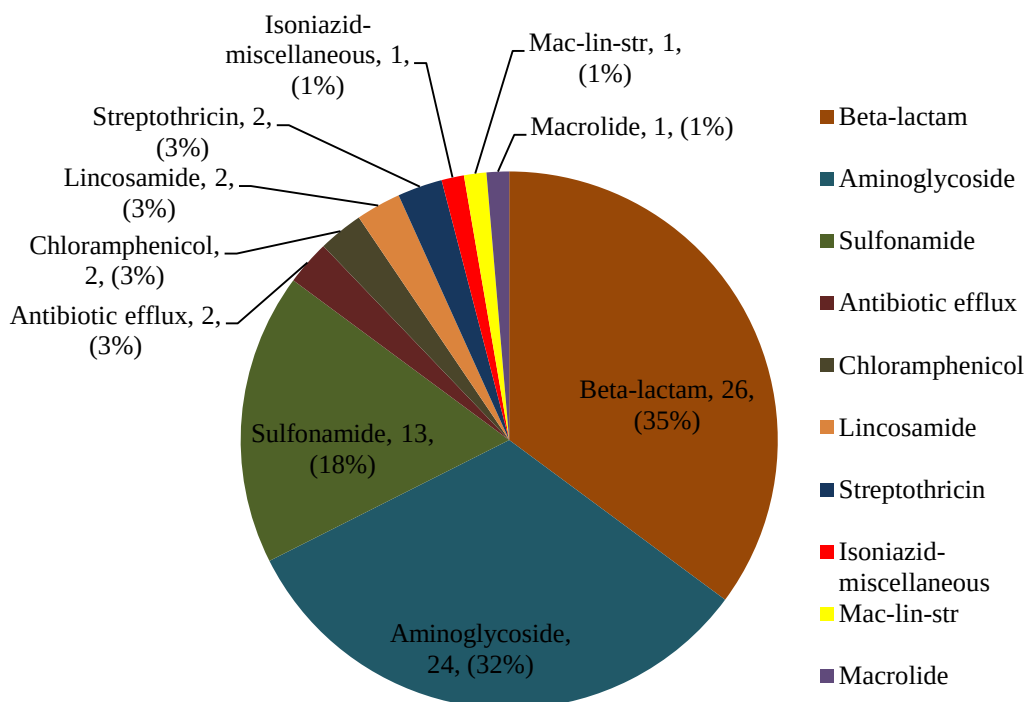


Figure 5.4m. Classes of Antibiotic resistance genes in Inc P

Table 5.4m. Inc P (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
18	56	10	56/18 = (3.1)

*; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.14. Inc A & C (AR profile)

The Inc A & C plasmids were characterized by high numbers of ARGs to Beta-lactams (33%), followed by aminoglycosides (28%), sulfonamides (15%) and efflux pumps (12%); (Figure 5.4n). The corresponding table, Table 5.4n shows the Inc A & C antibiotic resistance frequency determined from the population size of 58 Inc A & C plasmids; the resistance genes sequences (ARG ORF determined by CARD) is 405 and the classes of resistance genes found in Inc A & C is 14. This information supplements the Inc A & C pie chart and is used to calculate the average number of resistance genes per plasmid within Inc A & C which is determined to be 6.98

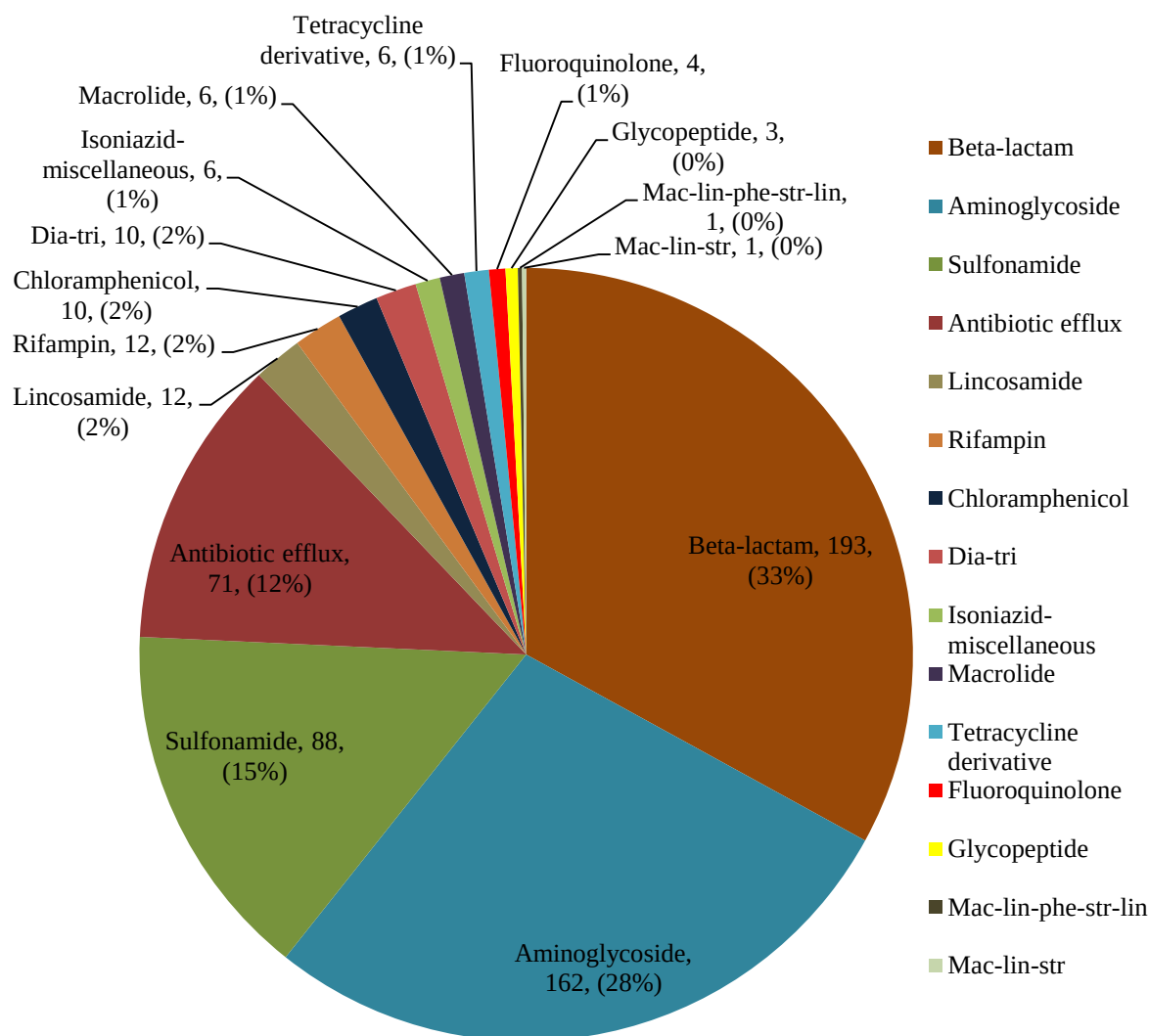


Figure 5.4n. Classes of Antibiotic resistance genes in Inc A & C

Table 5.4n. Inc A & C (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
58	405	14	405/58 = (6.98)

*, number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.15. Inc T (AR profile)

The Inc T plasmids were characterized by high numbers of ARGs to Beta-lactams (40%), followed by aminoglycosides (20%), efflux pumps (20%) and sulfonamides (20%); (Figure 5.4o). The corresponding table, Table 5.4o shows the Inc T antibiotic resistance frequency

determined from the population size of 2 Inc T plasmids; the resistance genes sequences (ARG ORF determined by CARD) is 3 and the classes of resistance genes found in Inc T is 4. This information supplements the Inc T pie chart and is used to calculate the average number of resistance genes per plasmid within Inc T which is determined to be 0.75

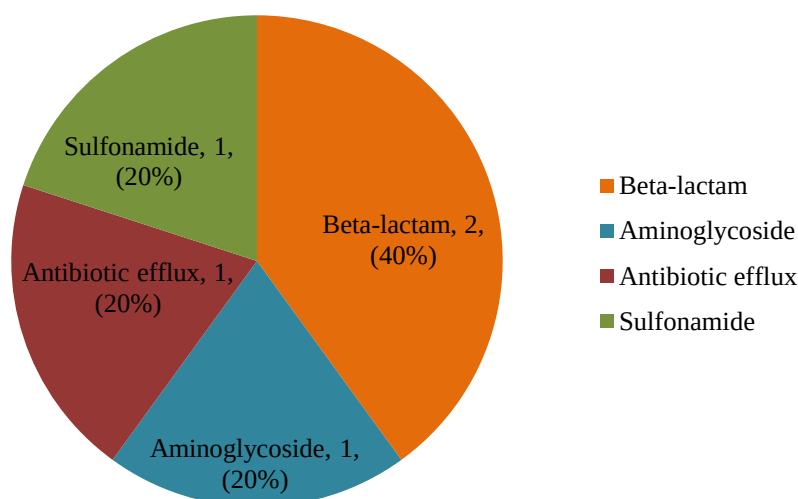


Figure 5.4o. Classes of Antibiotic resistance genes in Inc T

Table 5.4o. Inc T (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
2	3	4	3/4 = (0.75)

*; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.16. Inc K & B (AR profile)

The Inc K & B plasmids were characterized by high numbers of ARGs to Beta-lactams (34%), followed by aminoglycosides (27%), sulfonamides (15%) and efflux pumps (6%); (Figure 5.4p). The corresponding table, Table 5.4p shows the Inc K & B antibiotic resistance frequency determined from the population size of 12 Inc K & B plasmids; the resistance genes sequences (ARG ORF determined by CARD) is 39 and the classes of resistance genes found in Inc K & B is 9. This information supplements the Inc K & B pie chart and is used to calculate the average number of resistance genes per plasmid within Inc K & B which is determined to be 3.25

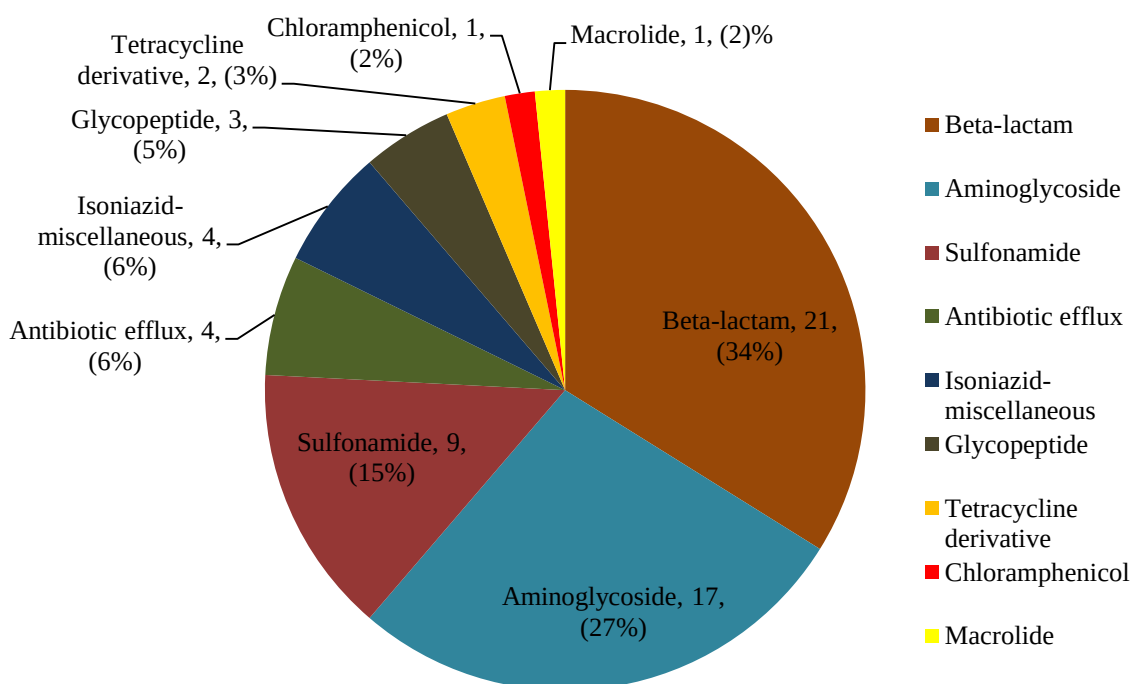


Figure 5.4p. Classes of Antibiotic resistance genes in Inc K & B

Table 5.4p. Inc K & B (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
12	39	9	39/12 = (3.25)

*, number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.17. Inc B & O (AR profile)

The Inc B & O plasmids were characterized by high numbers of ARGs to Beta-lactams and aminoglycosides both at (33%), followed by sulfonamides (12%) and efflux pumps (7%); (Figure 5.4q). The corresponding table, Table 5.4q shows the Inc B & O antibiotic resistance frequency determined from the population size of 22 Inc B & O plasmids; the resistance genes sequences (ARG ORF determined by CARD) is 74 and the classes of resistance genes found in Inc B & O is 10. This information supplements the Inc B & O pie chart and is used to calculate the average number of resistance genes per plasmid within Inc B & O which is determined to be 3.36

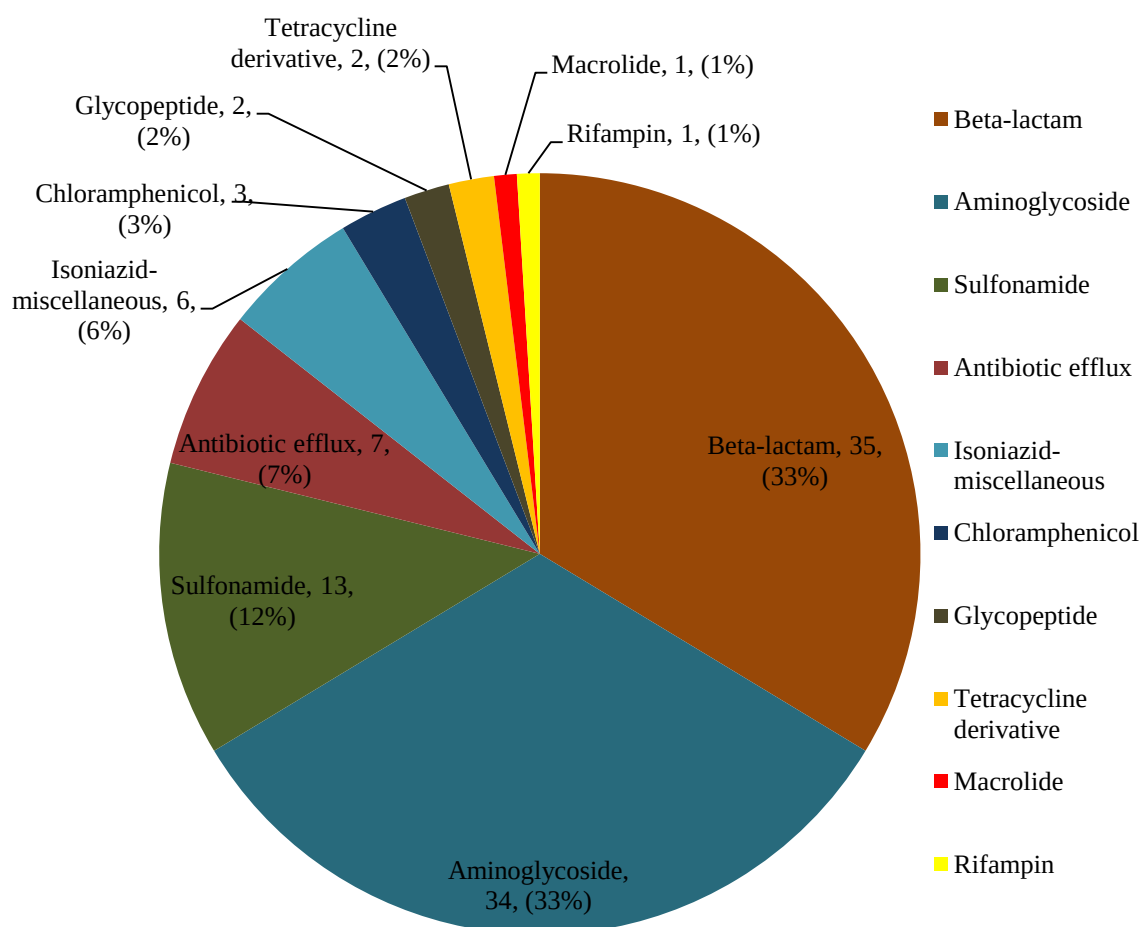


Figure 5.4q. Classes of Antibiotic resistance genes in Inc B & O

Table 5.4q. Inc B & O (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
22	74	10	74/22 = (3.36)

*, number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.18. Inc R (AR profile)

The Inc R plasmids were characterized by high numbers of ARGs to Beta-lactams (43%), followed by aminoglycosides (21%), efflux pumps (14%) and sulfonamides (11%); (Figure 5.4r). The corresponding table, Table 5.4r shows the Inc R antibiotic resistance frequency determined from the population size of 23 Inc R plasmids; the resistance genes sequences

(ARG ORF determined by CARD) is 159 and the classes of resistance genes found in Inc R is 11. This information supplements the Inc R pie chart and is used to calculate the average number of resistance genes per plasmid within Inc R which is determined to be 6.91

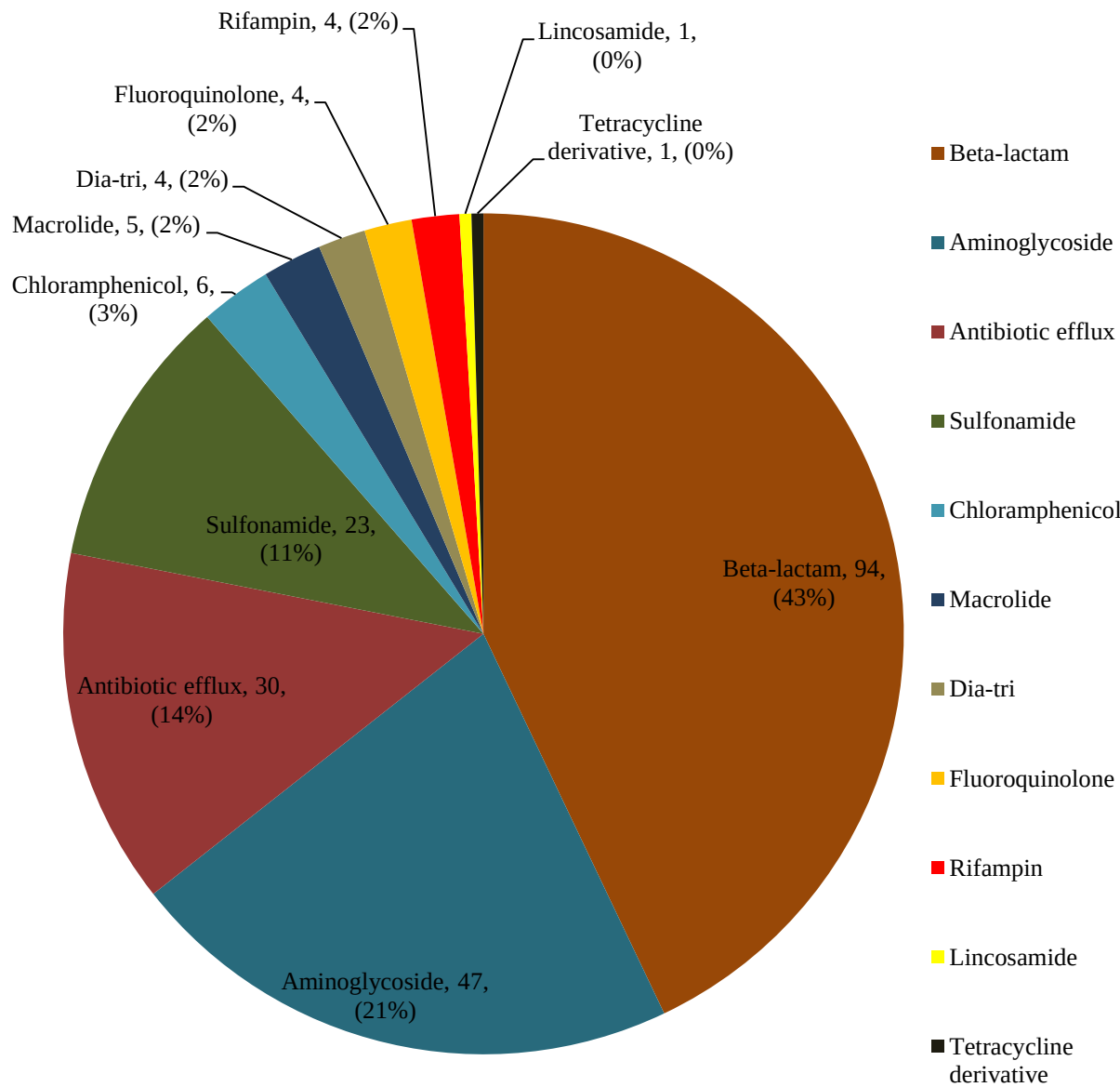


Figure 5.4r. Classes of Antibiotic resistance genes in Inc R

Table 5.4r. Inc R (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
23	159	11	159/23 = (6.91)

*; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.19. Inc U (AR profile)

The Inc U plasmids were characterized by high numbers of ARGs to Beta-lactams (30%), followed by aminoglycosides (20%), sulfonamides (14%) and efflux pumps (12%); (Figure 5.4s). The corresponding table, Table 5.4s shows the Inc U antibiotic resistance frequency determined from the population size of 11 Inc U plasmids; the resistance genes sequences (ARG ORF determined by CARD) is 48 and the classes of resistance genes found in Inc U is 14. This information supplements the Inc U pie chart and is used to calculate the average number of resistance genes per plasmid within Inc U which is determined to be 4.36

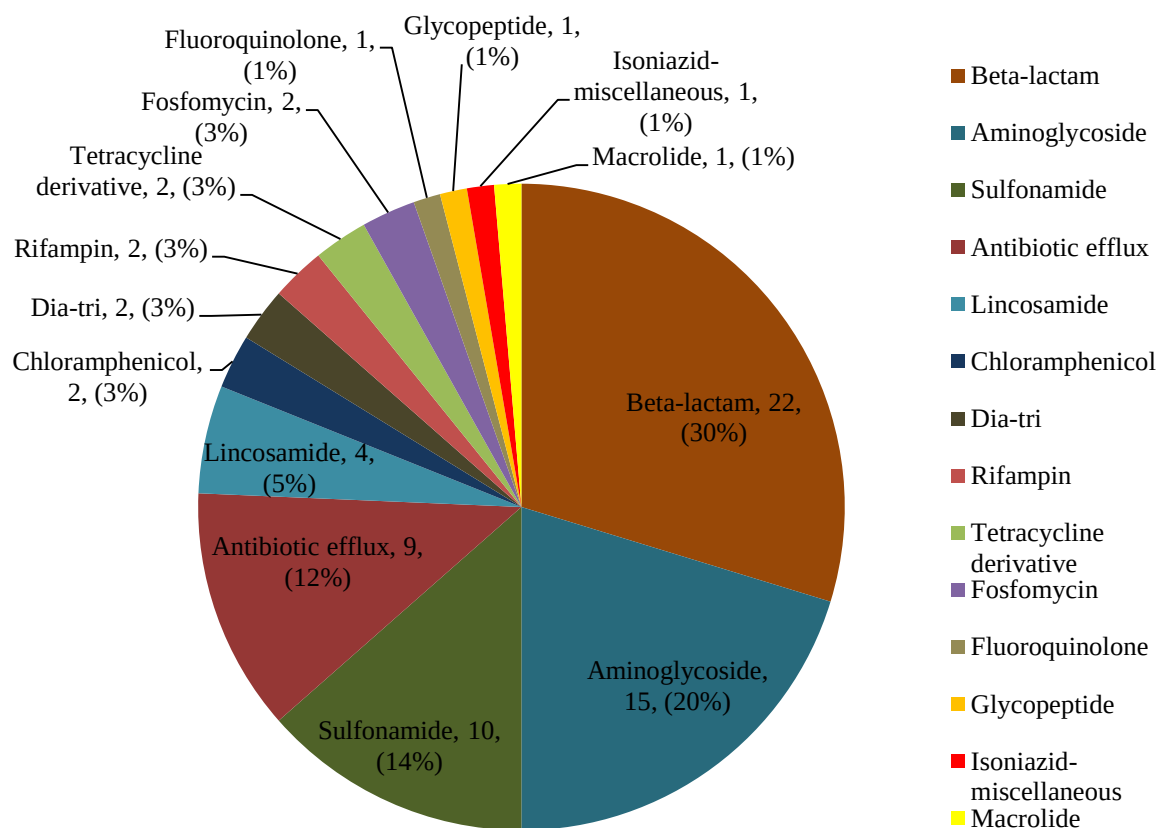


Figure 5.4s. Classes of Antibiotic resistance genes in Inc U

1056 Table 5.4s. Inc U (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
11	48	14	48/11 = (4.36)

1057 *; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid
1058 population

1059

1060 *5.4.20. ColE1 (AR profile)*

1061

1062 The ColE1 plasmids were characterized by high numbers of ARGs to Beta-lactams (37%),
1063 followed by aminoglycosides (25%), efflux pumps (10%) and sulfonamides (8%); (Figure
1064 5.4t). The corresponding table, Table 5.4t shows the ColE1 antibiotic resistance frequency
1065 determined from the population size of 83 ColE1 plasmids; the resistance genes sequences
1066 (ARG ORF determined by CARD) is 150 and the classes of resistance genes found in ColE1
1067 is 12. This information supplements the ColE1 pie chart and is used to calculate the average
1068 number of resistance genes per plasmid within ColE1 which is determined to be 1.81

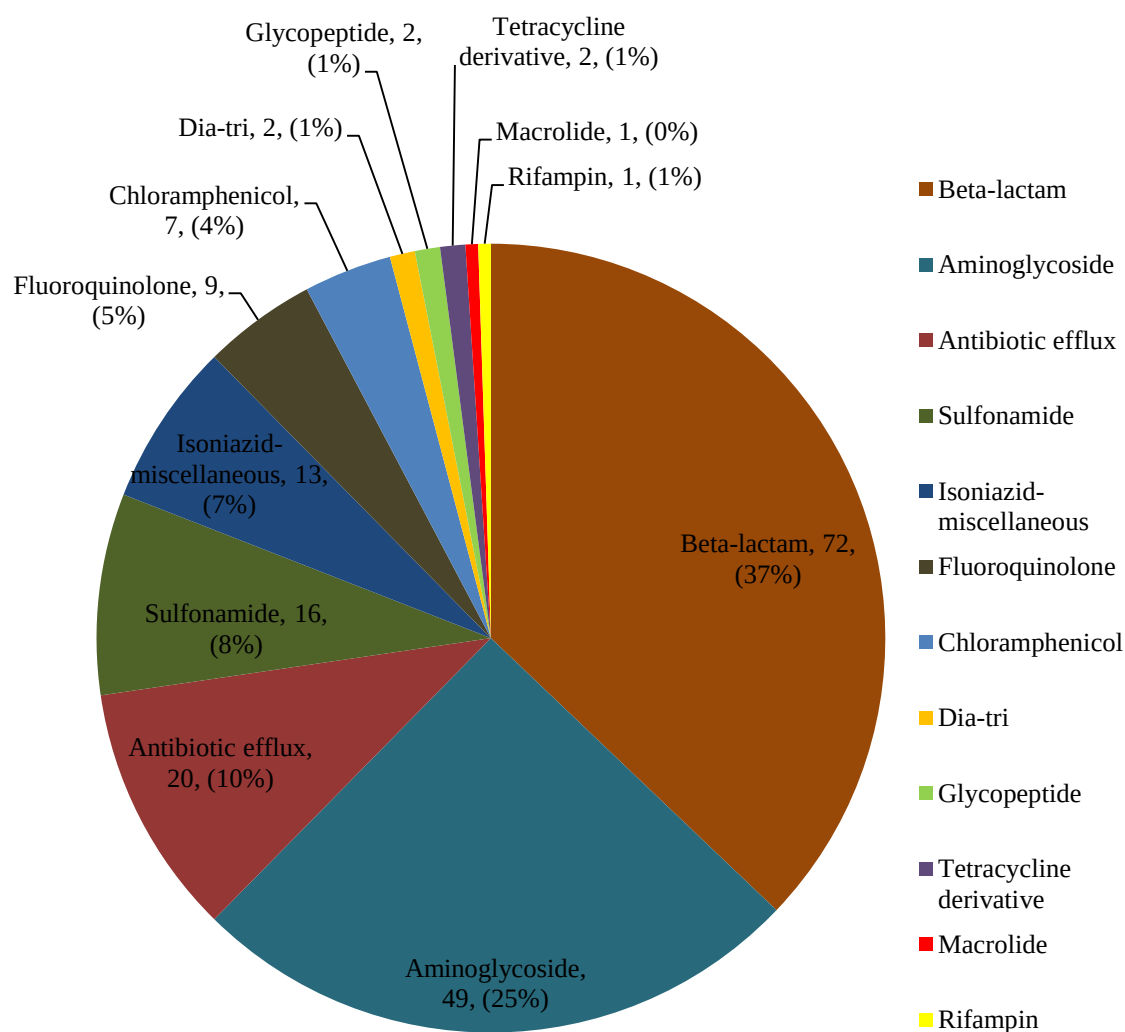


Figure 5.4t. Classes of Antibiotic resistance genes in ColE1

Table 5.4t. ColE1 (AR frequency)

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
83	150	12	150/83 = (1.81)

*, number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

5.4.21. Comprehensive AR profile of all typeable plasmids.

All typeable plasmids were characterized by high numbers of ARGs to Beta-lactams (37%), followed by aminoglycosides (24%), efflux pumps (13%) and sulfonamides (11%); (Figure 5.4u). The corresponding table, Table 5.4u shows the antibiotic resistance profile of all

typeable plasmids. This consists of total number of all typeable plasmids determined in each Inc/rep group which is 521, 2053 resistance genes sequences (ARG ORF determined by CARD) and 18 classes of resistance genes found in all typeable plasmids. This information supplements the pie chart of all typeable plasmids and is used to calculate the average number of resistance genes per plasmid within all typeable plasmids which is determined to be 3.94

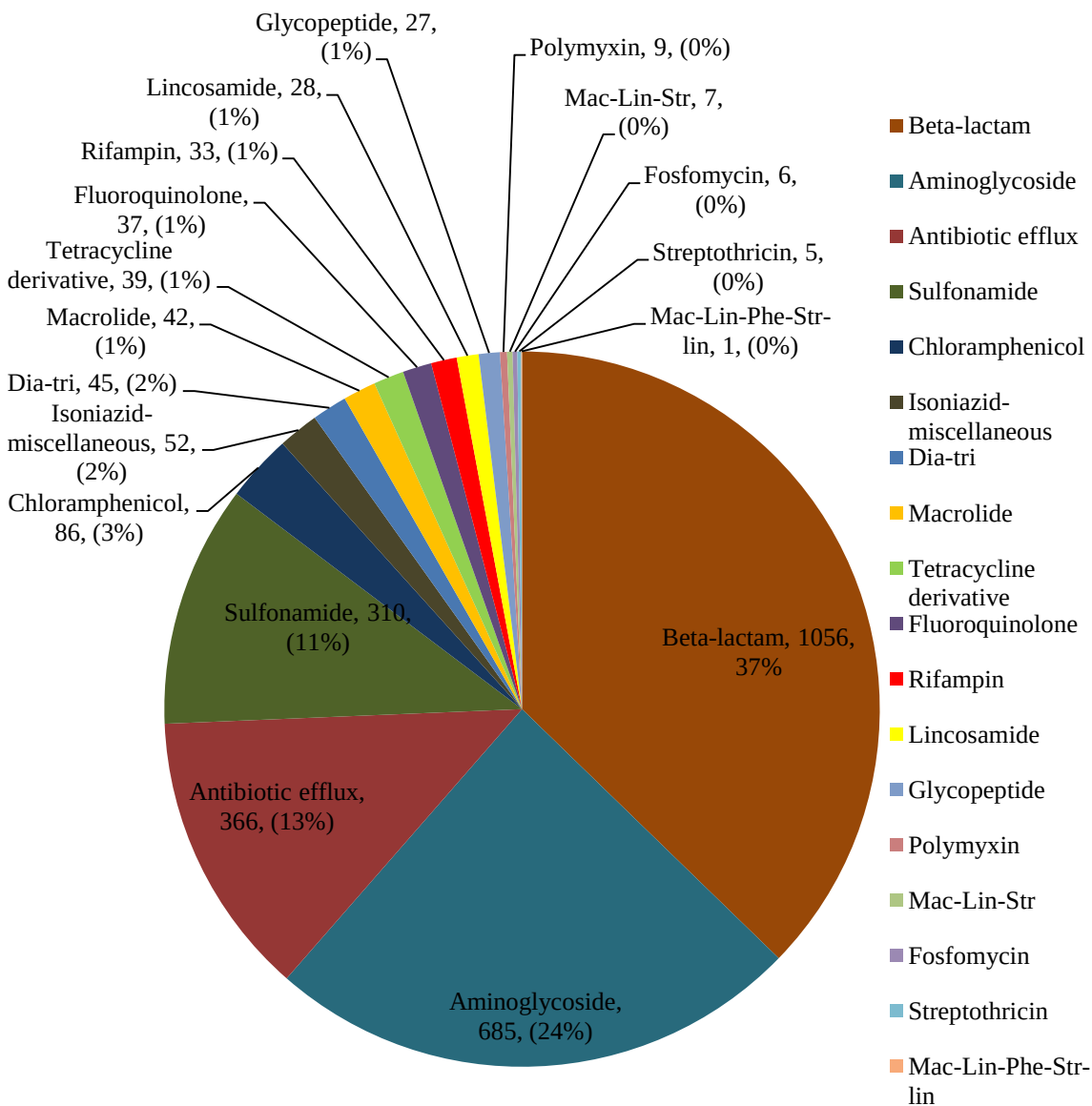


Figure 5.4u. Classes of Antibiotic resistance genes in all typeable plasmids

Table 5.4u. Comprehensive AR frequency of all typeable plasmids

Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
521	2053	18	2053/521 = (3.94)

*; number of ARG ORF determined by CARD., #; Average number of ARG ORF/plasmid population

In summary the association of how different classes of resistance genes are acquired in different Inc/rep groups was made to determine if there is a possible preferential selection of plasmids of a given Inc/rep group to acquire genes of certain resistance gene classes more than others. The exercise shows that β -lactams, aminoglycosides, antibiotic efflux and sulfonamides were the predominating resistant gene classes in more or less equal proportions across different Inc/rep groups. In all instances β -lactams were always more than other gene classes. In total these three resistant gene classes occupied between 75% and 95% of the total resistant genes in respective Inc/rep groups. The AR gene frequency per incompatibility group ranged between 0.75 (lower limit) to 10.2 (upper limit) and the AR gene frequency average of all the Inc groups was 3.94. The Inc/rep groups with the lowest AR gene frequency are Inc T, all Inc X subtypes, Inc Y and ColE1 whilst Inc HI1 and Inc HI2 Inc/rep groups have greater than average gene frequencies. The occurrence of higher gene frequencies is related to plasmid length because ColE1 plasmids are known to be short and they carry fewer genes and Inc HI1 and Inc HI2 are long and as such they have the capacity to carry many genes (Shintani *et al.*, 2015).

As already shown, the number of resistance plasmids that are typeable into different Inc/rep groups using the PBRT technique in relation to the total number of resistance plasmids was relatively low (35%). Due to this, the quantifiable resistance genes determined by CARD are not represented in their entirety in the population of plasmids that could be allocated into respective Inc/rep groups. Despite the fact, these are the shortcomings of the method but with those that are typeable it was determined that the three prominent resistance gene classes still predominated in typeable plasmids. β -lactams were the most abundant in typeable plasmids at 37%, followed by aminoglycosides at 24% and lastly efflux pumps had reduced to 13%. This implies that the majority of the efflux pump class of resistant genes are not found in plasmids of *Enterobacteriaceae* (G-negative bacteria) that are typeable using the PBRT technique. In

contrast, the most abundant antibiotic resistance gene classes in *Enterobacteriaceae* are the Extended spectrum β -lactamases (ESBLs) determined by CARD in their entirety as β -lactams, these findings are also in agreement with Carattoli (2009).

5.5. Plasmid analysis by reference environment

Plasmids of known Inc/rep region were allocated to reference environment or isolation sources as shown in Table 5.5. Approximately 82% plasmids were found associated with a host, 4.3 % of the plasmids were associated with managed environments, 1.9 % plasmids were associated with natural environments, and lastly 11.8% plasmids have unknown environments

Table 5.5. Allocation of Plasmids by incompatibility against their isolation source (reference environment)

Inc/rep group	Natural environment	Host associated environment	Managed environment	No documented isolation source
Inc W	0	0	0	3
Inc F's	0	102	2	4
Inc HI1	0	12	2	1
Inc HI2	0	8	1	1
Inc I1	0	37	0	3
Inc X	0	4	0	0
Inc X1&X2	0	16	0	3
Inc X3	0	19	0	3
Inc X4	0	2	0	0
Inc L&M	0	18	2	1
Inc N	0	31	9	3
Inc Y	0	0	1	0
Inc P	6	8	2	2
Inc A&C	1	41	3	11
Inc T	0	2	0	0
Inc K&B	0	10	0	2
Inc B&O	0	16	0	6
Inc R	0	19	0	4
Inc U	0	7	3	1
ColE1	3	69	0	11
Total (%)	10(1.9%)	418(82 %)	22(43%)	60(11.8%)

Despite the efficiency of the method used in determining the plasmids and classifying them based on the reference environment, the results tend to be biased towards the host associated environment linked to clinical isolates. This is true because antibiotic resistance has previously been associated with clinical settings, only recently research has shifted focus to non-clinical environments.

5.6. Distribution of Plasmid Incompatibility Groups in different environments

The main aim of the study was to determine how the incompatibility of antibiotic resistance plasmids would be found to correlate with how they distribute in different environments. The Chord association layout would visually represent this correlation efficiently. Each chord diagram described below comprises of a group of four Inc groups.

The chord diagram in, Figure 5.6a. comprises of Inc W, F, HI1 and HI2; the results reveal the Inc W group is not associated with any environment (no documented isolation source), this is probably due to the fact that there are only 3 plasmids in the group which is by comparison a few plasmids. Inc F plasmids were found only in host associated and managed environments. There were 100 Inc F plasmids in host associated compared to 2 in managed environments while there were 12 Inc HI1 plasmids in host associated and only 2 in managed environment and 8 Inc HI2 were found in host associated compared to 1 in managed environments.

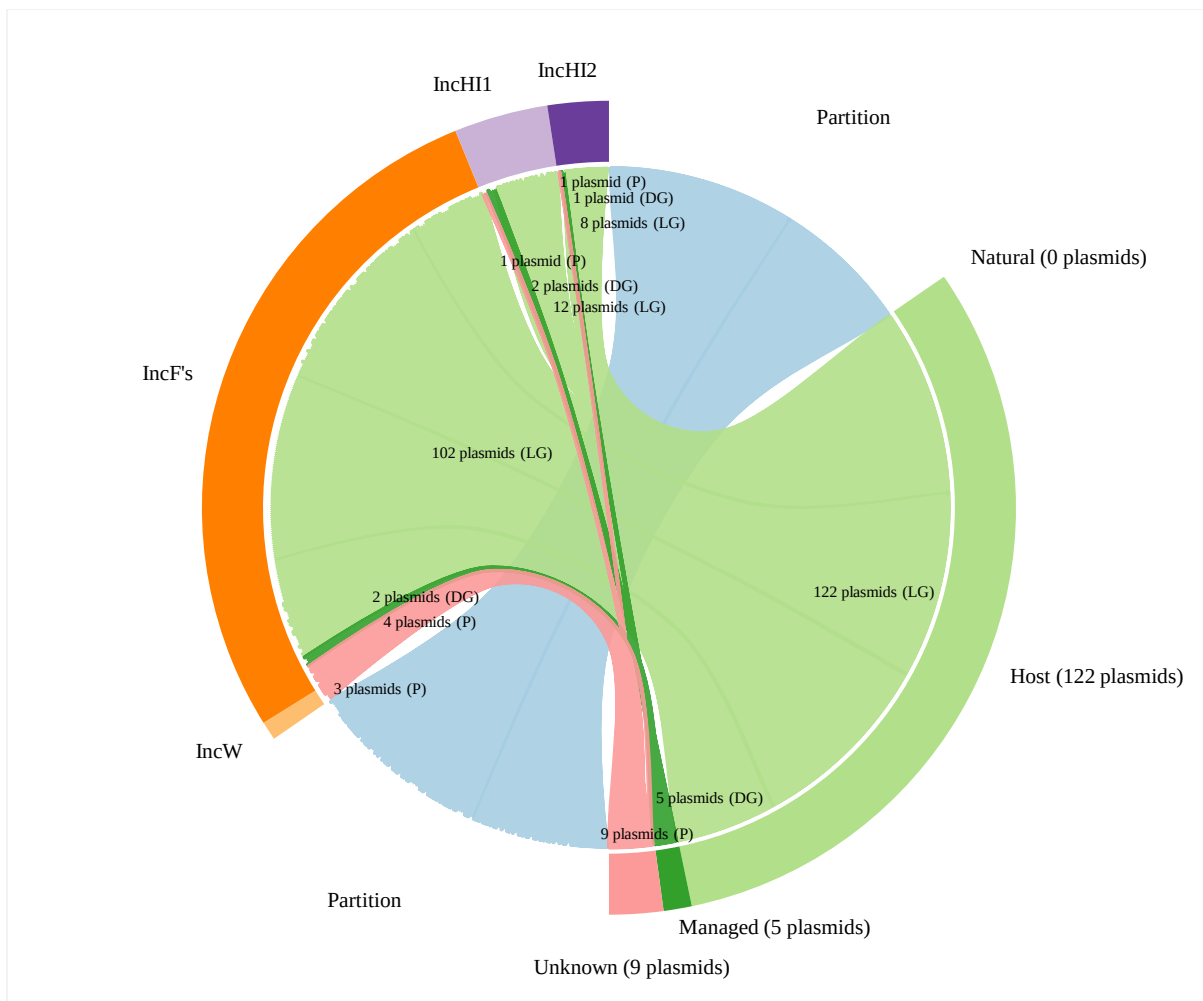


Figure 5.6a. Chord Diagram of Inc/rep groups against environment categories (Inc W, Inc F's, Inc HI1 & Inc HI2). The environment categories are segregated into the right portion of the diagrams, denoted as; Natural (dark blue-DB), Host (light green-LG), Managed (Dark Green-DG), Unknown (Pink-P).

Chord Diagram of Figure 5.6b. which encompasses Inc I1, Inc X, Inc X1&X2 and Inc X3. (NB: Analysis of replicon typing markers. The typing of Inc X1 and Inc X2 plasmids yielded an identical set of plasmids which were grouped as Inc X1&X2). All Inc/rep groups had plasmids associated exclusively with host associated environments; Inc I1 (37 plasmids), Inc X (4 plasmids), Inc X1&X2 (16 plasmids) and Inc X3 (19 plasmids).

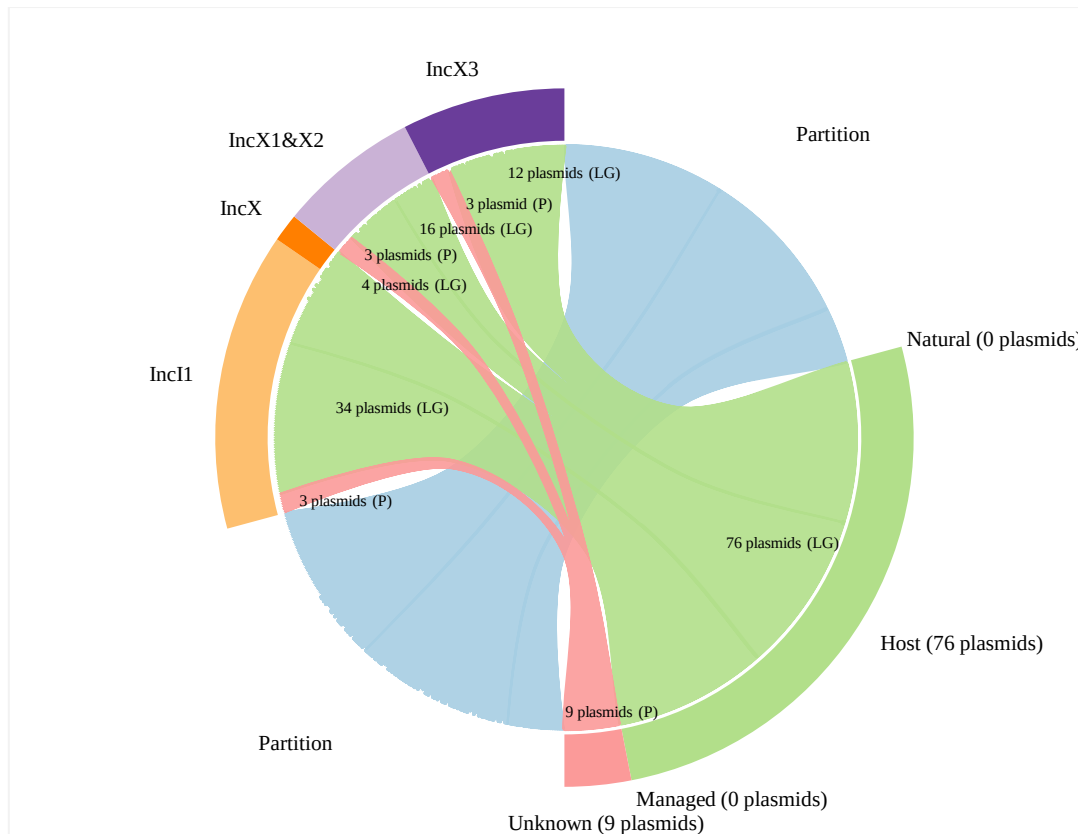


Figure 5.6b. Chord Diagram of Inc/rep groups against environment categories (Inc I1, IncX, IncX1&X2 and IncX3). The environment categories are segregated into the right portion of the diagrams, denoted as; Natural (dark blue-DB), Host (light green-LG), Managed (Dark Green-DG), Unknown (Pink-P).

In Chord Diagram of Figure 5.6c. Inc X4 plasmids (2 plasmids) were exclusively from host associated environments, similar to other Inc X subgroups. Inc L&M plasmids could be found only in host associated and managed environments with the majority being host associated; 18 plasmids to 2 plasmids of managed environments (18:2). Inc N plasmids could also be found only in host associated and managed environments in similar proportions, being 31 host associated plasmids to 9 plasmids of managed environments (31:9). In Inc Y only one plasmid identified with the replicon was associated with managed environments.

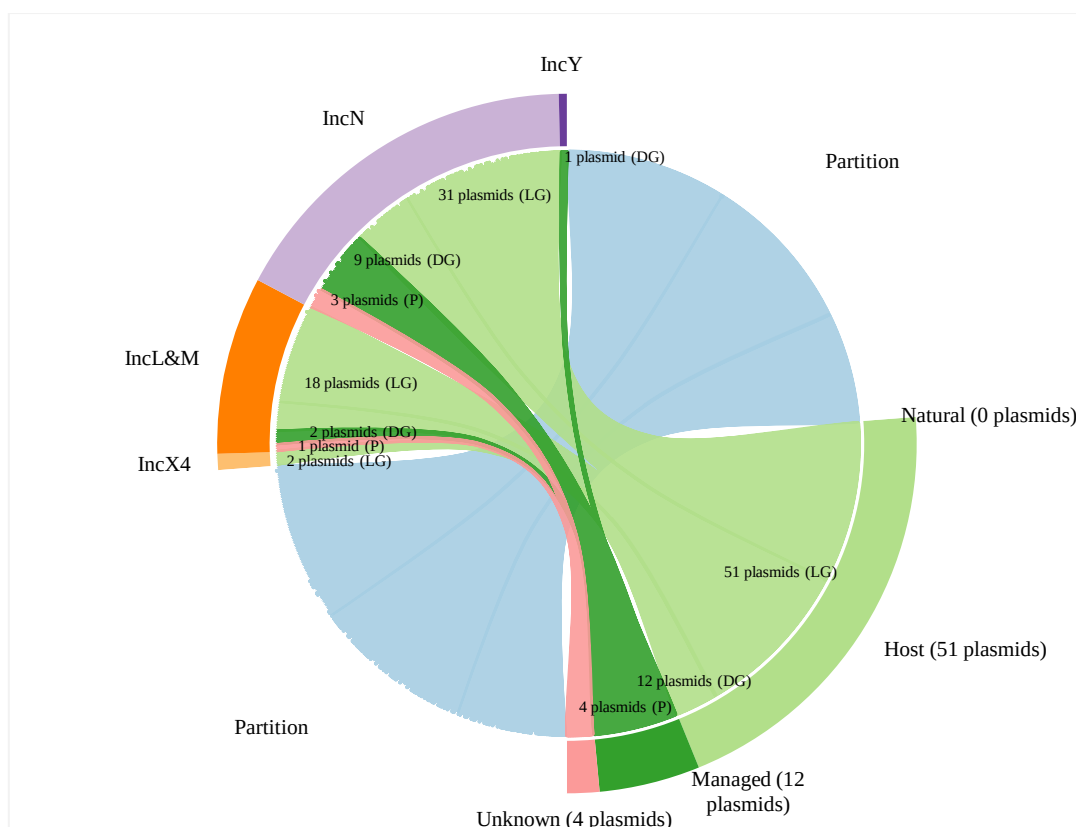


Figure 5.6c. Chord Diagram of Inc/rep groups against environment categories (Inc X4, Inc L&M, Inc N & Inc Y). The environment categories are segregated into the right portion of the diagrams, denoted as; Natural (dark blue-DB), Host (light green-LG), Managed (Dark Green-DG), Unknown (Pink-P).

In Chord Diagram of Figure 5.6d.; Inc P was the only Inc/rep group to be associated with all the 3 broad categories of environments; Natural, Host associated and Managed environments in near equal proportions which makes it statistically significant. This somehow indicates a broad diversity of this group of plasmids. By ratio [Natural: Host associated: Managed], Inc P plasmids were 6:8:2. A striking outcome is that 6 plasmids from a total of 16 were found in natural environments which emphasizes the abundance of Inc P plasmids in natural sources. Inc A & C plasmids were also associated with all the 3 broad categories of environments however it wasn't as statistically significant as Inc P plasmids because they were not in equal proportions as Inc P plasmids. By ratio [Natural: Host associated: Managed], Inc A & C plasmids were 3:41:1. Inc T only had 2 plasmids associated with host associated environments. Inc K&B also only had 10 plasmids associated with host associated environments.

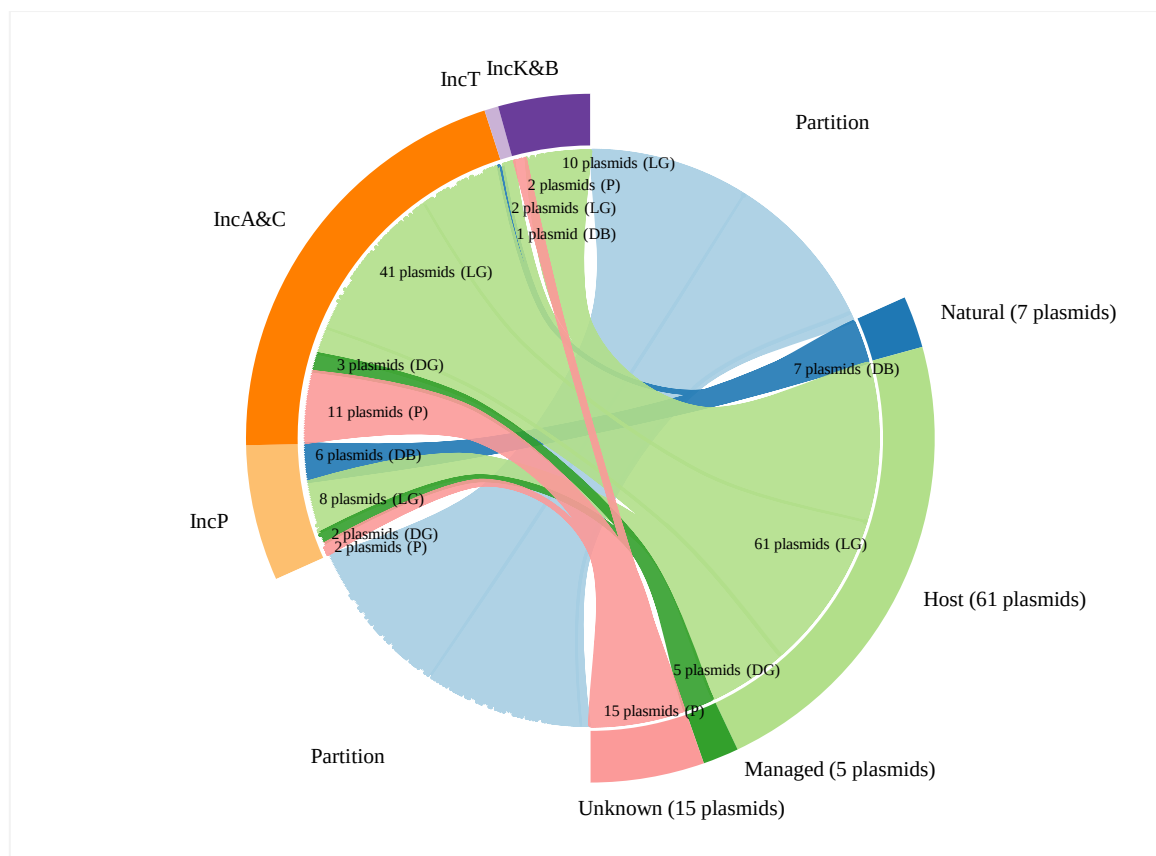


Figure 5.6d. Chord Diagram of Inc/rep groups against environment categories (Inc P, Inc A & C, Inc T and Inc K & B). The environment categories are segregated into the right portion of the diagrams, denoted as; Natural (dark blue-DB), Host (light green-LG), Managed (Dark Green-DG), Unknown (Pink-P).

In Chord Diagram in Figure 5.6e. Inc B&O with 16 plasmids and Inc R with 19 plasmids were associated exclusively with host associated environments. Inc U plasmids were associated with both host associated (7 plasmids) and managed environments (3 plasmids) with the majority being host associated. On the other hand ColE 1 like plasmids were found in host associated (69 plasmids) and natural environments (3 plasmids) with the majority being host associated (69:3).

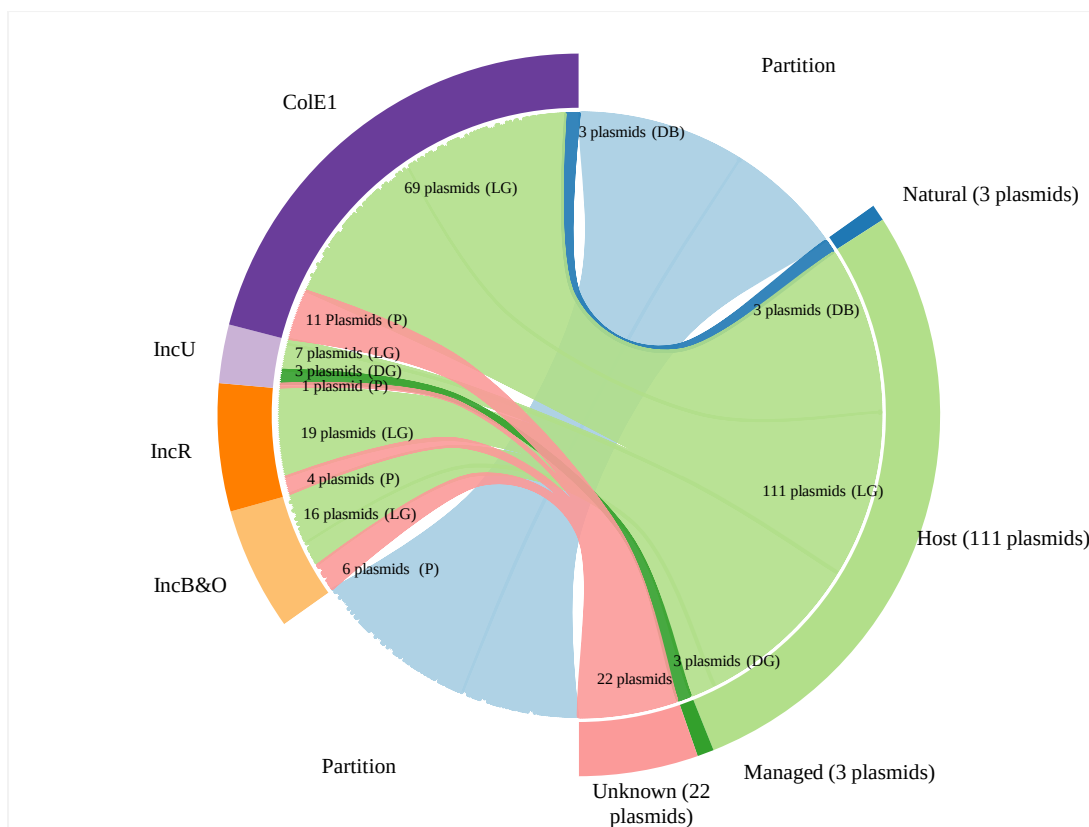


Figure 5.6e. Chord Diagram of Inc/rep groups against environment categories (Inc B&O, Inc R, Inc U & ColE1). The environment categories are segregated into the right portion of the diagrams, denoted as; Natural (dark blue-DB), Host (light green-LG), Managed (Dark Green-DG), Unknown (Pink-P).

In summary, there seems to be no clear notable variability on the correlation proportions of plasmids of a specific incompatibility group being associated with specific environmental sources for most replicon groups, an exception being Inc P plasmids. The proportions of plasmids in all the 3 broad categories of environments (Natural, Host associated and Managed environments) is generally consistent despite the different number of plasmids isolated for each Inc/rep group.

1227 5.7. *Comparative genomic analysis of completely sequenced antibiotic resistance plasmid*
1228 *groups*
1229

1230 The relationships between the different plasmid groups is depicted in the phylogenetic tree
1231 shown in Figure 5.7. Firstly, the tree shows that the root or reference out group plasmid is of
1232 Inc U origin; INC_U(b). It then shows that the most divergent recent terminal node plasmids
1233 are of Inc A&C origin; INC_A&C(a) and INC_A&C(b). The tree also shows that a majority
1234 of the Inc groups fall into lineages that have no interfering Inc group in between, exceptions
1235 being ColE1, Inc X, Inc I1, Inc U, Inc R, Inc B&O and Inc K&B. Of these, Inc B&O, Inc
1236 K&B and Inc I1 fall into a single clade; with Inc B&O and Inc K&B having the same exact
1237 reference plasmids. Inc N, Inc L&M, Inc Y, Inc P and Inc A&C fall into the same clade with
1238 the former being the outgroup and preceding in order to the latter being the terminal node. Inc
1239 W, Inc F's, Inc HI1 and Inc HI2 fall into the same clade, with Inc W and Inc F's being
1240 closely related in the terminal node and Inc HI1 being the outgroup of Inc W and Inc F's and
1241 Inc HI2 forming the outgroup of the clade. ColE1 and Inc X plasmids have no definite
1242 phylogeny but seem to distribute across a range of other Inc groups. With ColE1 being the
1243 most extreme. Inc T falls into a single clade being the second least diverse after the reference
1244 outgroup Inc U.

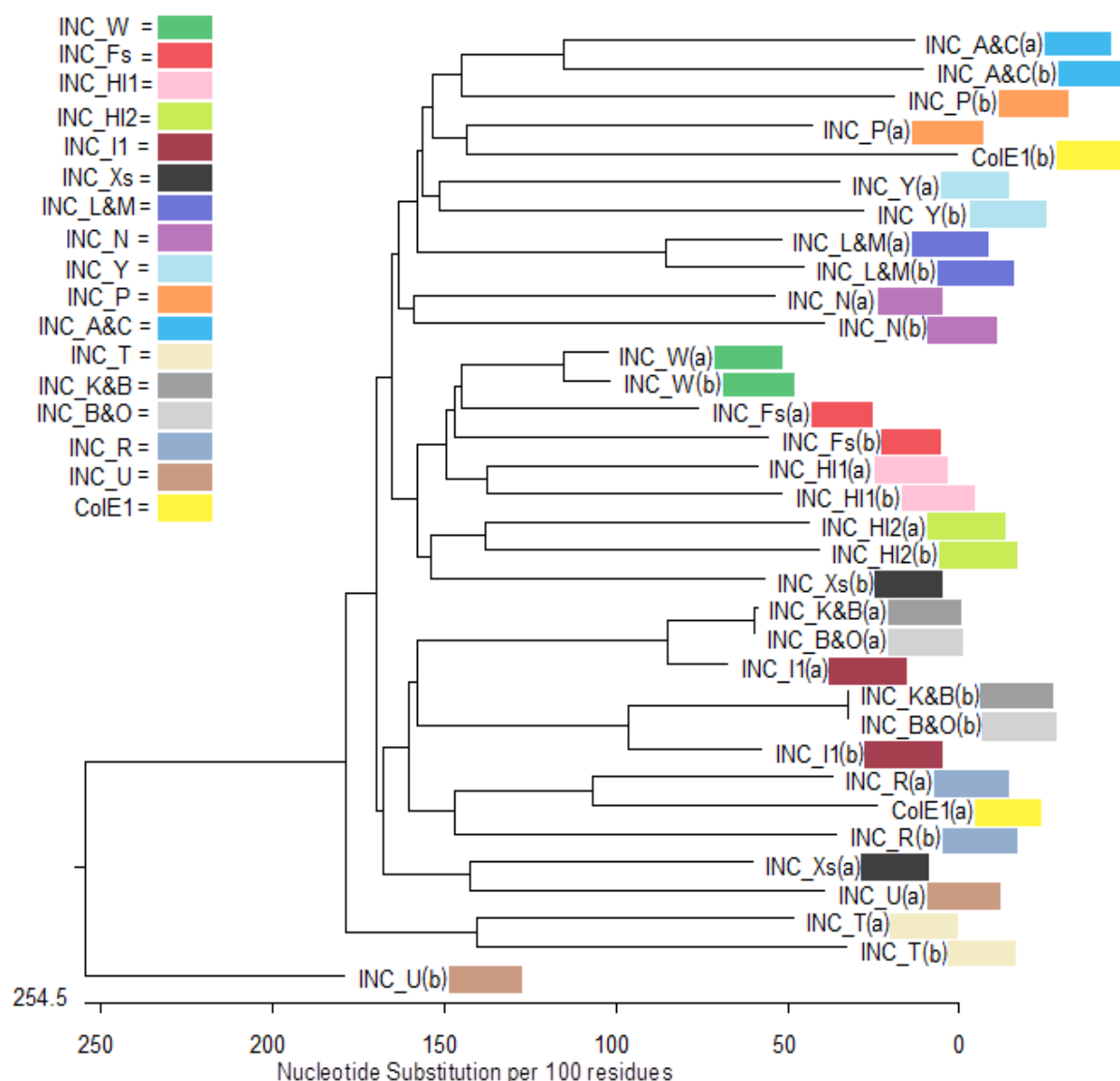


Figure 5.7. Phylogenetic tree of all Inc/rep groups with two representative plasmids from each group; one represents the most recent terminal node plasmid (a) and the other representing the root or reference outgroup plasmid (b). Emphasis is on representing cross relationships of Inc/rep groups where applicable. (Phylogenetic trees of each Inc/rep group are included as part of the appendices for reference, Appendix A1 to Appendix A35)

The phylogenetic tree results may be influenced by the genetic makeup of different plasmid acquired resistance genes. It is possible that the resistance genes amongst the plasmid groups increase evolutionary relatedness from phylogenetic interpretation. This data is important because it has shown the replicon that is most closely related to the ancestral plasmid which is Inc U.

5.8 New primer pair sequences for detection of plasmid groups

New primer sequences were developed to diagnose and complement generic primer sequences that were initially designed for the PBRT. All new primers were tabulated against their previous counterparts to compare differences and relative number of new primers in each respective Inc/rep group. All notable differences are highlighted according to a scheme described at the bottom of Table 5.8.

Table 5.8. New primer sequences developed to complement generic primer sequences

Inc/rep	Generic replicon typing primers (5'- 3' direction)	New replicon typing primers indicating mutations from generic primers (5'- 3' direction)
1 INC W ^a	FW: CCTAAGAACAACAAAGCCCCG RV: GGTGCGCGGCATAGAACCGT	NO PRIMER DIAGNOSIS REQUIRED (FW & RV)
2 INC F(repB) ^a	FW: GATCGTTTAAGGAATTTTG RV: GAAGATCAGTCACACCATCC	FW1: GATCGTTT T AAGGAATTTTG FW2: GATCGTTTAAGGAATTT R FW3: GATCGTTTA CA GATT CTGC FW4: CTGCGAGAGTAGAGTTTTTG RV1: CA TGATTAG CC CACACCATCC RV2: CA TATCAGTCAC CGCT TCC RV3: CCG TCAGTCACAC CT TCC
3 INC FIA ^a	FW: CCATGCTGGTTCTAGAGAAGGTG RV: GTATATCCTTACTGGCTTCCGCAG	FW1: CCATGCKGGTT YY AGAGRRGG TR FW2: CCATGCKGGTTC Y AG R GA R GG WK RV1:GTATATCCTTACT R GCTTCCGCAG RV2:GTATATCCTTACTGGCTTC M G C R G
4 INC FIB ^a	FW: GGAGTTCTGACACACGATTTTCTG RV: CTCCCGTCGCTTCAGGGCATT	FW1: GGAGTTCTGAC R CACGATTTTCTG FW2: GGR R GTTCTGACAY M CGATTTTCTG FW3: GGR R GTTCTGACACACGATTTTCTG FW4: GGR GY CTGACACACGATTT K CTG FW2: GG MSW TCTRA Y RM A MA G A K KK CTG RV1:CTCCCG K CGCTTCAGGGCATT
5 INC FIC ^a	FW: GTGAAGTGGCAGATGAGGAAGG RV: TTCTCCTCGTCGCCAAACTAGAT	FW1: GTGAAGTGGC R GATGAGGAAGG FW2: GTGAR CTK GCW GATGAR G AM GG FW3: KKWRM KK R C W CMKRR R GARR G FW4: GGTTGAGTACTGCGAGAGAGAG RV1: YYT Y CT M GTCGCCAAACT W G A W RV2: YYT C YYM GTCGCCAAACT W K M T RV3: YYT YY SK W C K CSMM A MWR K W K

			RV4: TCTTTTCGTACTCGCCAAAGTTC
			FW1: CTGYCGTAAGCTGATGGC FW2: CTGYSRWAARCTGATGGR
6	INC FII ^a	FW: CTGTCGTAAGCTGATGGC RV: CTCTGCCACAACTTCAGC	RV1:CTCTGCCAYAACTTCAGC RV2:CTCYGCCACAACTTCAGC RV3:CTCYGCCACAAACKTCAGC RV4:CTCYGCCAMAACTTCAGC RV5:CTCTGMCASRAACKTCAGS
7	INC FIA ^b	FW: CCATGCTGGTTCTAGAGAAGGTG RV: GTATATCCTTACTGGCTTCCGCAG	FW1: CCATGCKGGTTYAGAGGGGTA FW2: CCATGCKGGTTCYAGCGAGGAT FW3: CCAYGSKGRTWYTWKWRWARKTR FW4: TTTTCACAAAGTTATCCMTS FW5: TTCACATGGARCTGTCATGGCGGA RV1:GTATATCCTTACTRGCTTCCGCAG RV2:GTATATCCTTACTGGCTTCMGRG
8	INC FIB ^b	FW: TCTGTTTATTCTTTTACTGTCCAC RV: CTCCCGTCGCTTCAGGGCATT	FW1: KCTGTTTATTCTTTTACTGTCCAC NO PRIMER DIAGNOSIS REQUIRED (RV)
9	INC FIBs ^b	FW: TGCTTTTATTCTTAAACTATCCAC RV: CTCCCGGCGCTTCAGGGCATT	FW1:TCTGTTTATTCTTTTACTGTCCAC FW2:GCTGTTTATTCTTTTACTGTCCAC FW3:TGSYTKYWTTYWTWMWYWAWCYAW FW4:TCGYTKYWTTYWTWMWMAWCMMAW FW5: ACCACCCATCTGGTGCAGCC RV1:CTCCCGKCGCTTCAGGGCATT RV2:CTCCCGGCGYTCAGGGCATT RV3:YTCCYKCGCTKCAGGGCRTY
10	INC FIC ^b	FW: GTGAAGTGGCAGATGAGGAAGG RV: TTCTCCTCGTCGCCAACTAGAT	FW1: GTGAAGTGGCRGATGAGGAAGG FW2: GTGARCTKGCWGATGARWAMGG FW3: GWGWACTGSSAGAKRRRRRRGR RV1:TYTYCTWGTGCCAAACTWGAW RV2:TYTCTYMGTCGCCAAACTWKMT RV3: TTYTCSIMSTCGCCAAASTWGAW
11	INC FII ^b	FW: CTGATCGTTTAAGGAATTTT RV: CACACCATCCTGCACTTA	FW1: CTGATCGTTTAAGGAATTTT FW2: CTGMKMGWKTARRGTTTTTK FW3: CTGATCGTTTARRGAWTTYT RV1:CACACCATCCTGCRCTTA RV2:CACACCWTCCTGCACTTA RV3: CACMSCWTCTGCRCTTA
12	INC FII(repA) ^b	FW: CCAAACGTATTACCGCCAGGTAA RV: CGGTTACGTGACAGAATCATGCGC	FW1: CCAAACRTATTACCGCCAGGTAA FW2: CCAAACKTATTACCGCCAGGTAA FW3: CCAAA MRTASTAYYSMCAGGTWA FW4: CCAAACSYASTAYYSMCAGGTWA FW5: GGCTTCASBTCCCGTTTTC RV1:YKKYYRCGYGACAGMAKCATGCGC

			RV2: YKRYR CGTGACA KWATCATGCGC RV3: TCSTATTCSGTCTGRTARSTRATS
13	INC FII[Fd] ^b	FW: CTGATCGTTTTAAGGAATTTT RV: GGAATGTAGACCCCGAT	NO PRIMER DIAGNOSIS REQUIRED (FW & RV)
14	INC FIIk ^b	FW: TCTTCTCAATCTTGCGGA RV: GCTTATGTTGCACRGAAGGA	NO PRIMER DIAGNOSIS REQUIRED (FW & RV)
15	INC FIIIs ^b	FW: CTAAAGAATTTTGATGGCTGGC RV: CAGTCACTCTGCCTGCAC	FW1: MTAAAGAATTTTGATGGCTGGC FW2: YTAARGAATTTTGATGGCTGGC FW3: CTRAAKAATTTTGGGYKKGC FW4: SWRARKARWKTTKW TGGYKKGC RV1: CAGTCACYKCTKCCTGCRC RV2: CAGTCACWYCWK CCTGCAC RV3: YAGKACACWYCWK CCTGCAC RV4: YAKTCAMWYCY GCCTGCAC
16	INC FIIy ^b	FW: TGGYAGGGAAGTGGTTCTG RV: GTRAGTCACACCTTCCCGC	NO PRIMER DIAGNOSIS REQUIRED (FW) RV1: RTRAGTCACMSCTTCCYGC
17	INC HI 1 ^a	FW: GGAGCGATGGATTACTTCAGTAC RV: TGCCGTTTCACCTCGTGAGTA	FW1: GGAGCGATGGMTKACTTCAGTAC FW2: KGARCGRTGRMYKACTTCWGTMC RV1: TGCCGTWTCACCTCGTGARTA RV2: YGCMGTCTACCCWCGTGARTA
18	INC HI 2 ^a	FW: TTTCTCCTGAGTCACCTGTAAACAC RV: GGCTCACTACCGTTGTCATCCT	NO PRIMER DIAGNOSIS REQUIRED (FW) RV: GGCTCACTACCGTTGTCAVCS
19	INC I 1 ^a	FW: CGAAAGCCGGACGGCAGAA RV: TCGTCGTTCCGCCAAGTTCGT	FW1: CGAAAGYCGGACGGCAGAA FW2: CGAAAGCCGGACGGYAGAA FW3: SGAAAGCCRGAMRRCRGMA FW4: SGRAAGYCRGAMRRYRGMA FW5: SGRAAGYCRGAMRRRCRGMA RV1: TCGTC KTTCCGCCAAGTTCGT RV2: TCGTCGYTCCGCCAAGTTCGT RV3: TCKYCGTYCCGCCAAGTTCGT RV4: TCKKCGTTCCGCCAAGTTCGW RV5: TYSWCGAKYCCGCCAAGTTCGW RV6: YYRWCGTTCCGCCAAGTTCGW RV7: TSKKCGTTCCGCCAAGTTCGW
20	INC X ^a	FW: AACCTTAGAGGCTATTTAAGTTGCTGAT RV: TGAGAGTCAATTTTATCTCATGTTTATAGC	NO PRIMER DIAGNOSIS REQUIRED (FW) RV1: TGAGAGTCAATT TTTTATCTCATGTTTATAGC RV2: TGAGAGTCAATT TTTTATCTCATGTTTATAGC
21	INC X1 ^d	FW: GCTTAGACTTTGTTTTATCGTT RV: TAATGATCCTCAGCATGTGAT	FW1: GCTTWGRYTKGTTTTATCGTT FW2: GCTTMGRCTTTGTTTTATCRK RV1: TAATGATCCTCARCATGTGAT RV2: WAAVGA VCCTCARCATGTAT

			RV3: TAATGAYCCTCARRATGTKWT
			FW1: GYGAMGAAATYMWGGAAGCTA FW2: GYGAMGAAATYMWGGARGCTA FW3: GYGAMGARATYMARARGCTA FW4: GYGAMGARATMMARGAAGCWA FW5: GYGAWGARATYMARARGCTA
22	INC X2 ^d	FW: GCGAAGAAATCAAAGAAGCTA RV: TGTTGAATGCCGTTCTTGTCAG	RV1: TGYTGAAYGCCRTTYTTRTCMAG RV2: TGYTGAAYRCCGTTYTTRTCSAG RV3: TGYTGAAYRCCRTTYTTRTCSAG RV4: TGYTGWAYRCCGTTCTTRTCMAG
			FW1: GTTTTCTCMACSCCCTTGTTMA FW2: GTKKWYTCMACGCCYTTRTTCA FW3: GTTTTYTCMACRCCYTGTGTTMA FW4: GTWIWYICYWYGCCMITRWKYR FW5: GTWWYCTSYWYKCCMYKTTYA
23	INC X3 ^d	FW: GTTTTCTCCACGCCCTTGTTCA RV: CTTTGTGCTTGGCTATCATAA	RV1: CTTTGTGCTKGGCTATCATAA RV2: YTTTGTGCTTGGCTATCATAA RV3: YTTTGTKCTTGGMTAYCAYWR RV4: YTTTGTKCTTGGMTATCATWR RV5: YTTYGTGCTKGSCTATCAYAM RV6: YTTYGKSYYGKMTAYCATRA RV7: MTTTGTGMTKGYTAYCATRA
24	INC X4 ^d	FW: AGCAAACAGGGAAAGGAGAAGACT RV: TACCCCAAATCGTAACCTG	NO PRIMER DIAGNOSIS REQUIRED (FW & RV)
25	INC L & M ^a	FW: GGATGAAAACATCAGCATCTGAAG RV: CTGCAGGGGCGATTCTTTAGG	FW1: GGRTGAAAAATATCWSCKYKGMRR RV1: CTGCWGGGGYGATTCTTTAGG RV2: CAGACCCCATACCCCTGAG
26	INC N ^a	FW: GTCTAACGAGCTTACCGAAG RV: GTTTCAACTCTGCCAAGTTC	FW1: GTCTAACGAGCTTACCGARG RV1: GTTTCAAYTCTGCCARRTTC
27	INC Y ^a	FW: AATTCAAACAACACTGTGCAGCCTG RV: GCGAGAATGGACGATTACAAAACCTT	FW1: AATTCAAAAYAACACYGTCAGCCTG RV1: GCGCTTGTGAATAATGAGGTAGTTCT
			FW1: TGGCGGGCCTTCCTGGCCCGTTT
28	INC P ^a	FW: CTATGGCCCTGCAAACGCGCCAGAAA RV: TCACGCGCCAGGGCGCAGCC	RV1: TCACGCGCAACAGGGCGCAGCC RV2: CTGCCCCCTCATCTGTCAACG RV3: CGCCCCYCAWSTGTCAGCA
29	INC A & C ^a	FW: GAGAACCAAAGACAAAGACCTGGA RV: ACGACAAACCTGAATTGCCTCCTT	FW1: RAGARCYARAGACAWAGACYTGGA FW2: RARARCSAAAGACAAAGMYSISGA RV1: ACGACAAACCTGRATTGCYTCTT RV2: RCGACAAACMWGRATTGCYTCYTT RV3: WCGRCARAYCTGAATWGCYTCYTT
30	INC T ^a	FW: TTGGCCTGTTTGTGCCTAAACCAT RV: CGTTGATTACACTTAGCTTTGGAC	FW1: TTGGCCTKITTTGTGCCTAAACCAT NO PRIMER DIAGNOSIS REQUIRED

31	INC K & B ^a	FW: GCGGTCCGGAAGCCAGAAAAC RV: TCTTTCACGAGCCCGCCAAA	FW1: RSGRTCCGGR [*] AAG [*] YCAGAAAAC
			FW2: RSGRTCCGGR [*] AAG [*] YCAGAAAAY
32	INC B & O ^a	FW: GCGGTCCGGAAGCCAGAAAAC RV: TCTGCGTTCGCCAAGTTCGA	FW3: AGCAAAGCGGCGAAAGCCGGA
			RV1: T [*] TTTC [*] WSSRKY [*] CCGCCAAR [*]
33	INC R ^c	FW: TCGCTTCATTCCTGCTTCAGC RV: GTGTGCTGTGTTATGCCTCA	RV2: TCTTTC [*] RYSRKY [*] CCGCCAAR [*]
			RV3: TCTTCTAACGTT [*] CCGCCAAR [*]
34	INC U ^c	FW: TCACGACACAAGCGCAAGGG RV: TCATGGTACATCTGGGCGC	FW1: GCGGCGAAAGCCGGACGGC
			FW2: RSGRTCCGGR [*] AAG [*] YCAGAAAAY
35	ColE1 ^c	FW: GTTCGTGCATACAGTCCA RV: GGCGAAACCCGACAGGACT	RV1: T [*] TGCGTTCGCCAAGTTCGA
			RV2: TCT [*] SCGT [*] YCCGCCAAGTTCG [*] W
			RV3: TCK [*] CGTTCGCCAAGTTCG [*] W
			RV4: YYWR [*] CGTTCGCCAAGTTCGA
			NO PRIMER DIAGNOSIS REQUIRED
			RV1: GCTGCCTCATCGCTAACTTTG
			FW1: TCACGACACAR [*] SCSSRM [*] RGG
			FW2: TCAGACGACAGGGATTCTTC
			FW3: TCACCGACAGCATCAGGTTG
			RV1: TCR [*] TGGTACATCTGGSCG [*] S
			RV2: AGATACGCCAGCGGCC
			FW1: GTTCGTGCATAY [*] AGTCCA
			FW2: GTTC [*] TGCATACAGTCCA
			FW3: GTTCGTGCAY [*] ACAGTCCA
			FW4: GTTC [*] TGCATACAG [*] YCCA
			FW5: GTTCGTGCAY [*] ACAG [*] YCCA
			FW6: GTTC [*] TGC [*] W [*] TACAG [*] YCCA
			FW7: GTTC [*] TGCAY [*] ACAG [*] YCCA
			RV1: GR [*] CGAAACCCGACAGGACT
			RV2: GGCGAAACCC [*] Y [*] GACAGGACT
			RV3: GGCGAAACCCGAY [*] RS [*] GR [*] CT
			RV4: AAGCCCTTATAGTACAATA

^a = Replicon primers developed by Carattoli et al 2005, ^b = Replicon primers developed by Villa *et al* 2010, ^c = Replicon primers developed by García-Fernández *et al* 2009, ^d = Replicon primers developed by Johnson *et al* 2012. Yellow highlight indicates insertion, a deletion is indicated by asterisk [*]. A purple highlight indicates substitution mutation and a green highlight indicates a newly developed primer sequence.

Table 5.9. is a summary of the IUPAC system for nucleotide nomenclature that is used here to define the degenerate base pairs.

Table 5.9. IUPAC system for nucleotide nomenclature.

SINGLE LETTER CODE	NUCLEOTIDE/S	EXPLANATION
A	A	A denine
C	C	C ytosine
G	G	G uanine
T	T	T hymine
I	I	I nosine
R	A or G	p u R ine
Y	C or T	p Y rimidine
M	A or C	a M ino
K	G or T	K eto
S	C or G	S trong interaction
W	A or T	W weak interaction
H	A or C or T	not G, H follows G alphabetically
B	C or G or T	not A, B follows A alphabetically
V	A or C or G	not T/U, V follows U alphabetically
D	A or G or T	not C, D follows C alphabetically
N	A or C or G or T	a N y

It has clearly been noted that the PBRT method has its limitations, since it cannot characterize all replicons of plasmids and that replicons are not exclusive to one plasmid (multiple replicon plasmids). Another problem is that the PBRT method works by probing DNA for specific sequences therefore mutations that occur over time make the primer annealing region of the amplicon mismatched with its corresponding primer. This mutation can accumulate to an extent that if left unchecked, the hybridisation sequence will no longer be recognizable. Mutations associated with this problem commonly include base substitutions and indel (insertion-deletion) mutations which result in single nucleotide polymorphisms. In rare cases such as in some Inc P plasmids, the mutations can involve genome rearrangements within a plasmid which would lead to the truncation of the replicon primer amplification region (amplicon). In this instance, only one primer would be effective with a distinguishable recognition sequence and the other having no recognition sequence i.e. totally absent. Upon primer alignments with target sequences (per Inc/rep group), SNP mutation were predominant source of primer errors (Appendices; clause 6.1.2 [Appendix B1 to Appendix B70]). From this diagnosis, 20 new primers were created and 148 primers with one or more degenerate base positions were redesigned from the generic primer sequences as shown in Table 5.8. These primers comprise all Inc/rep groups including Inc F and subgroups developed both by Carattoli then Villa and colleagues (Carattoli *et al* 2005), (Villa *et al* 2010).

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS AND RECOMMENDATIONS

The plasmid sequence data used in this project was collected from the NCBI nucleotide reference database at a single point in time (November 15 2015) when the complete set of plasmid sequence data was representative of all the known sequences at that point in time. Due to the fact that all the sequences used in this project were only sampled at a point in time, they are only representative of that particular point in time and lose accuracy over time with the introduction of new sequence data. However, the information used in this research is still relevant to the analysis of plasmid-acquired resistance in general because the deviation caused by the introduction of new sequence data is minimal. The plasmid reference database sequences have been updated and restructured continuously in NCBI since the time of sampling. As a general outlook, the quantity of available plasmids online in NCBI retrieved and used for research purposes is accumulating overtime; Smillie and colleagues reported 1730 plasmids (early 2009), Pal and colleagues reported 4582 plasmids (June 2014), Shintani and colleagues reported 4602 plasmids (August 2014), this study retrieved 5682 complete sequences (November 2015) adding to the exponential increase of complete plasmid accessions added in NCBI overtime since 2009. These increases correspond with increasing knowledge and applications of next-generation sequencing and bioinformatics.

The preliminary analysis and filtering of data to allow only the most relevant plasmid sequences to be used for addressing the project objectives was crucial in removing sequences that would distort results accuracy and relevance. Before the beginning of this project no direct method had been described thoroughly in literature as a standard protocol for data sorting. Conventionally, there isn't any standard protocol because each study will focus on pre-processing sequence data for the research interests of that particular study. This study was able to determine the frequency of specific ARGs in plasmid groups and revealed that the genes associated with resistance to Beta-Lactams, Aminoglycosides and Antibiotic efflux pumps occur in majority of plasmids. This is because all these antibiotics were the first generation antibiotics to be used as therapeutic agents. Other antibiotic resistance gene classes may have not been found at higher frequencies relative to others due to late introduction of the associated antibiotics in clinical settings e.g. fluoroquinolone, chloramphenicol and macrolides. There were no unique ARGs classes associated with each

1330 incompatibility group, and this information could suggest that ARGs are moving randomly
1331 from one plasmid to another and not necessarily in related groups.

1332 The comparative genomic analysis revealed that the replicon most closely associated to
1333 plasmids that emerged early in existence is Inc U. However, it should be noted that this
1334 interpretation is based solely on resistance plasmids that were selected in the analysis. For a
1335 more compelling conclusion, all natural plasmids including those that have no resistance also
1336 have to be considered. The design of new sequence typing primers to complement those that
1337 already exist in the PBRT was successfully performed. Although these newly designed
1338 primers have not been tested *in vitro* for sensitivity and specificity, the study has shown the
1339 need to revise and develop new markers for PCR studies. There are other molecular typing
1340 methods for identification of plasmids that have been developed as an alternative for the
1341 shortcomings of the replicon typing method based on mobility of plasmids. This method is
1342 called mob typing and it was conceptualised by Francia and colleagues (2004) through the
1343 study of plasmid conjugation. The method is useful for typing incompatibility in conjugative
1344 and mobilizable plasmids only. The implementation of this technique as a typing approach
1345 was beyond the scope of this research since this research was based solely on replicon typing.
1346 However the use of this approach is open for investigation in future research

1347 Based on the methods utilized in this study for investigating the ecology and diversity of
1348 plasmids derived from sequence data, the majority of antibiotic resistance plasmids were
1349 found in host associated environments particularly in human patients from clinical settings. It
1350 should be noted that the majority of these plasmids were sequenced using a genomics
1351 approach which relies on primary isolation and culture of host organism. The results suggest
1352 no clear or notable variability on the proportions of plasmids of a specific incompatibility
1353 group being associated with specific environmental sources. The major limitation within this
1354 bioinformatics study was the plasmid sequences that were not completely sequenced and
1355 those not clearly described in terms of source or isolation. It is recommended that sequence
1356 data submission to public sequence databases be improved for future bioinformatics studies
1357 that aim to look particularly at evolution of plasmids encoding antibiotic resistance.

REFERENCES

- Abraham, E. P., & Chain E. (1940). An enzyme from bacteria able to destroy penicillin. *Nature* 46 (3713): 837–837.
- Alanis, A.J. (2005). Resistance to antibiotics: *Are we in the post-antibiotic era?* *Arch Med Res* 36, 697-705.
- Alvarado, A., Garcillán-Barcia, M.P., and De la Cruz, F. (2012). A degenerate primer MOB typing (DPMT) method to classify gamma-proteobacterial plasmids in clinical and environmental settings. *PLoS One*. 2012;7(7)
- Bennett, P.M. (1999). Integrons and gene cassettes: *a genetic construction kit for bacteria*. *J Antimicrob Chemother* (1999) 43 (1): 1-4.
- Bennett, P.M. (2004). Genome plasticity: *insertion sequence elements, transposons and integrons, and DNA rearrangement*. *Methods Mol Biol*. 2004; 266:71-113.
- Bennett, P.M. (2008) Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *British Journal of Pharmacology* 153, S347-S357.
- Bergquist, P. L., Saadi, S., Maas, W. K., (1986). Distribution of basic replicons having homology with RepFIA, RepFIB, and RepFIC among IncF group plasmids. *Plasmid*. 1986 Jan;15(1):19-34.
- Carattoli, A. (2013). Plasmids and the spread of resistance. *Int J Med Microbiol*. 2013 Aug;303(6-7):298-304.
- Carattoli, A. (2009). Resistance Plasmid Families in *Enterobacteriaceae*. *Antimicrob. Agents Chemother*. June 2009 vol. 53 no. 6 2227-2238.
- Carattoli, A., Bertini, A., Villa, L., Falbo, V., Hopkins, K. L., and Threlfall, E. J., (2005). Identification of plasmids by PCR-based replicon typing. *J Microbiol Methods*. 2005 Dec; 63(3):219-28.
- Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., MøllerAarestrup, F., and Hasman, H., (2014). In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother*. 2014 Jul; 58(7): 3895-903.
- CDC. (2013). Centre for Disease Control and Prevention: *Antibiotic resistance threats in the United States*, 2013 report. Online: available at; <<http://www.cdc.gov/drugresistance/threat-report-2013/pdf/ar-threats-2013-508.pdf#page=13>>.

1392 Couturier, M., Bex, F., Bergquist, P. L., and Maas, W. K., (1988). Identification and
1393 classification of bacterial plasmids. *Microbiol Rev* 52:375-95.

1394 Datta, N., and Hedges, R. W., (1972). R factors identified in Paris, some conferring
1395 gentamicin resistance, constitute a new compatibility group. *Ann Inst Pasteur*
1396 (Paris) 123:849-52.

1397 Davies, J., & Davies, D. (2010) Origins and evolution of antibiotic resistance. *Microbiol*
1398 *Mol Biol Rev* 74, 417-433.

1399 D'costa, V.M., Mcgrann, K.M., Hughes, D.W., & Wright, G.D. (2006) Sampling the
1400 antibiotic resistome. *Science* 311, 374-377.

1401 Dimitriu, T., Misevic, D., Lindner, A. B., and Taddei, F. (2015). Mobile genetic
1402 elements are involved in bacterial sociality. *Mob Genet Elements*. 2015 Jan-Feb;
1403 5(1): 7–11.

1404 Francia M, V., Varsaki, A., Garcillán-Barcia, M. P., Latorre, A., Drainas, C., and De
1405 la Cruz. F (2004) A classification scheme for mobilization regions of bacterial
1406 plasmids. *FEMS Microbiol Rev*. 2004 28; 79–100.

1407 Funnell, B. & Phillips G. (2004). Preface, in *Plasmid Biology*, Funnell B., Phillips G.,
1408 editors. Washington, DC: ASM Press, xi.

1409 García-Fernández, A., Fortini, D., Veldman, K., Mevius, D., and Carattoli, A.,
1410 (2009). Characterization of plasmids harbouring *qnrS1*, *qnrB2* and *qnrB19* genes
1411 in *Salmonella*. *J Antimicrob Chemother*. 2009 Feb; 63(2):274-81.

1412 Garcillán-Barcia M, P., Francia, M, V., and De la Cruz, F., (2009) The diversity of
1413 conjugative relaxases and its application in plasmid classification. *FEMS*
1414 *Microbiol Rev*. 33(3):657-87

1415 Garcillán-Barcia, M. P., Alvarado, A., and De la Cruz, F., (2011). Identification of bacterial
1416 plasmids based on mobility and plasmid population biology. *FEMS Microbiol Rev*.
1417 2011 Sep;35(5):936-56.

1418 Hall, R. M. (1997). Mobile gene cassettes and integrons: moving antibiotic resistance
1419 genes in gram-negative bacteria. *Ciba Found Symp*. 1997; 207:192-205.

1420 Hanza, H., (2008). Linear plasmids in plant mitochondria: *peaceful coexistences or*
1421 *malicious invasions?*. *Mitochondrion*. 2008 Jan;8(1):15-25

1422 Hinnebusch, J., and Tilly, K., (1993). Linear plasmids and chromosomes in bacteria.
1423 *Mol Microbiol*. 1993 Dec;10(5):917-22.

1424 Hugenholtz, P & Tyson, G. W. (2008). Metagenomics. *Microbiology*. *Nature*, 455 (25),
1425 481-483.

1426 Jia, B., Raphenya, A.R., Alcock, B., Waglechner, N., Guo, P., Tsang, K.K., Lago,
1427 B.A., Dave, B.M., Pereira, S., Sharma, A.N., Doshi, S., Courtot, M., Lo, R.,
1428 Williams, L.E., Frye, J.G., Elsayegh, T., Sardar, D., Westman, E.L., Pawlowski,
1429 A.C., Johnson, T.A., Brinkman, F.S., Wright, G.D., McArthur, A.G., (2017).
1430 CARD 2017: *expansion and model-centric curation of the comprehensive*
1431 *antibiotic resistance database*. Nucleic Acids Res. 2017 Jan 4; 45.

1432 Johnson, T.J., Bielak, E.M., Fortini, D., Hansen, L.H., Hasman, H., Debroy, C.,
1433 Nolan, L.K., and Carattoli, A., (2012). Expansion of the IncX plasmid family for
1434 improved identification and typing of novel plasmids in drug-resistant
1435 Enterobacteriaceae. Plasmid. 2012 Jul;68(1):43-50.

1436 Kado, C.I., (1998). Origin and evolution of plasmids. Antonie Van Leeuwenhoek.
1437 1998 Jan;73(1):117-26

1438 Kline, B. C., (1985). A review of mini-F plasmid maintenance. Plasmid 1985, 14:1-6.

1439 Lederberg, J., (1952). *Cell genetics and hereditary symbiosis*, Phy. Rev 32, p403-414/5

1440 Liu, B., & Pop, M. (2009). ARDB-Antibiotic resistance genes database. Nucleic Acids Res
1441 37, D443-D44

1442 Luscombe, N. M., Greenbaum, D., and Gerstein, M., (2001). What is bioinformatics? An
1443 introduction and overview. Yearb Med Inform. 2001;(1):83-99.

1444 Marmur, J., Rownd, S. Falkow, La S. Baron, CaSchildkraut, and P. Doty. (1961). The
1445 nature of intergeneic episomal infection. Proc Natl Acad Sci U S A. 47:972-979.

1446 Martinez Villegas P. A. (2012). Exploratory Metagenomic Analysis of Antibiotic
1447 Resistance Genes in Bacterial Communities: A statistical approach for classification
1448 of the bacterial resistome. Online: available at;
1449 <<http://publications.lib.chalmers.se/records/fulltext/160023.pdf>>.

1450 Martínez, J. L., Coque, T. M., and Baquero, F., (2015). What is a resistance gene? *Ranking*
1451 *risk in resistomes*. Nat Rev Microbiol. 2015 Feb;13(2):116-23.

1452 McArthur, A. G., and Wright, G.D., (2015). Bioinformatics of antimicrobial resistance
1453 in the age of molecular epidemiology. Curr Opin Microbiol. 2015 Oct; 27:45-50.

1454 McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J.,
1455 Bhullar, K., Canova, M. J., De Pascale, G., Ejim, L., Kalan, L., King, A. M., Koteva,
1456 K., Morar, M., Mulvey, M. R., O'Brien, J. S., Pawlowski, A. C., Piddock, L. J.,
1457 Spanogiannopoulos, P., Sutherland, A. D., Tang, I., Taylor, P. L., Thaker, M., Wang,
1458 W., Yan, M., Yu, T., & Wright, G. D., (2013). The comprehensive antibiotic
1459 resistance database. Antimicrob. Agents Chemother. 2013; 57:3348–3357.

1460 Nikaido, H., (2009). Multidrug resistance in bacteria. Annu Rev Biochem. 2009;
1461 78:119-46.

1462 Novick, R. P., (1987). Plasmid incompatibility. *Microbiol Rev* 51:381-95.

1463 Novick, R. P., Clowes, R. C., Cohen, S. N., Curtiss III, R., Datta, N., and Falkow, S.
1464 (1976). Uniform nomenclature for bacterial plasmids: a proposal. *Bacteriol. Rev.*
1465 40:168-189.

1466 Orlek, A., Phan, H., Sheppard, A. E., Doumith, M., Ellington, M., Peto, T., Crook, D.,
1467 Walker, A. S., Woodford, N., Anjum, M. F., Stoesser, N. (2017). Ordering the
1468 mob: Insights into replicon and MOB typing schemes from analysis of a curated
1469 dataset of publicly available plasmids. *Plasmid*. 2017 May;91:42-52.

1470 Pal, C., Bengtsson-Palme, J., Kristiansson, E., Larsson, D.G., (2015). Co-occurrence
1471 of resistance genes to antibiotics, biocides and metals reveals novel insights into
1472 their co-selection potential. *BMC Genomics*. 2015 Nov 17;16:964.

1473 Perry, J.A., and Wright, G.D. (2014). Forces shaping the antibiotic resistome. *Bioessays*.
1474 2014 Dec;36(12):1179-84.

1475 Pruitt, K.D., Tatusova, T., and Maglott, D.R., (2007). NCBI reference sequences
1476 (RefSeq): a curated non-redundant sequence database of genomes, transcripts and
1477 proteins. *Nucleic Acids Res*. 2007 Jan; 35(Database issue).

1478 Rahube, T.O., Viana, L., Koraimann, G., and Yost, C. K., (2014). Characterization
1479 and comparative analysis of antibiotic resistance plasmids isolated from a
1480 wastewater treatment plant. *Front Microbiol*. 2014 Oct 28; 5:558.

1481 Rankin, D. J., Rocha, E. P. C., and Brown, S. P. (2011). What traits are carried on
1482 mobile genetic elements, and why?. *Heredity*. 2011 Jan; 106(1): 1–10.

1483 Riesenfeld, C. S., Schloss, P. D., and Handelsman, J. (2004). METAGENOMICS:
1484 Genomic Analysis of Microbial Communities. *Annu Rev Genet*, 325-352.

1485 Seshadri, R., Kravitz, S. A., Smarr, L., Gilna, P., & Frazier, M. (2007). CAMERA: A
1486 *Community Resource for Metagenomics*. *PLoS Biology*, V (3), 0394-0397.

1487 Shintani, M., Sanchez, Z. K., and Kimbara, K., (2015). Genomics of microbial plasmids:
1488 *classification and identification based on replication and transfer systems and host*
1489 *taxonomy*. *Front Microbiol*. 2015; 6: 242.

1490 Smillie, C., Garcillán-Barcia, M. P., Francia M.V., Rocha E. P., and De la Cruz, F.
1491 (2010). Mobility of plasmids. *Microbiol Mol Biol Rev*. 2010 Sep; 74(3):434-52.

1492 Villa, L., García-Fernández, A., Fortini, D., and Carattoli, A., (2010). Replicon
1493 sequence typing of IncF plasmids carrying virulence and resistance determinants.
1494 *J Antimicrob Chemother*. 2010 Dec;65(12):2518-29.

1495 Watanabe, T., and Fukasawa, T., (1961).Episome-mediated transfer of drug resistance
1496 in Enterobacteriaceae. III. *Transduction of resistance factors*. .J Bacteriol. 1961
1497 Aug;82:202-9.

1498 Werren, J. H., (2011). Selfish genetic elements, genetic conflict, and evolutionary
1499 innovation. Proc Natl Acad Sci U S A. 2011 Jun 28; 108 (Suppl 2): 10863–10870.

1500 WHO. (2014). World Health Organisation: Antimicrobial resistance: *global report on*
1501 *surveillance 2014*. Online: available at; <
1502 http://apps.who.int/iris/bitstream/10665/112642/1/9789241564748_eng.pdf?ua=1 >.

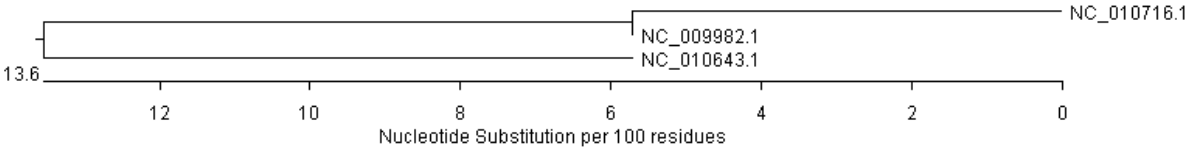
1503 Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O.,
1504 Aarestrup, F. M., and Larsen, M. V., (2012). Identification of acquired
1505 antimicrobial resistance genes. J Antimicrob Chemother. 2012 Nov;67(11):2640-
1506 4.

CHAPTER 6: SUPPLEMENTARY MATERIALS

6.1. APPENDICES

6.1.1. APPENDIX A: Phylogenetic trees (Analysis of Plasmid Incompatibility)

Appendix A1. Inc W phylogenetic tree

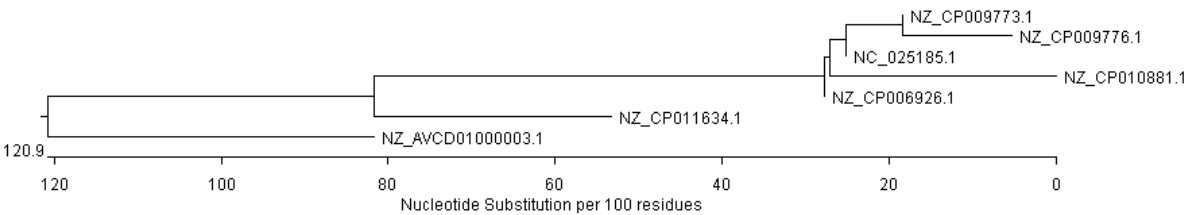


Appendix A2. Inc F(rep B) phylogenetic tree



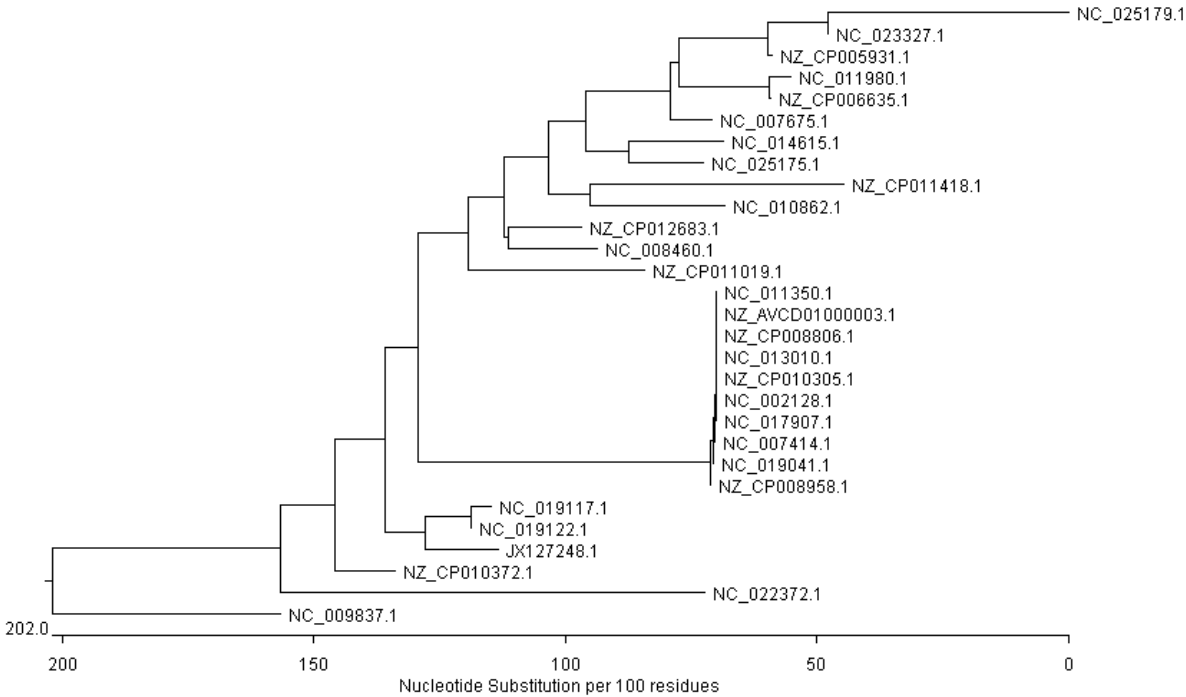
1517

1518 Appendix A3. Inc FIA-R phylogenetic tree



1519

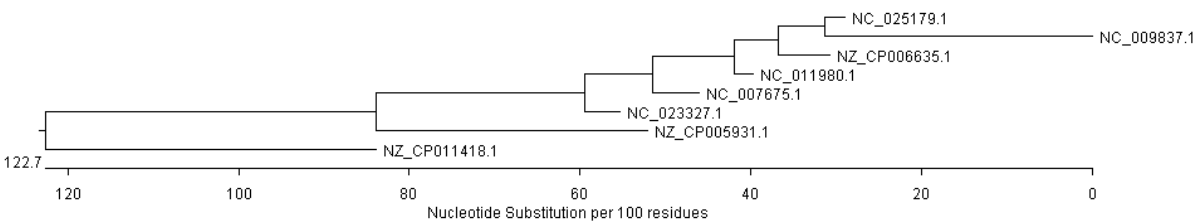
1520 Appendix A4. Inc FIB-R phylogenetic tree



1521

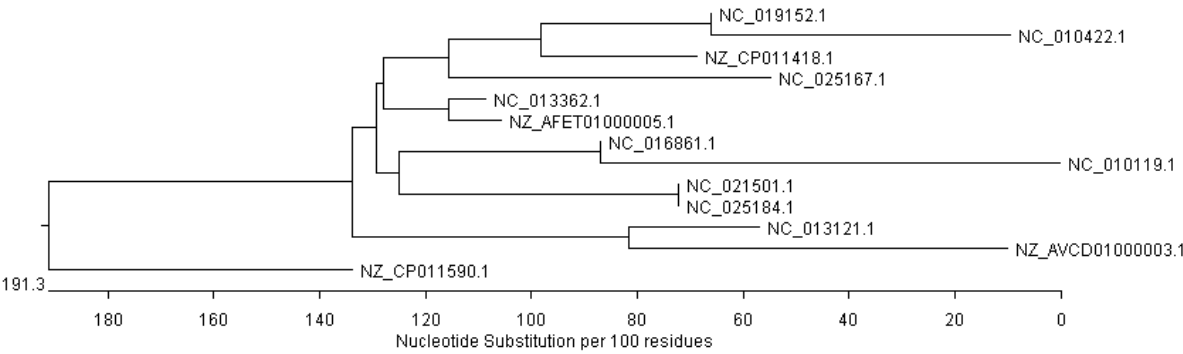
1522 Appendix A5. Inc FIC-R phylogenetic tree

1523



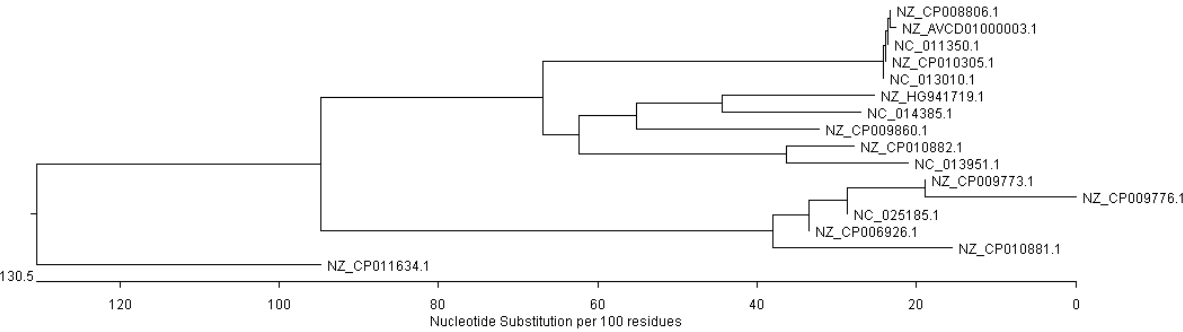
1524

1525 Appendix A6. Inc FII-R phylogenetic tree



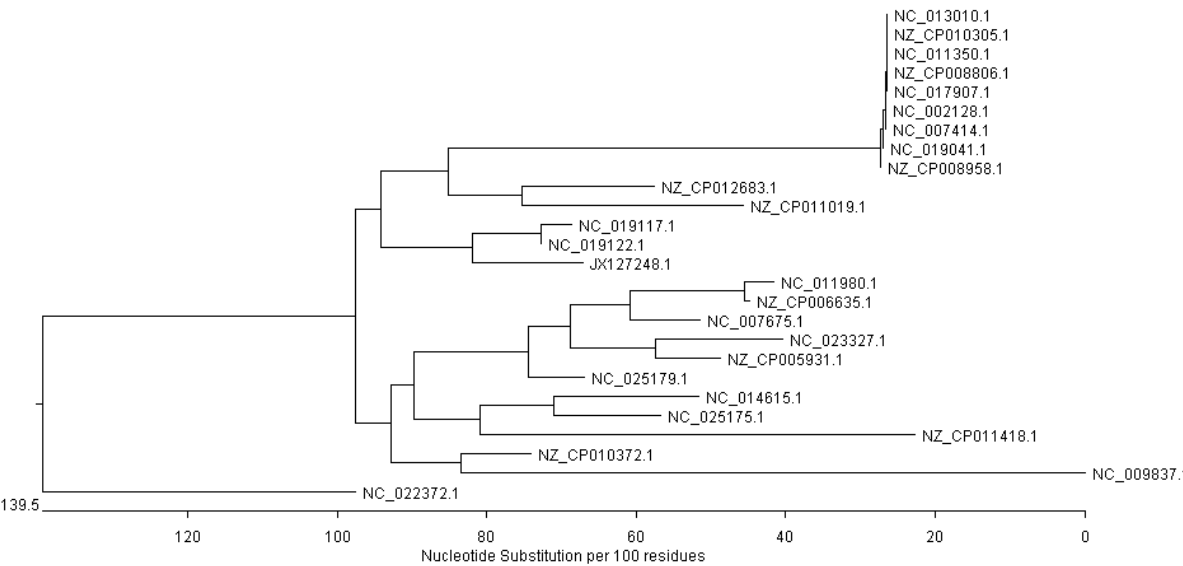
1526

1527 Appendix A7. Inc FIA2-R phylogenetic tree



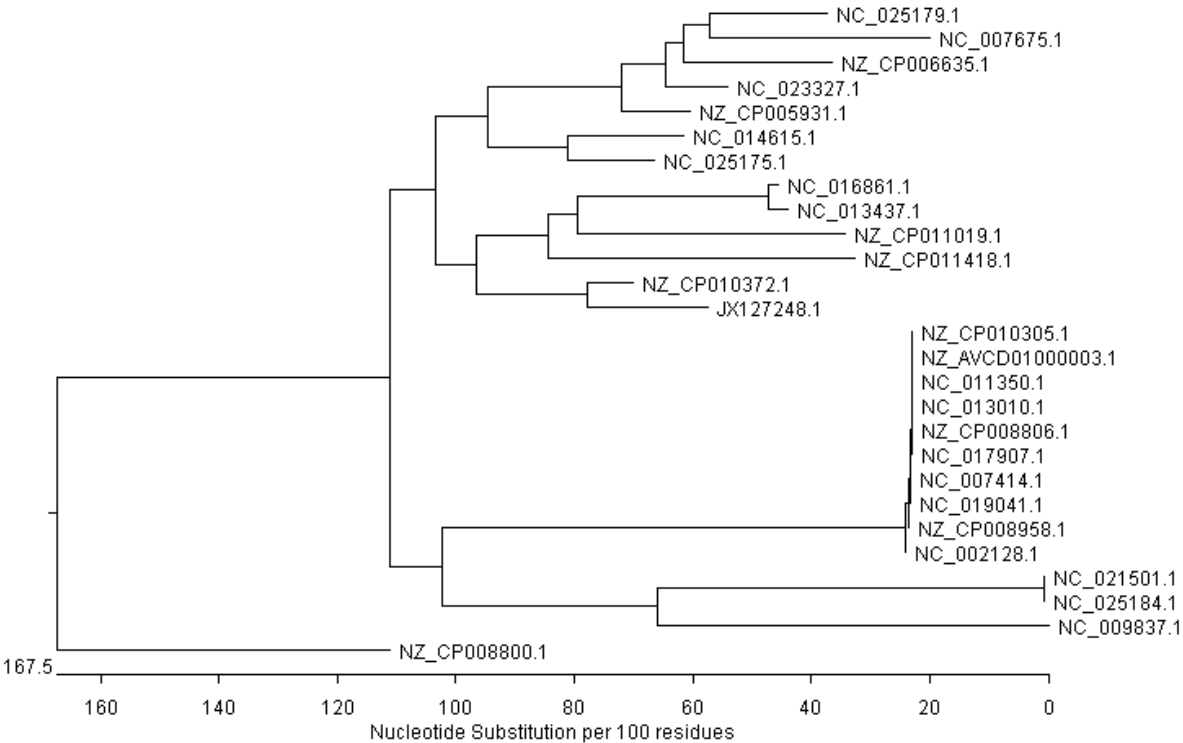
1528

1529 Appendix A8. Inc FIB2-R phylogenetic tree



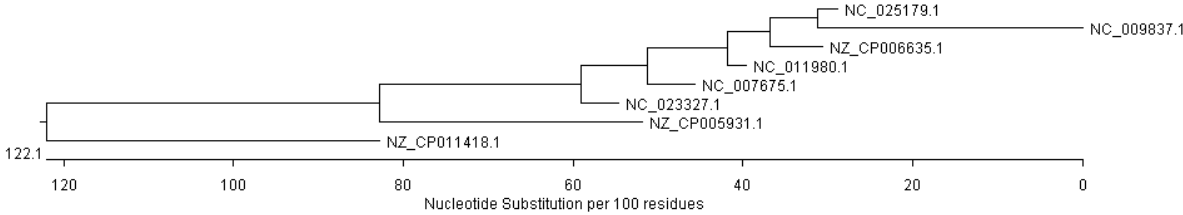
1530

1531 Appendix A9. Inc FIBs-R phylogenetic tree

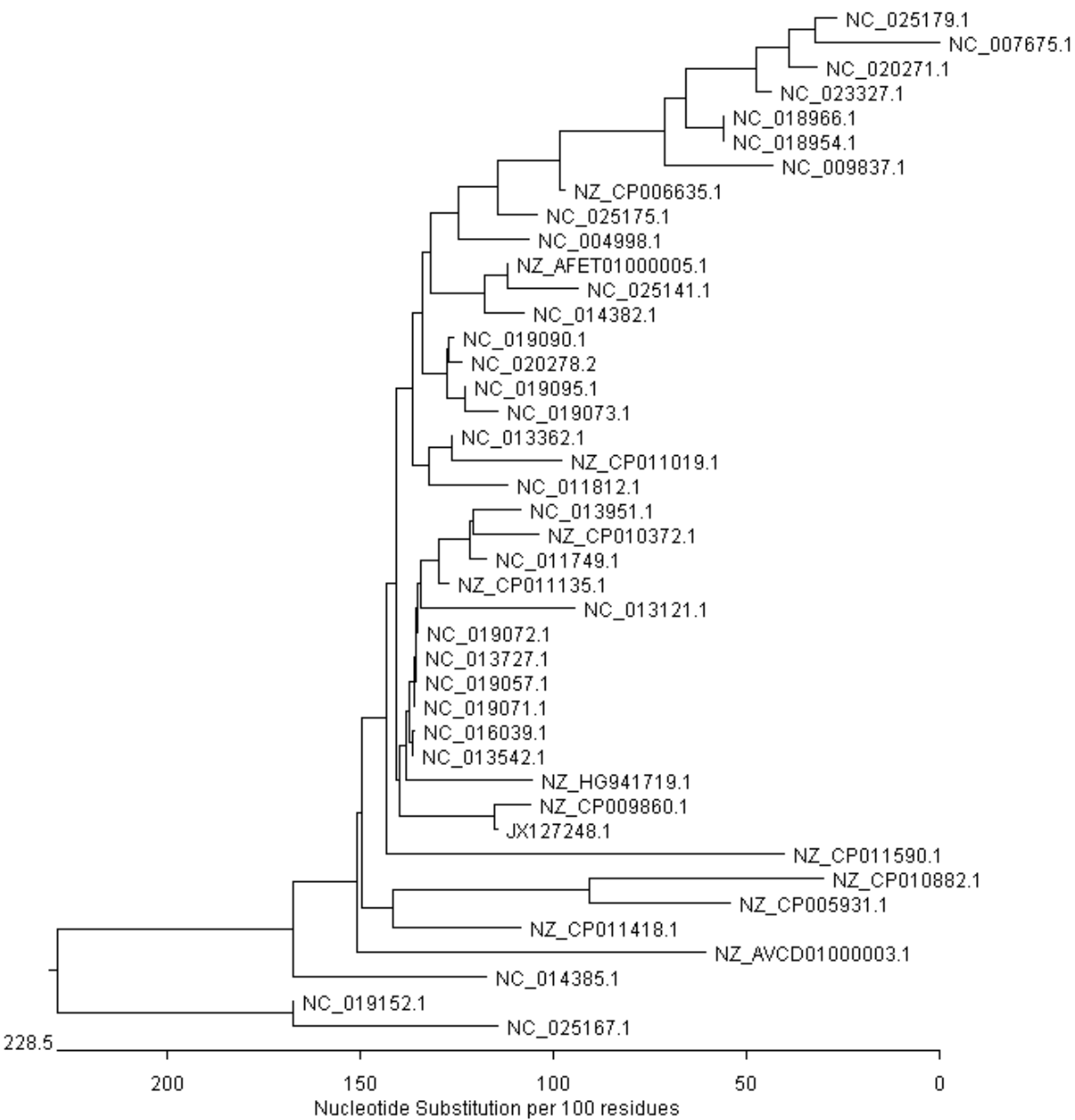


1532

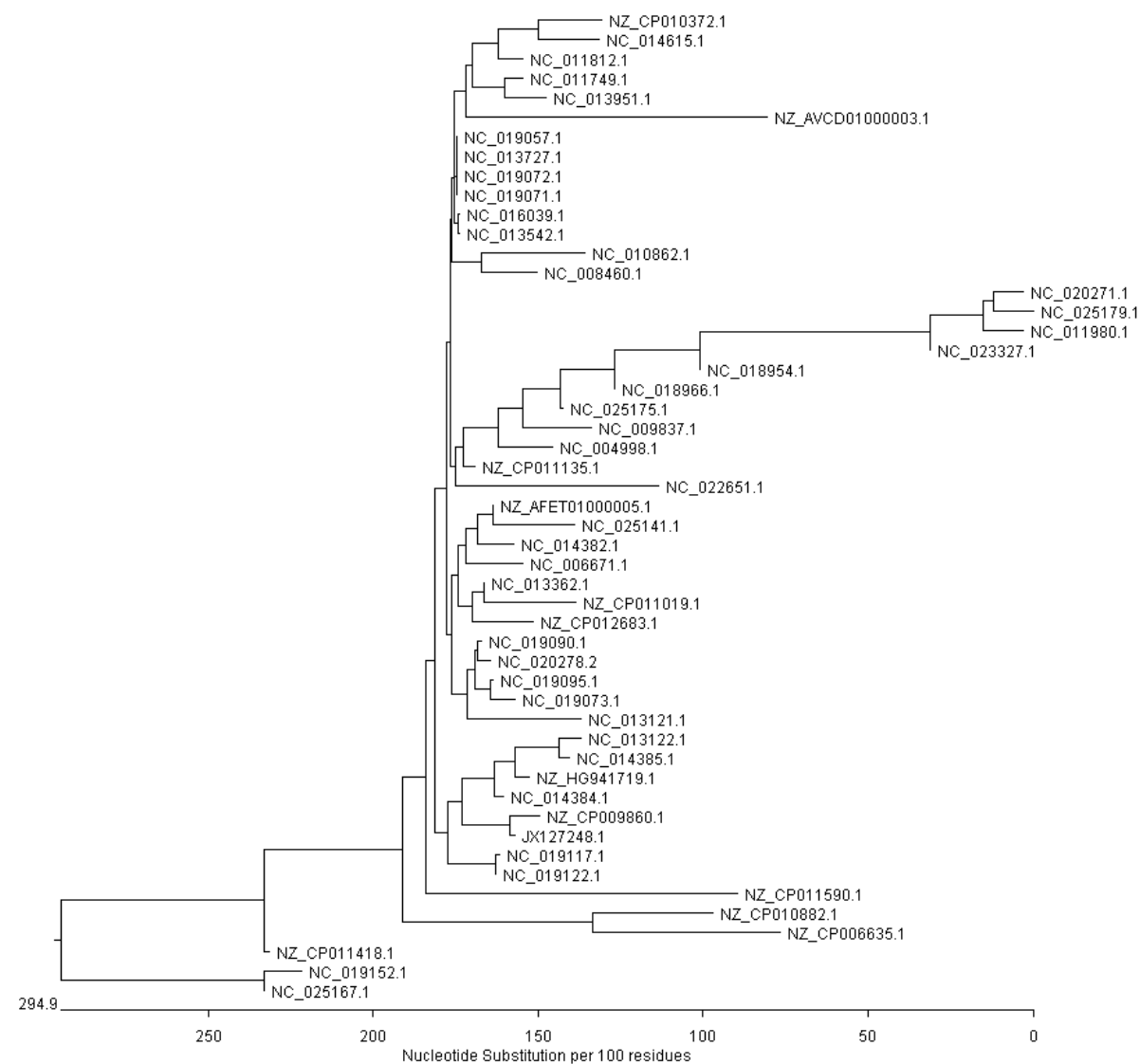
1533 Appendix A10. Inc FIC2-R phylogenetic tree



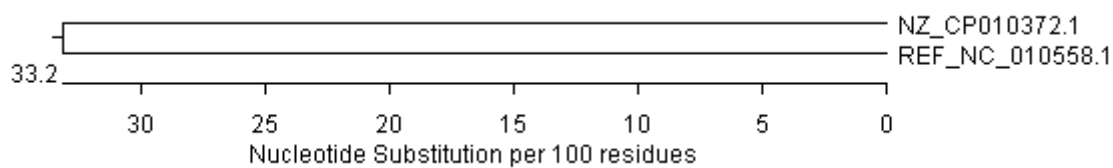
1534



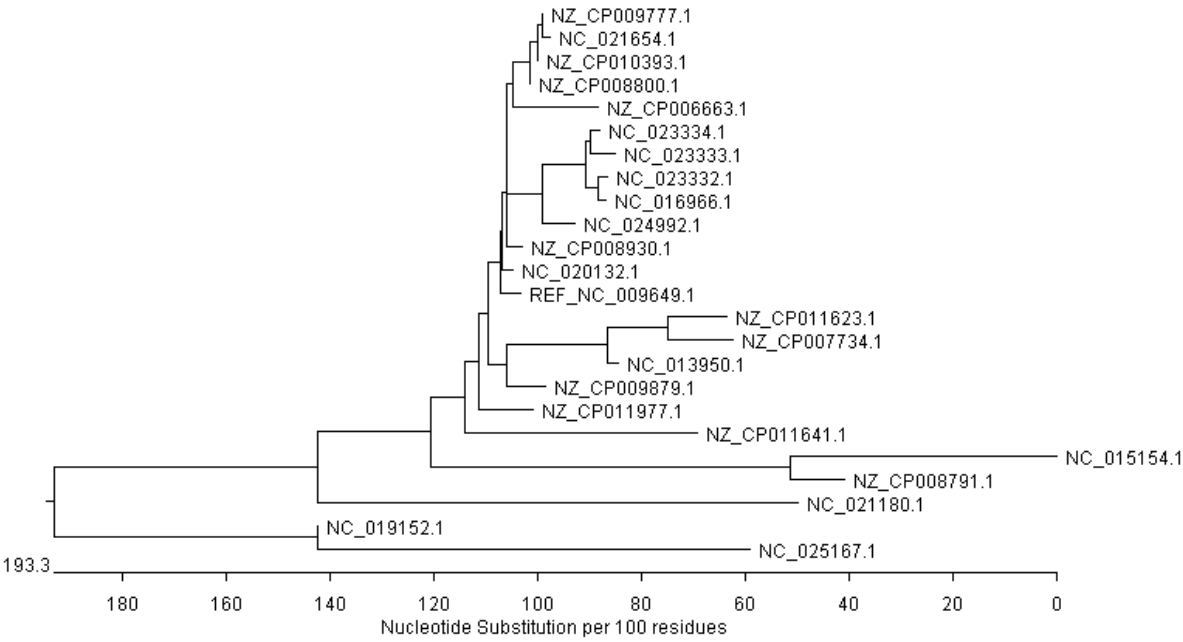
1538



1540

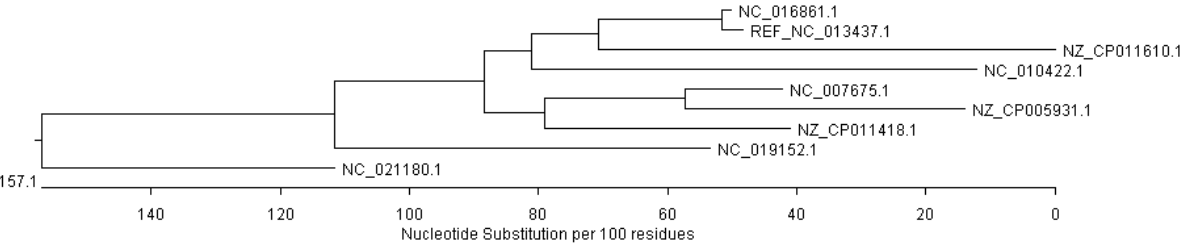


1541 Appendix A14. Inc FIIk phylogenetic tree



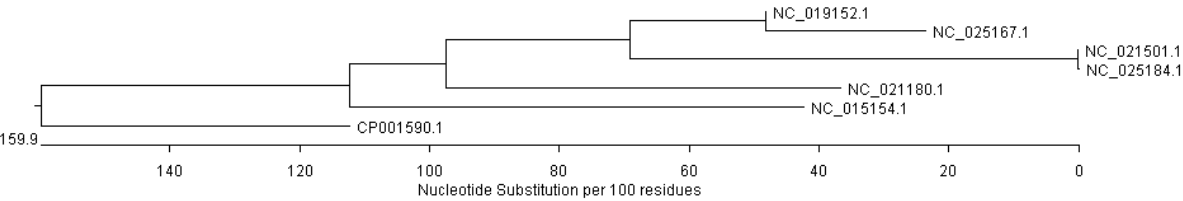
1542

1543 Appendix A15. Inc FIIs phylogenetic tree



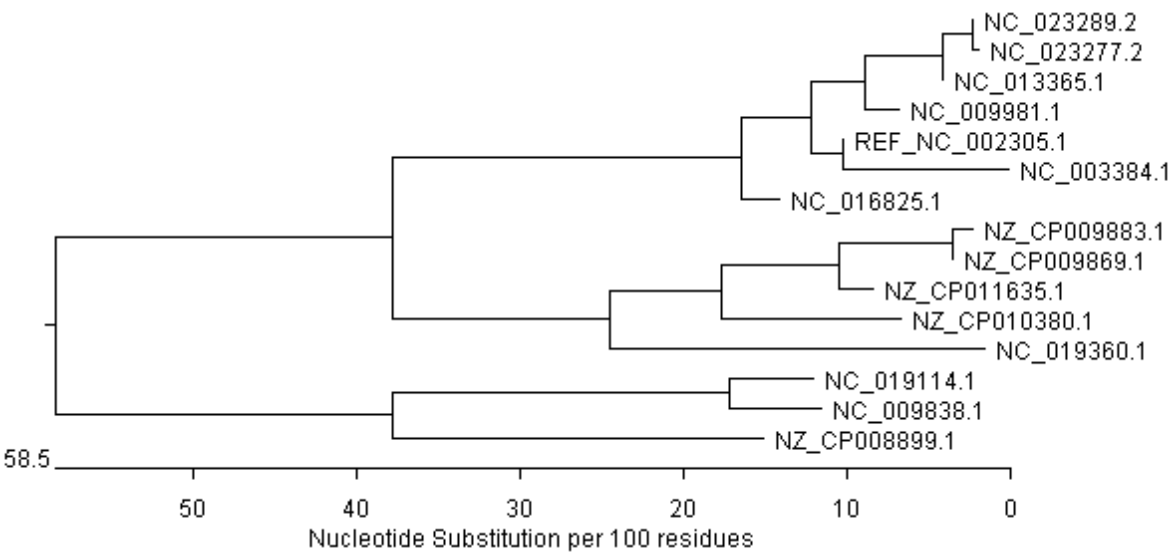
1544

1545 Appendix A16. Inc FIly phylogenetic tree



1546

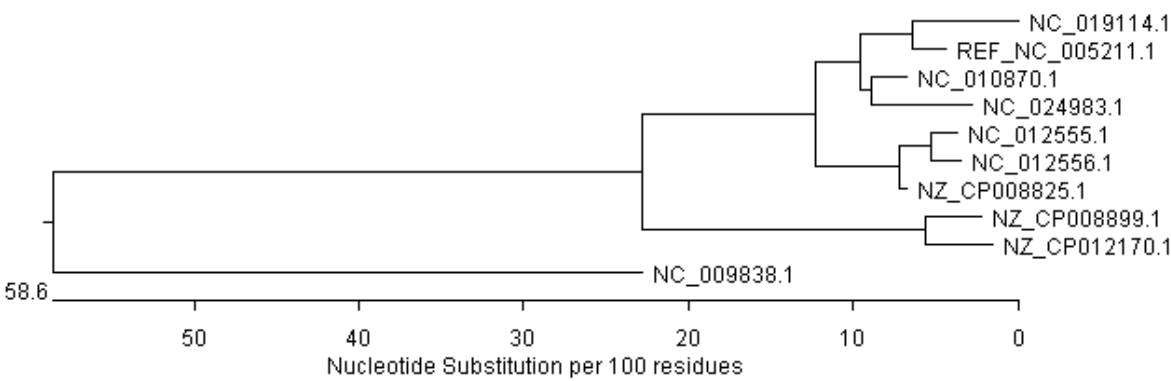
1547 Appendix A17. Inc HI1 phylogenetic tree



1548

1549

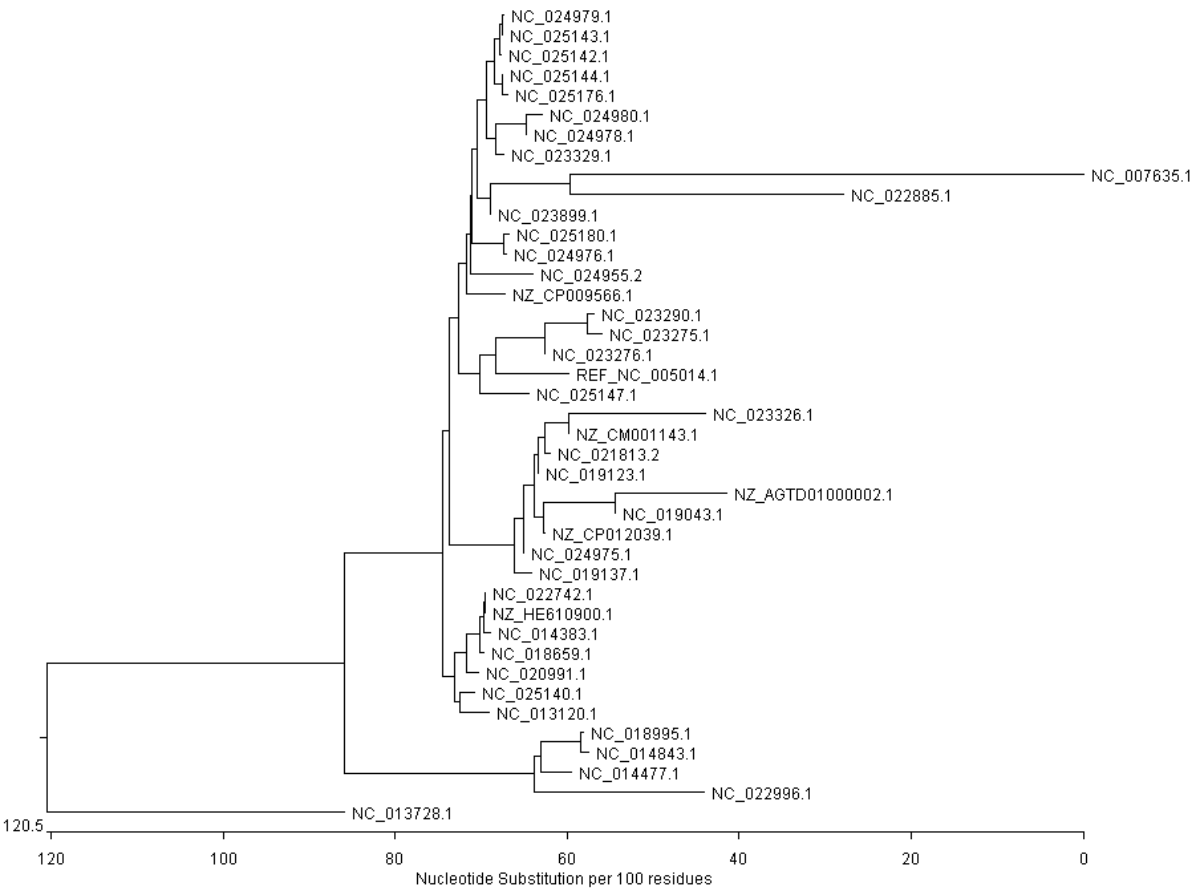
1550 Appendix A18. Inc HI2 phylogenetic tree



1551

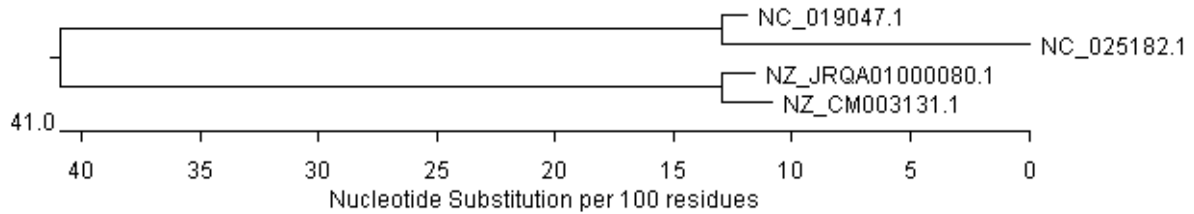
1552

1553 Appendix A19. Inc I1 phylogenetic tree



1554

1555 Appendix A20. Inc X phylogenetic tree



1556

1557

1558

1559

1560

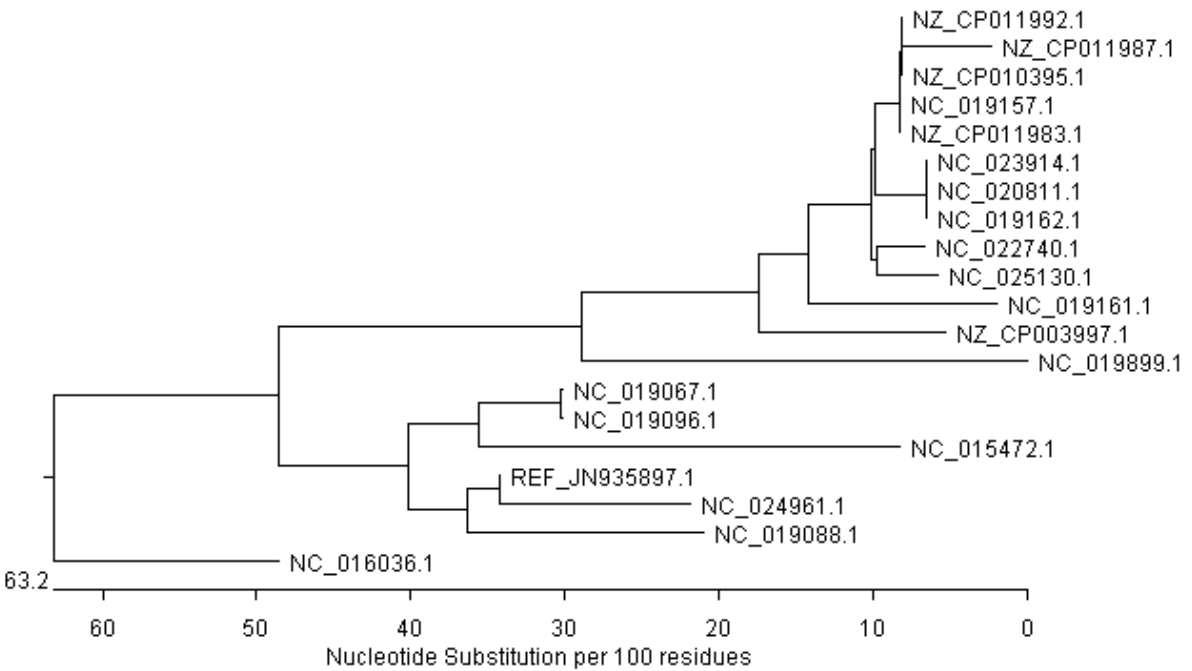
1561

1562

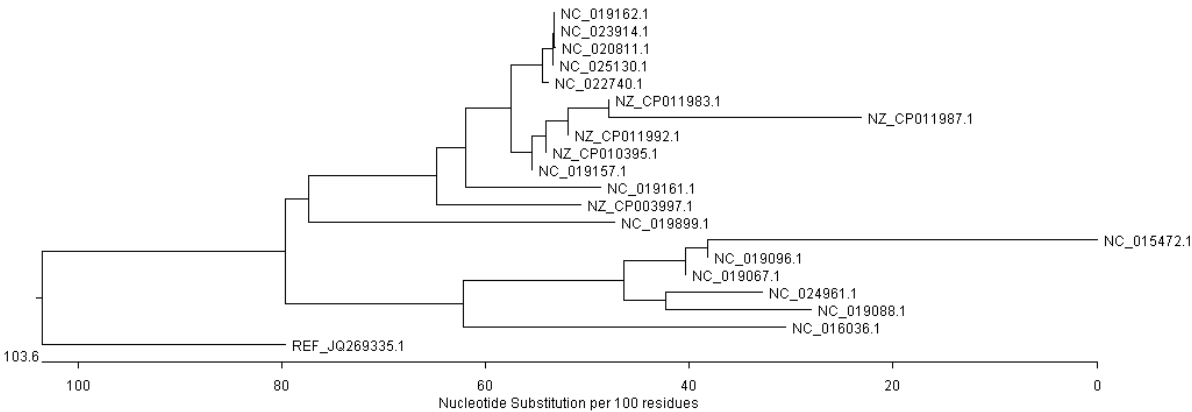
1563

1564

1565 Appendix A21. Inc X1 phylogenetic tree
1566

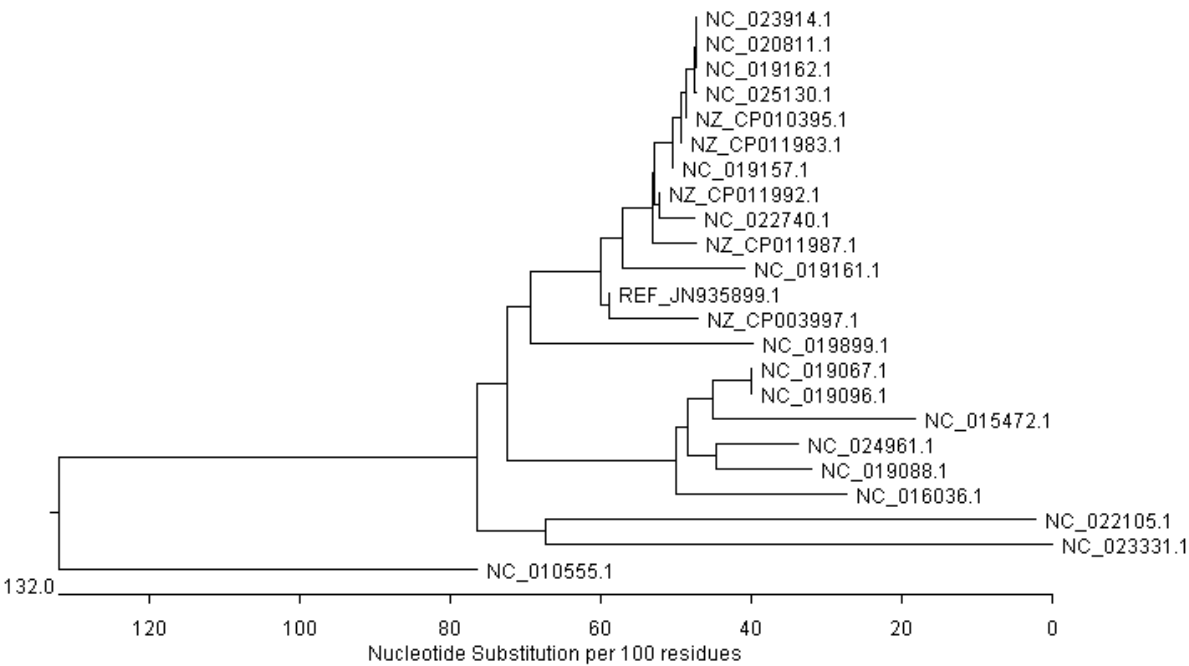


1567
1568 Appendix A22. Inc X2 phylogenetic tree
1569



1570
1571

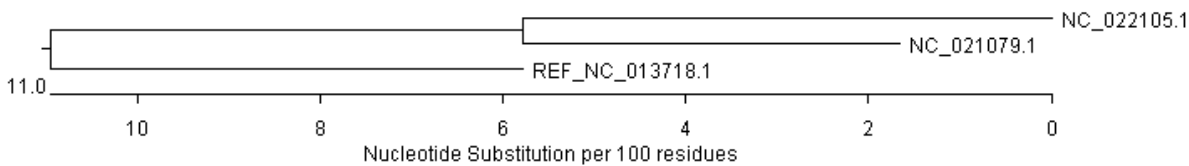
1572 Appendix A23. Inc X3 phylogenetic tree



1573

1574

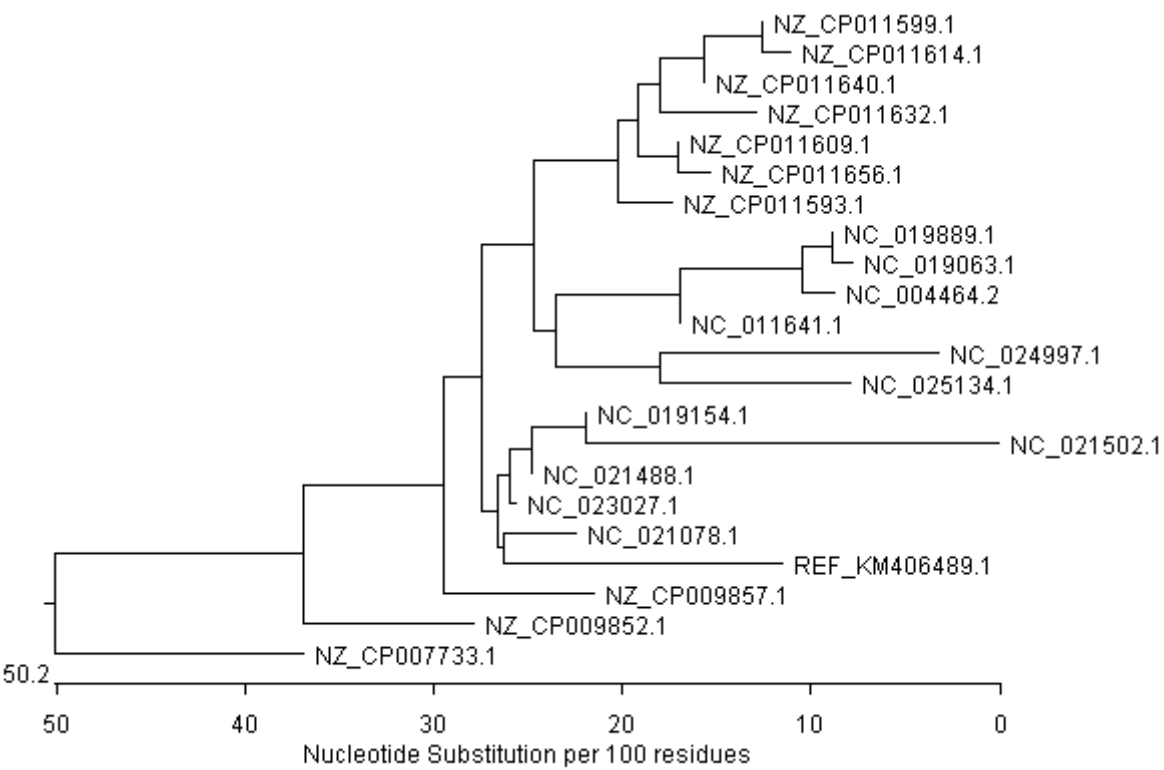
1575 Appendix A24. Inc X4 phylogenetic tree



1576

1577

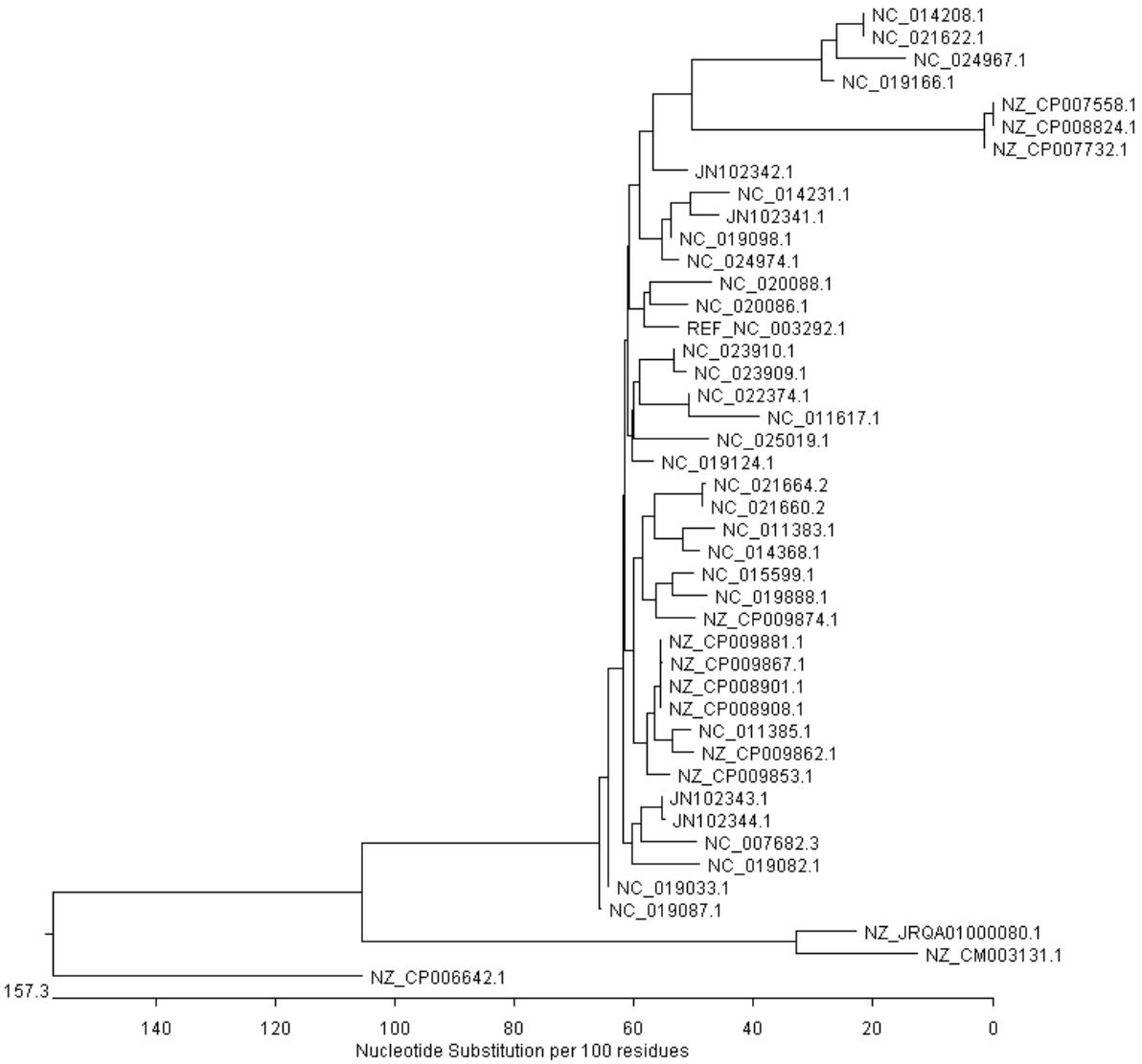
1578 Appendix A25. Inc L & M phylogenetic tree



1579

1580

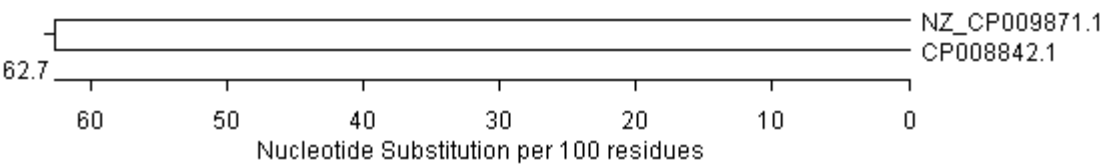
1581 Appendix A26. Inc N phylogenetic tree



1582

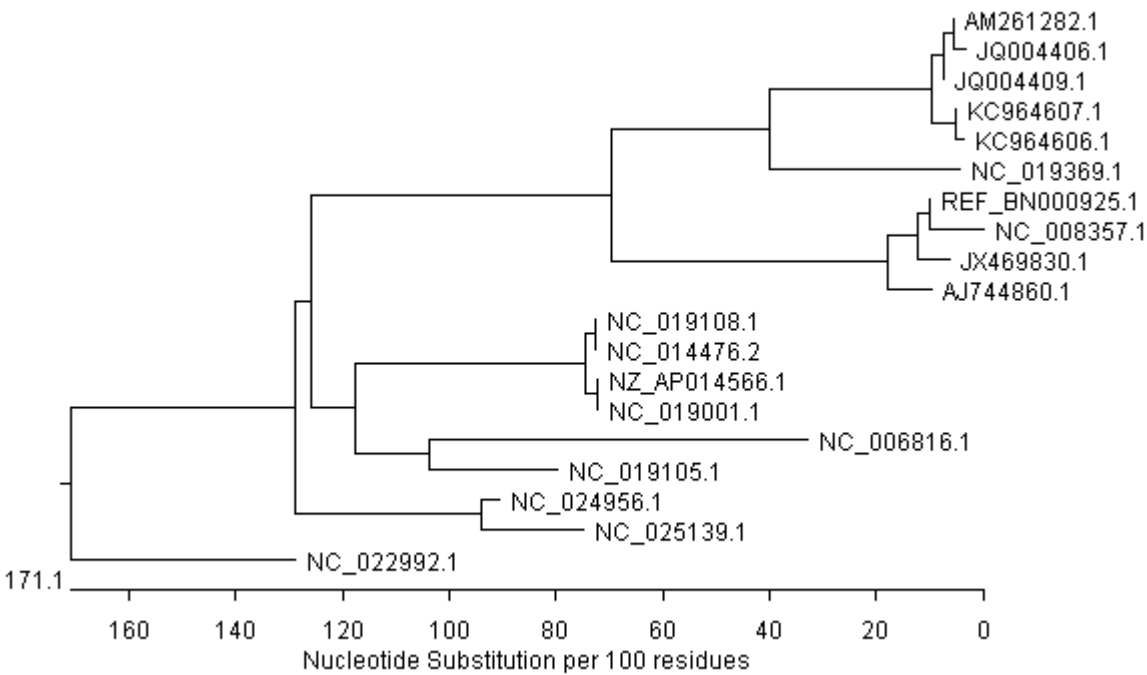
1583

1584 Appendix A27. Inc Y phylogenetic tree



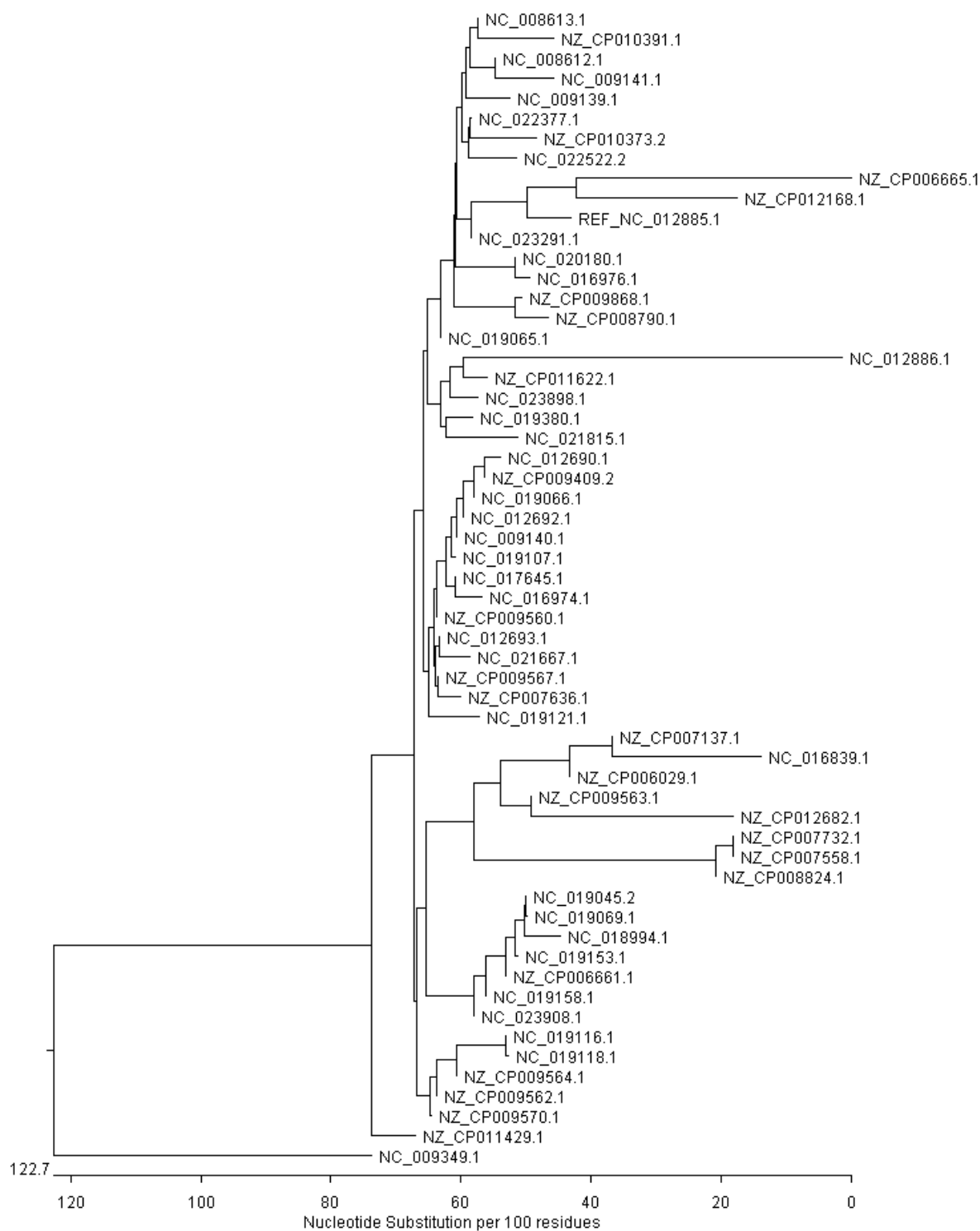
1585

1586 Appendix A28. Inc P phylogenetic tree

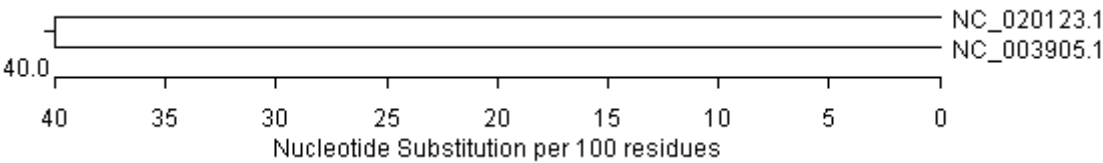


1587

1588 Appendix A29. Inc A &C phylogenetic tree

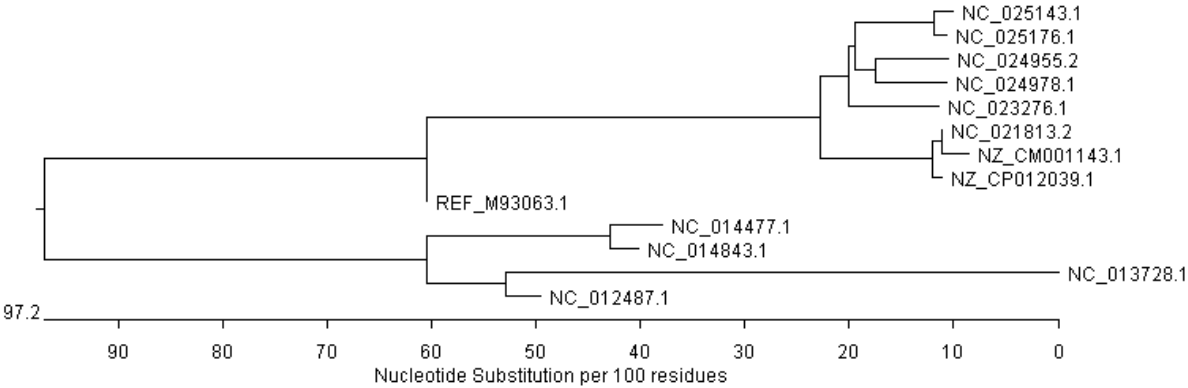


1591 Appendix A30. Inc T phylogenetic tree



1592

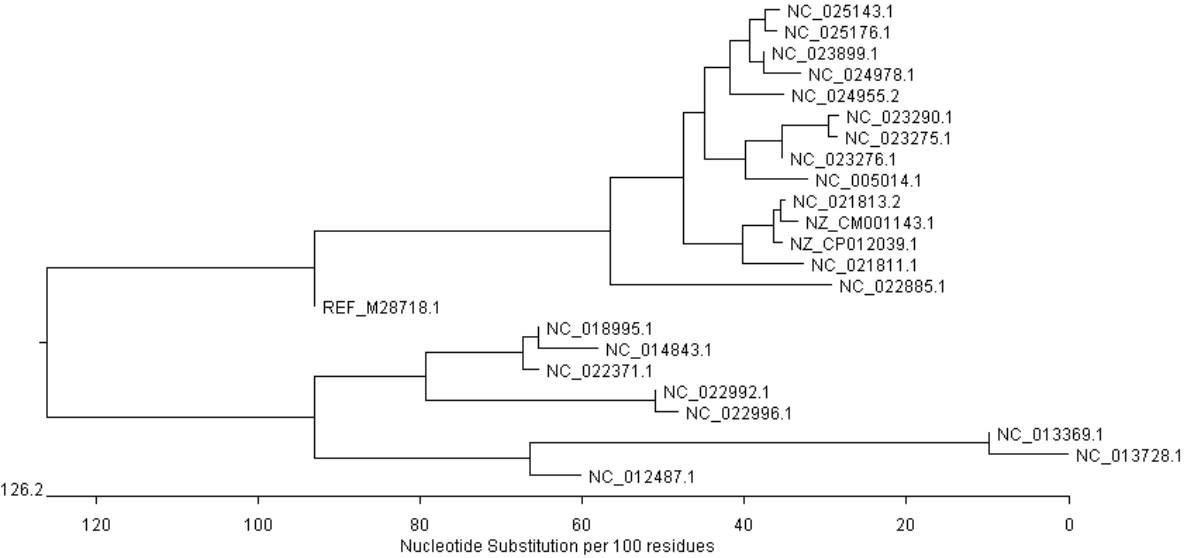
1593 Appendix A31. Inc K & B phylogenetic tree



1594

1595

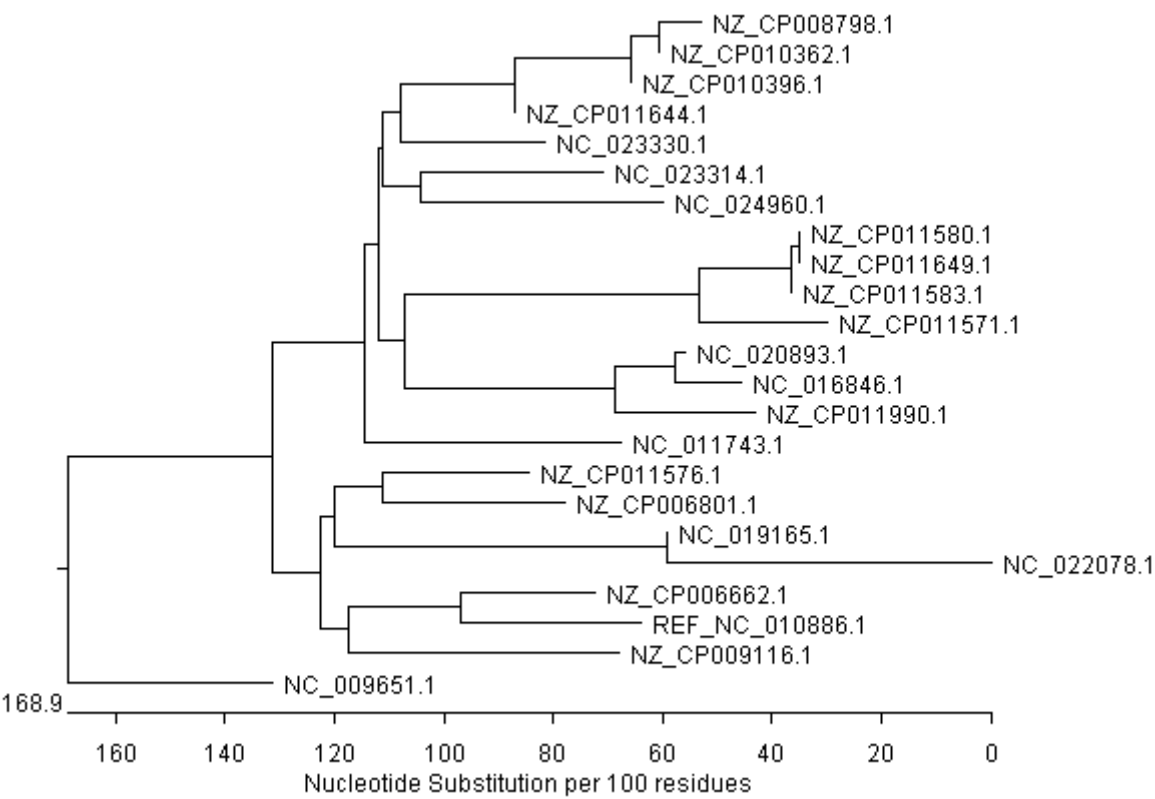
1596 Appendix A32. Inc B & O phylogenetic tree



1597

1598

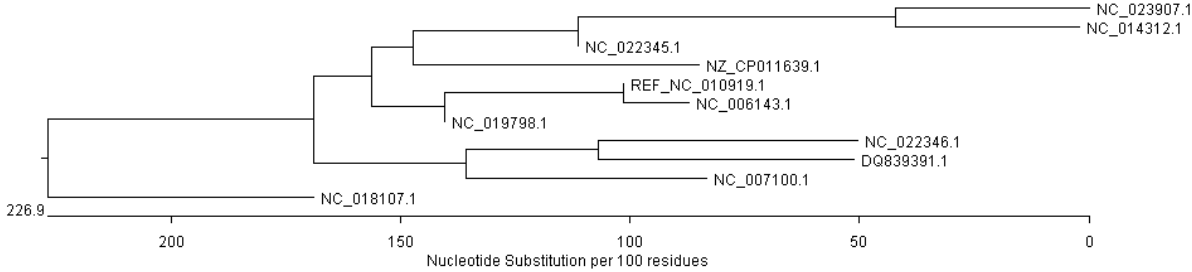
1599 Appendix A33. Inc R phylogenetic tree



1600

1601

1602 Appendix A34. Inc U phylogenetic tree

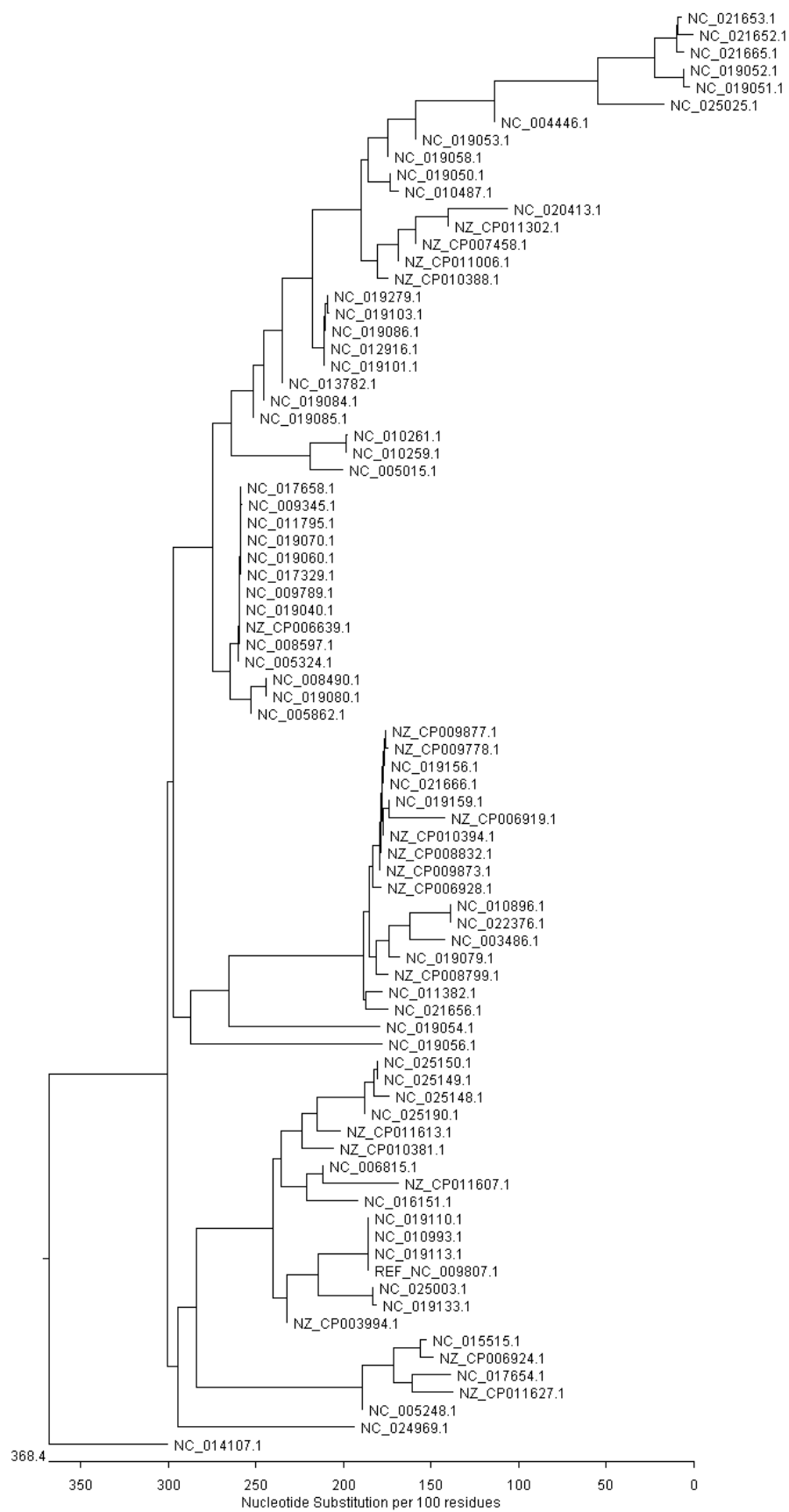


1603

1604

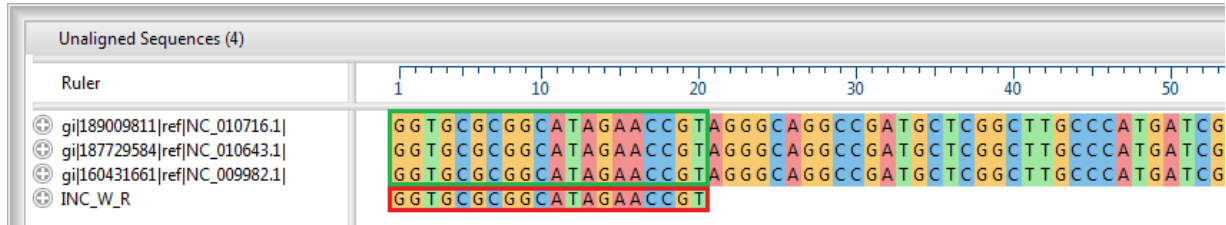
1605 Appendix A35. ColE1 phylogenetic tree

1606

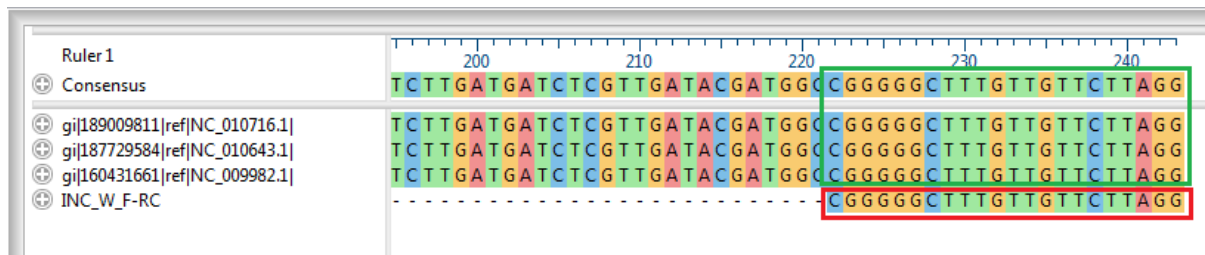


6.1.2. APPENDIX B: Primer MSA against target sequences (Primer diagnosis)

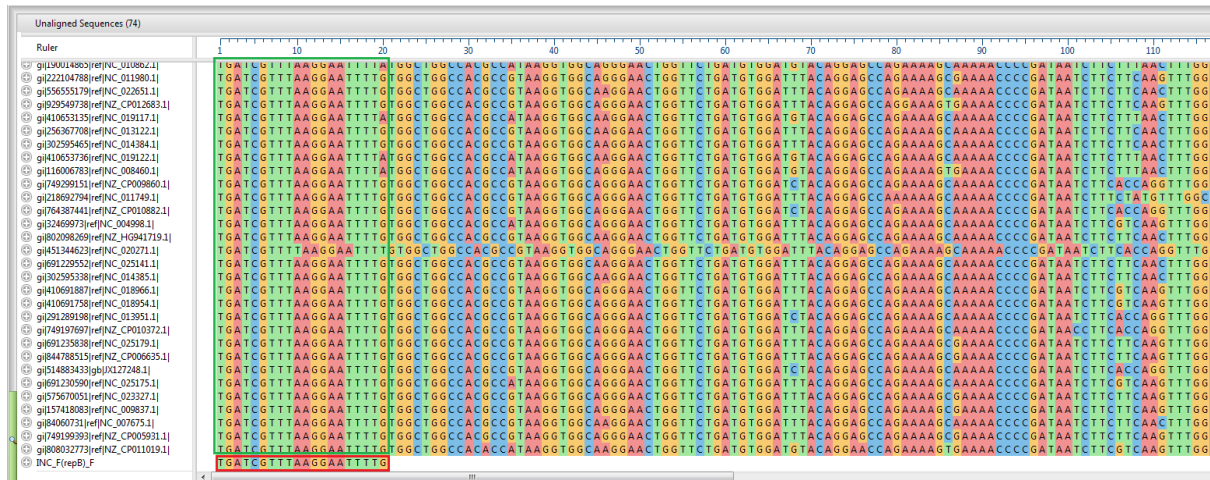
Appendix B1. Inc W^a Forward primer MSA against 3 amplicons



Appendix B2. Inc W^a Reverse primer MSA against 3 amplicons



Appendix B3. Inc F(rep B)^a Forward primer MSA against 74 amplicons



1622 Appendix B4.Inc F(rep B)^a Reverse primer MSA against 74 amplicons

1623

1624

1625 Appendix B5. Inc FIA^a Forward primer MSA against 7 amplicons

1626

1627 Appendix B6. Inc FIA^a Reverse primer MSA against 7 amplicons

1628

1629

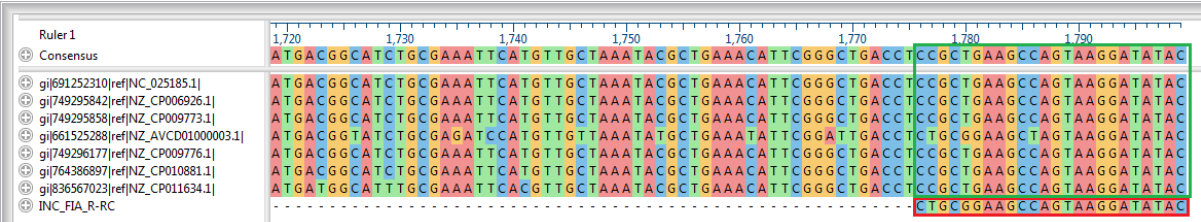
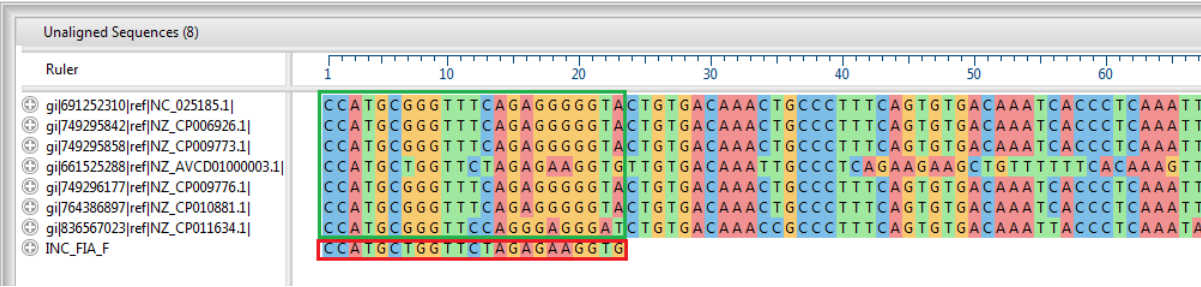
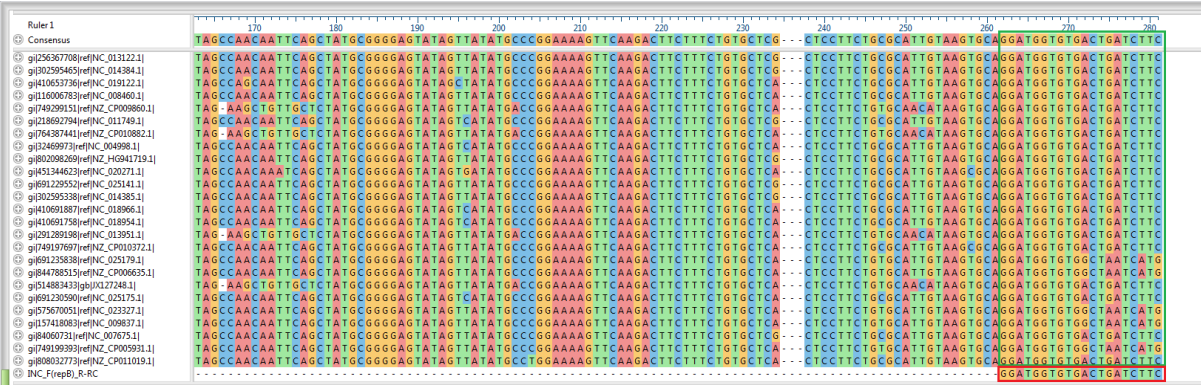
1630

1631

1632

1633

1634

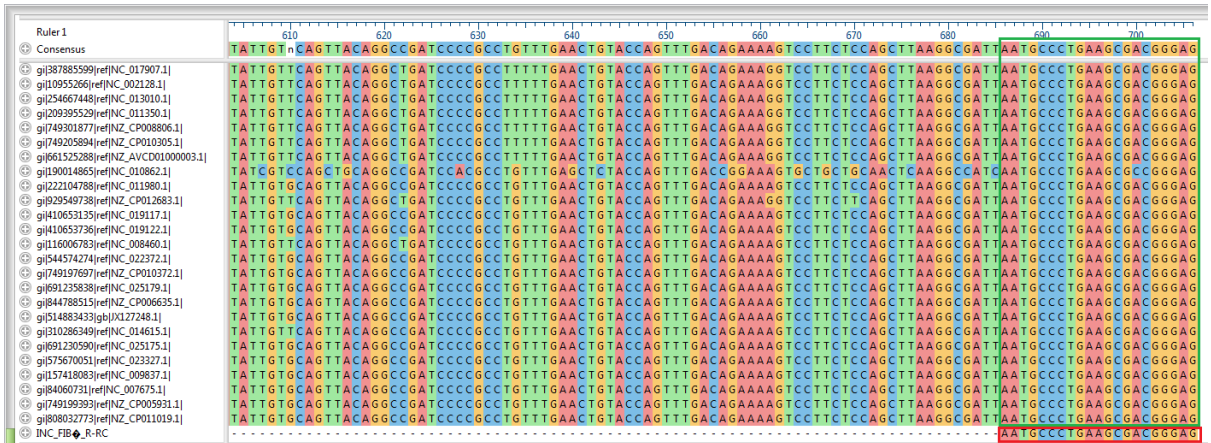


1635 Appendix B7. Inc FIB^a Forward primer MSA against 29 amplicons



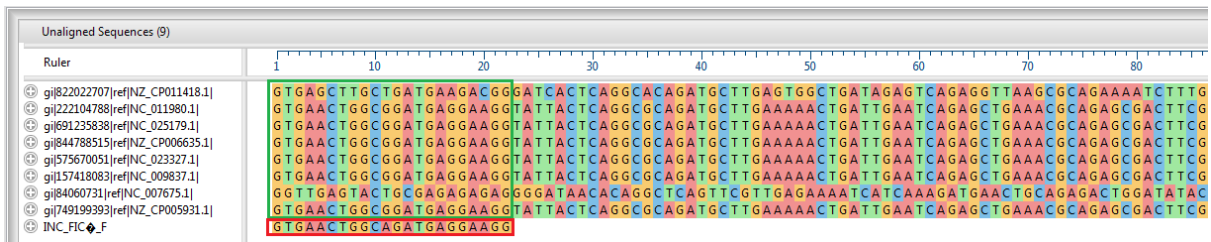
1636

1637 Appendix B8. Inc FIB^a Reverse primer MSA against 29 amplicons



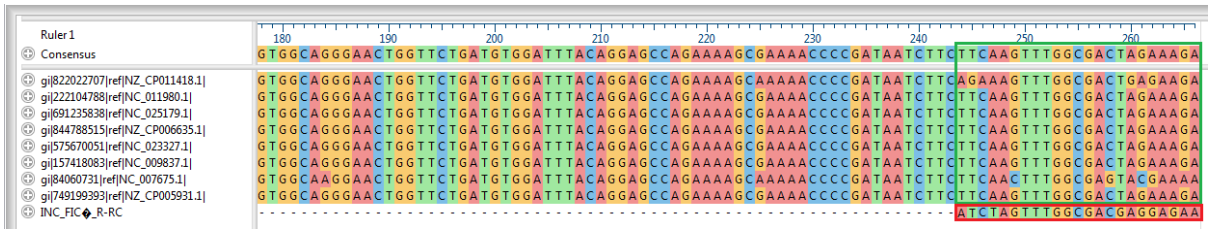
1638

1639 Appendix B9. Inc FIC^a Forward primer MSA against 8 amplicons



1640

1641 Appendix B10. Inc FIC^a Reverse primer MSA against 8 amplicons



1643

1644 Appendix B11. Inc FII^a Forward primer MSA against 13 amplicons

1645



1646 Appendix B12. Inc FII^a Reverse primer MSA against 13 amplicons

1647



1648

1649 Appendix B13. Inc FIA^b Forward primer MSA against 17 amplicons

1650



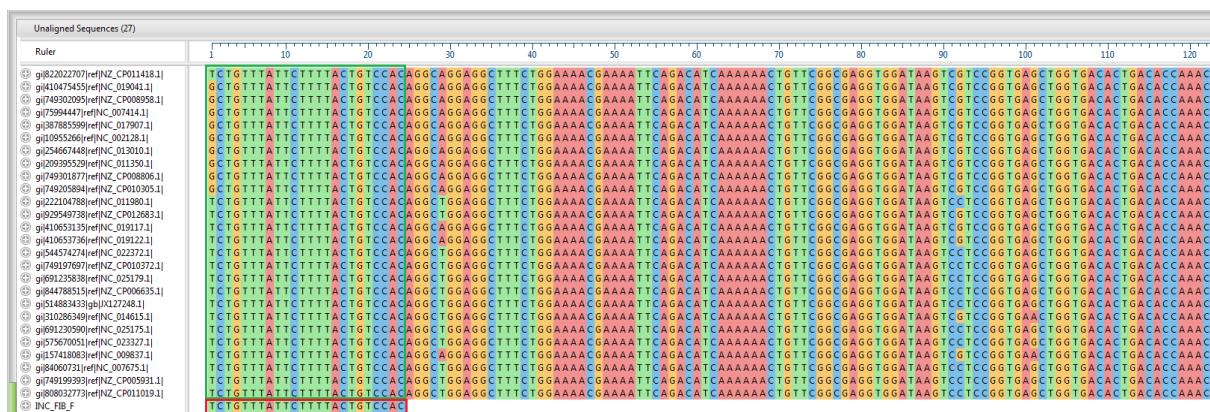
1651 Appendix B14. Inc FIA^b Reverse primer MSA against 17 amplicons

1652



1653

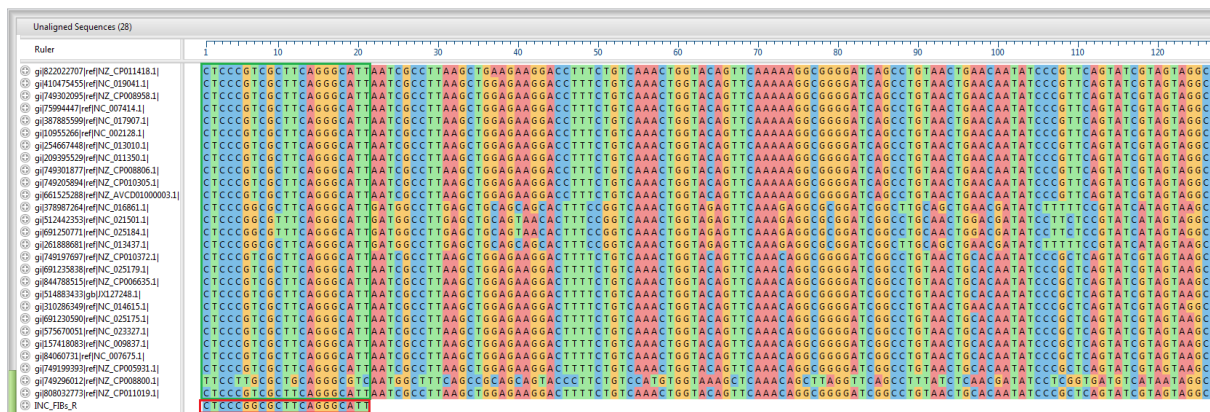
1654 Appendix B15. Inc FIB^b Forward primer MSA against 26 amplicons



1656 Appendix B16. Inc FIB^b Reverse primer MSA against 26 amplicons



1659 Appendix B17. Inc FIBs^b Forward primer MSA against 27 amplicons



1664 Appendix B18. Inc FIBs^b Reverse primer MSA against 27 amplicons



1666 Appendix B19. Inc FIC^b Forward primer MSA against 8 amplicons



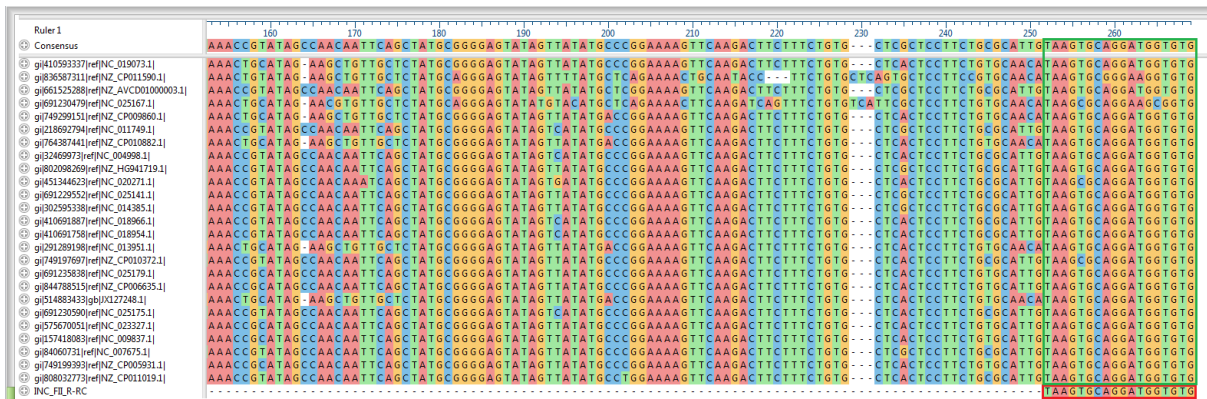
1668 Appendix B20. Inc FIC^b Reverse primer MSA against 8 amplicons



1670 Appendix B21. Inc FII^b Forward primer MSA against 42 amplicons



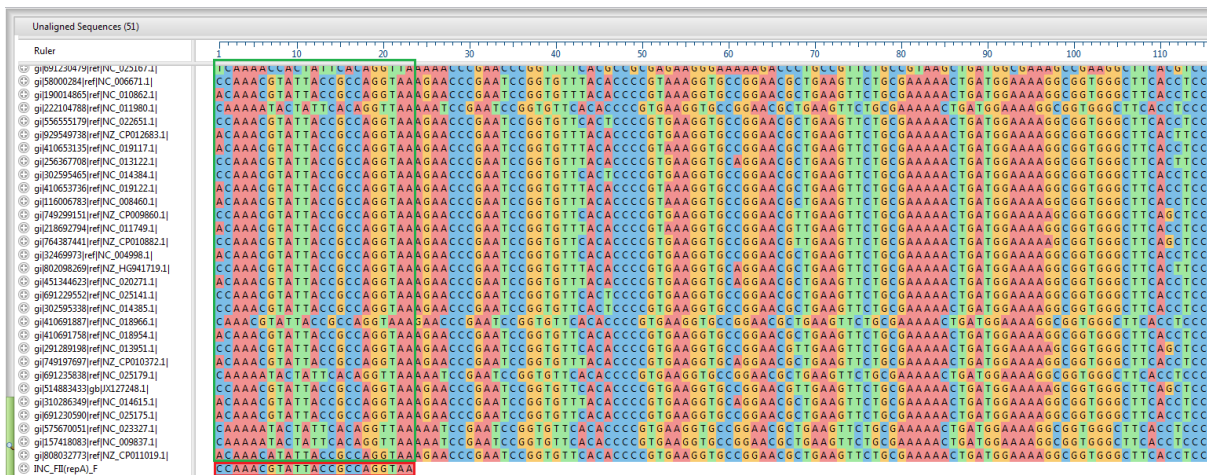
1672 Appendix B22. Inc FII^b Reverse primer MSA against 42 amplicons



1673

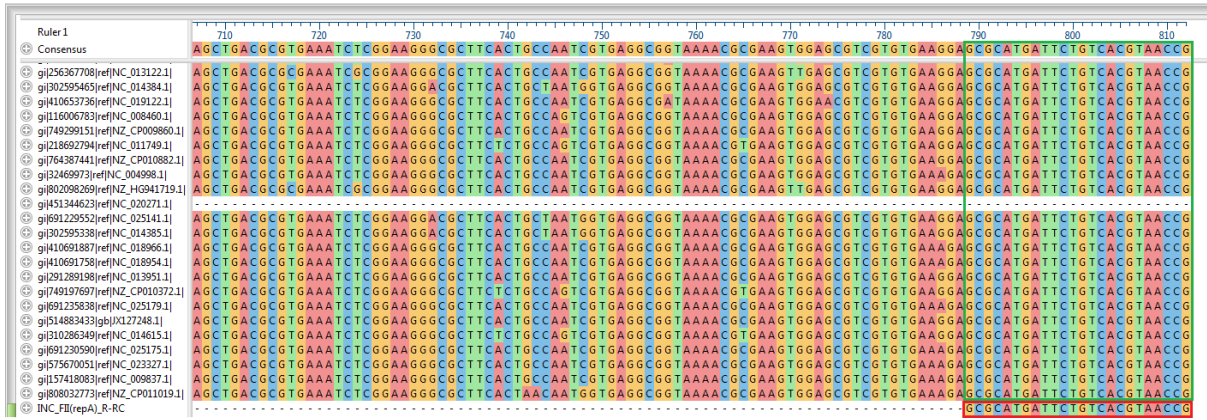
1674

1675 Appendix B23. Inc FII(rep A)^b Forward primer MSA against 50 amplicons



1676

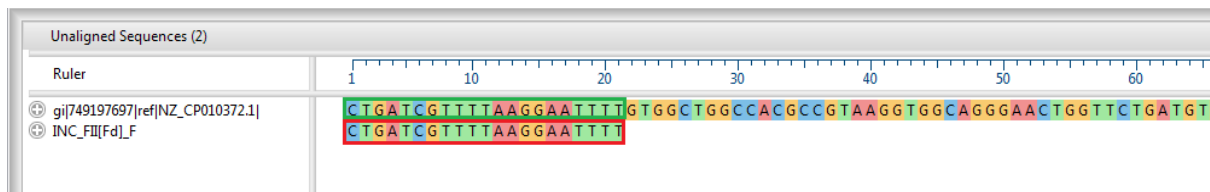
1677 Appendix B24. Inc FII(repA)^b Reverse primer MSA against 50 amplicons



1678

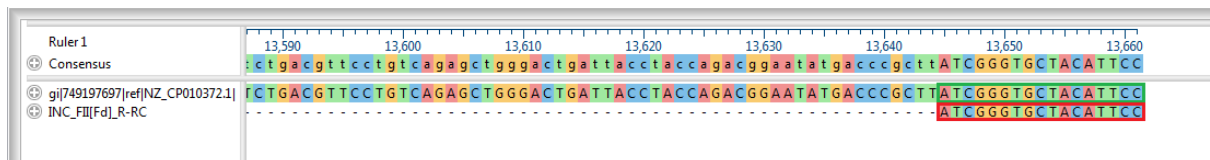
1679

1680 Appendix B25. Inc FII[Fd]^b Forward primer MSA against 1 amplicon



1681

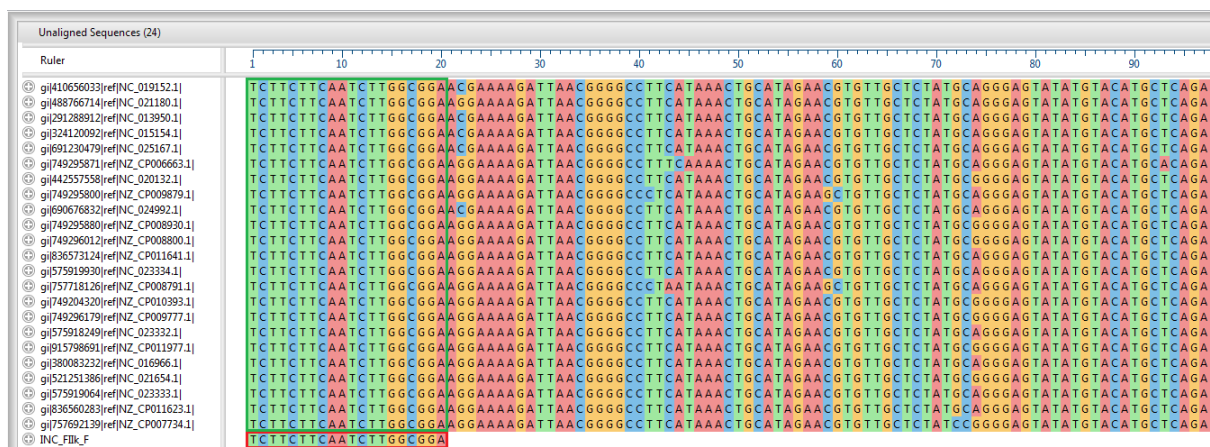
1682 Appendix B26. Inc FII[Fd]^b Reverse primer MSA against 1 amplicon



1683

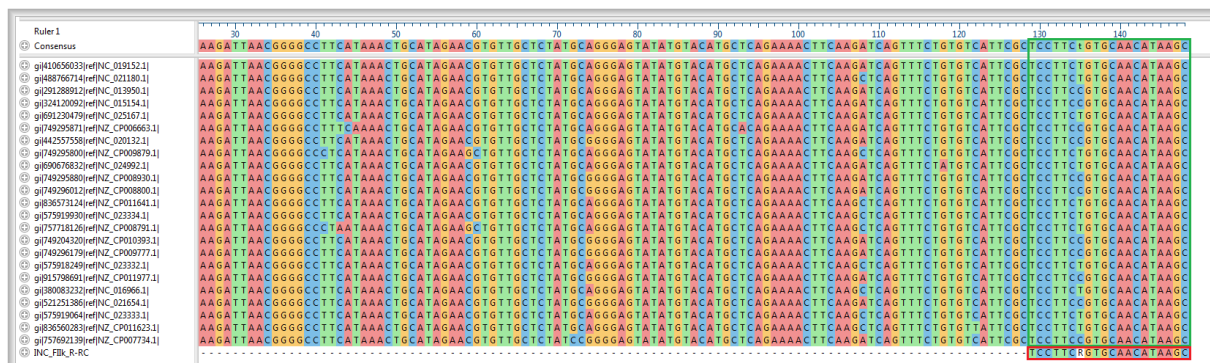
1684

1685 Appendix B27. Inc FIik^b Forward primer MSA against 23 amplicons



1686

1687 Appendix B28. Inc FIik^b Reverse primer MSA against 23 amplicons

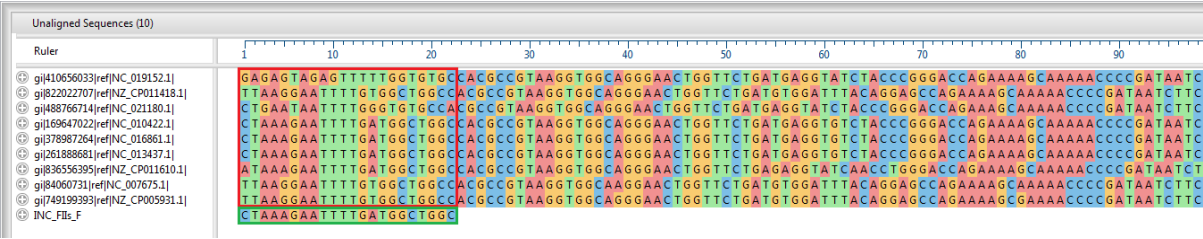


1688

1689

1690

1691 Appendix B29. Inc FIIIs^b Forward primer MSA against 9 amplicons



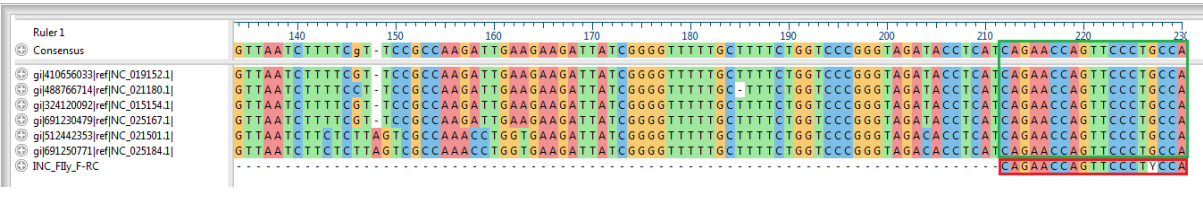
1692
1693 Appendix B30. Inc FIIIs^b Reverse primer MSA against 9 amplicons



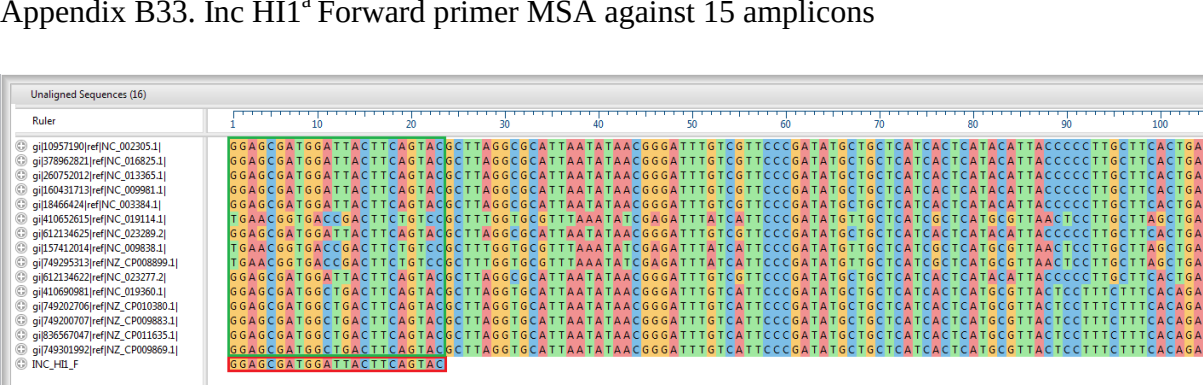
1694
1695 Appendix B31. Inc FIIy^b Forward primer MSA against 6 amplicons



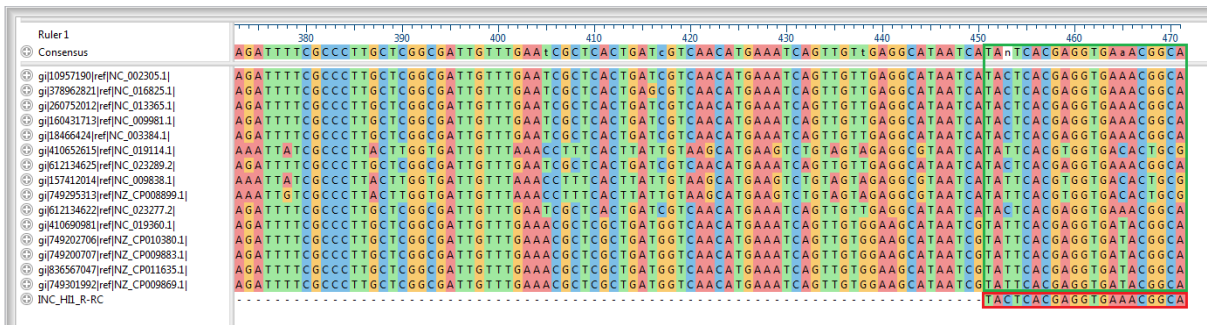
1696
1697 Appendix B32. Inc FIIy^b Reverse primer MSA against 6 amplicons



1698
1699
1700 Appendix B33. Inc HI1^a Forward primer MSA against 15 amplicons



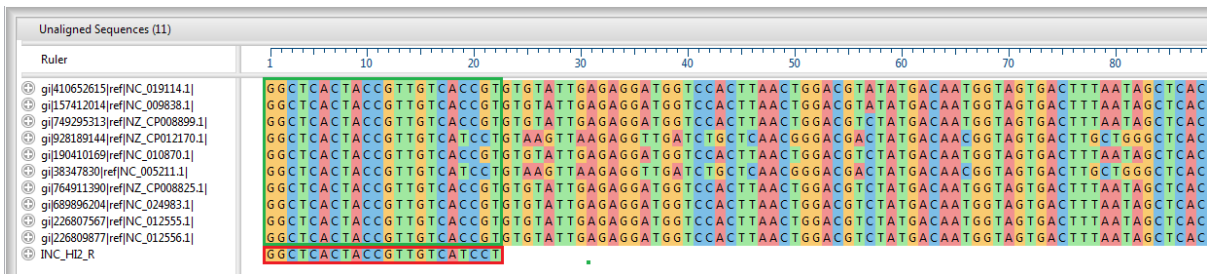
1702 Appendix B34. Inc HI1^a Reverse primer MSA against 15 amplicons



1703

1704

1705 Appendix B35. Inc HI2^a Forward primer MSA against 10 amplicons



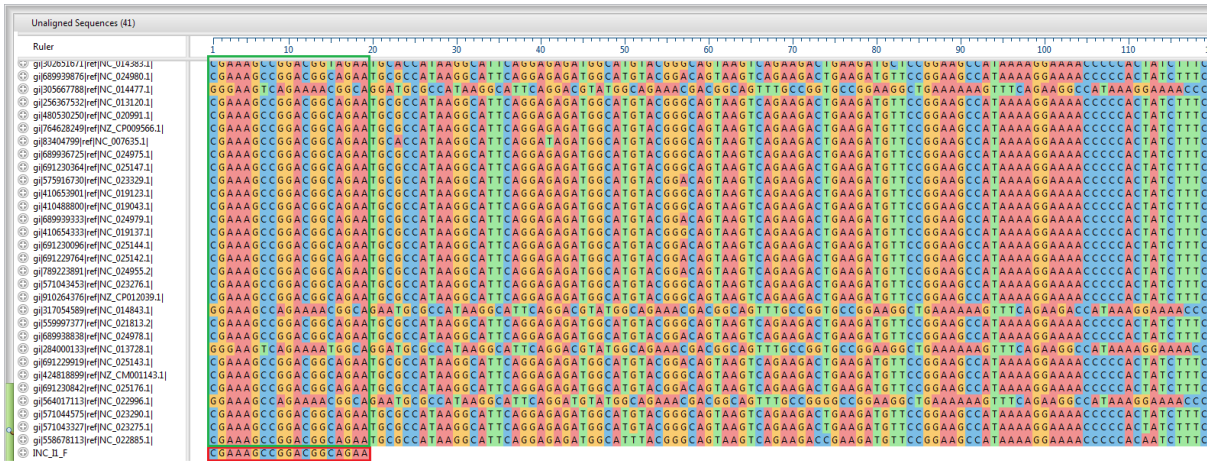
1706

1707 Appendix B36. Inc HI2^a Reverse primer MSA against 10 amplicons



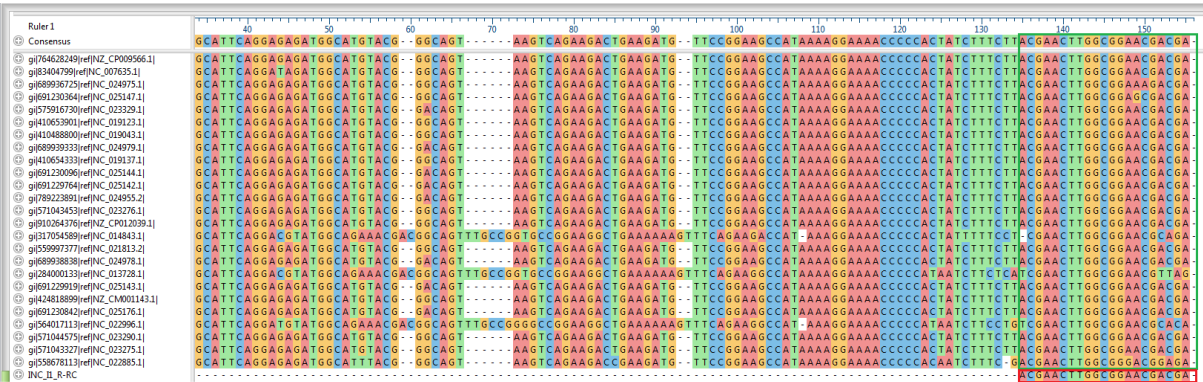
1708

1709 Appendix B37. Inc I1^a Forward primer MSA against 40 amplicons



1710

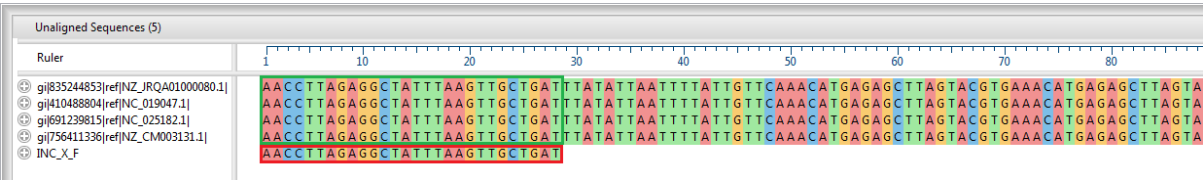
1711 Appendix B38. Inc I1^a Reverse primer MSA against 40 amplicons



1712

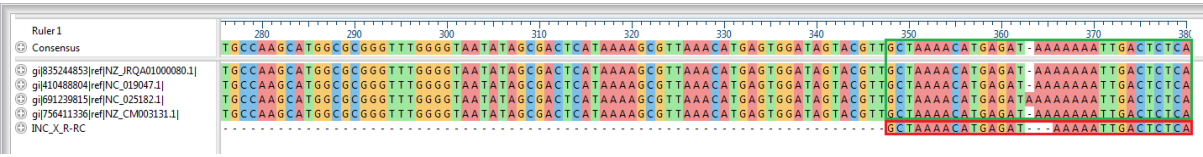
1713

1714 Appendix B39. Inc X^a Forward primer MSA against amplicons



1715

1716 Appendix B40. Inc X^a Reverse primer MSA against amplicons



1717

1718

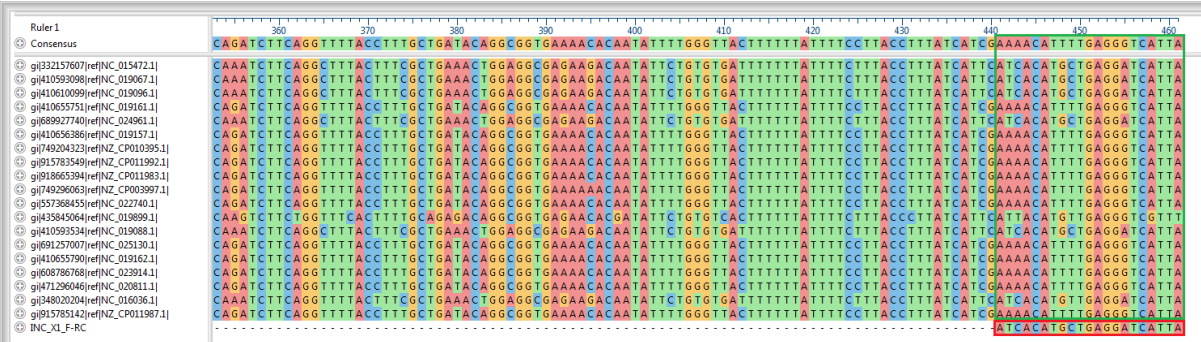
1719 Appendix B41. Inc X1^d Forward primer MSA against 19 amplicons



1720

1721

1722 Appendix B42. Inc X1^d Reverse primer MSA against 19 amplicons



1723

1724

1725 Appendix B43. Inc X2^d Forward primer MSA against 19 amplicons



1726

1727 Appendix B44. Inc X2^d Reverse primer MSA against 19 amplicons



1728

1729

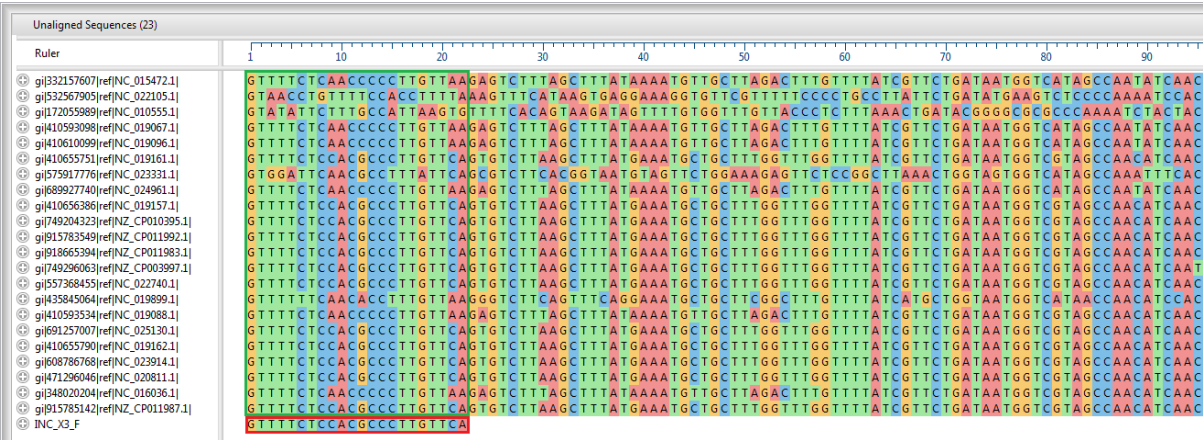
1730

1731

1732

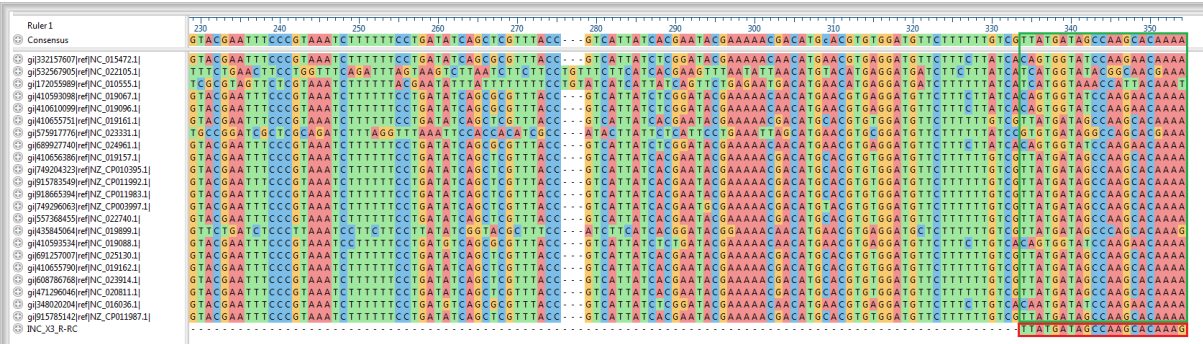
1733

1734 Appendix B45. Inc X3^d Forward primer MSA against 22 amplicons



1735

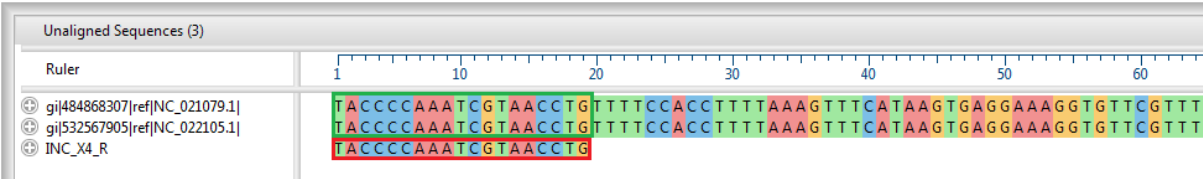
1736 Appendix B46. Inc X3^d Reverse primer MSA against 22 amplicons



1737

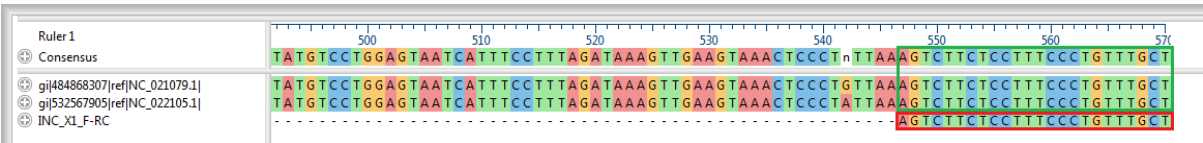
1738

1739 Appendix B47. Inc X4^d Forward primer MSA against 2 amplicons



1740

1741 Appendix B48. Inc X4^d Reverse primer MSA against 2 amplicons



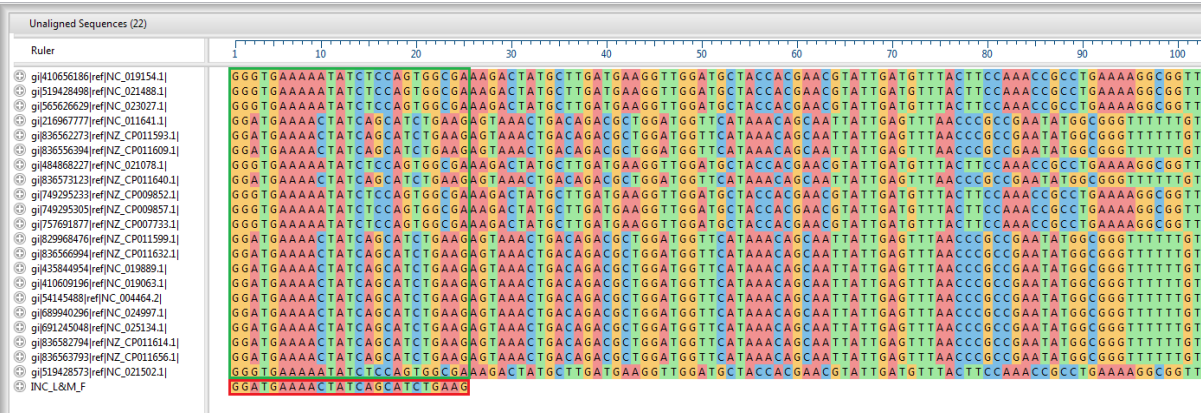
1742

1743

1744

1745

1746 Appendix B49. Inc L & M^a Forward primer MSA against 21 amplicons



1747

1748 Appendix B50. Inc L & M^a Reverse primer MSA against 21 amplicons



1749

1750 Appendix B51. Inc N^a Forward primer MSA against 44 amplicons



1751

1752

1753

1754 Appendix B52. Inc N^a Reverse primer MSA against 44 amplicons

1755

1756

1757 Appendix B53. Inc Y^a Forward primer MSA against 1 amplicons

1758

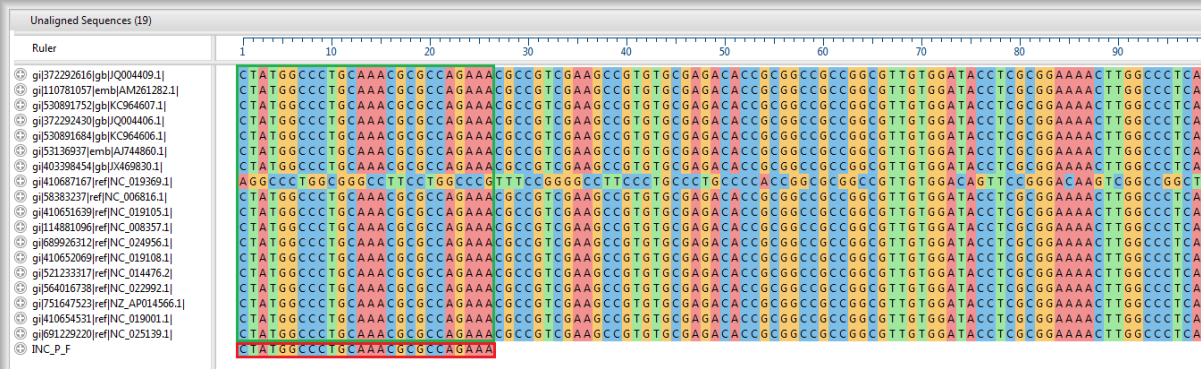
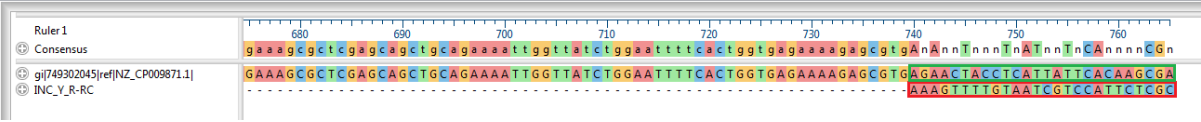
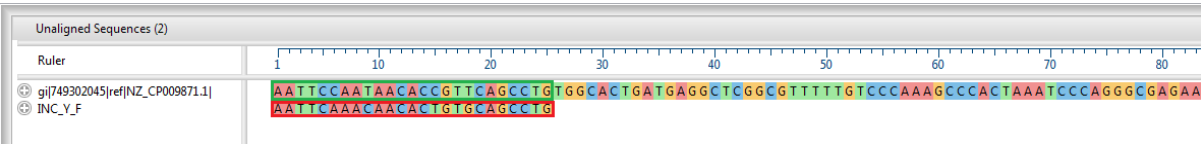
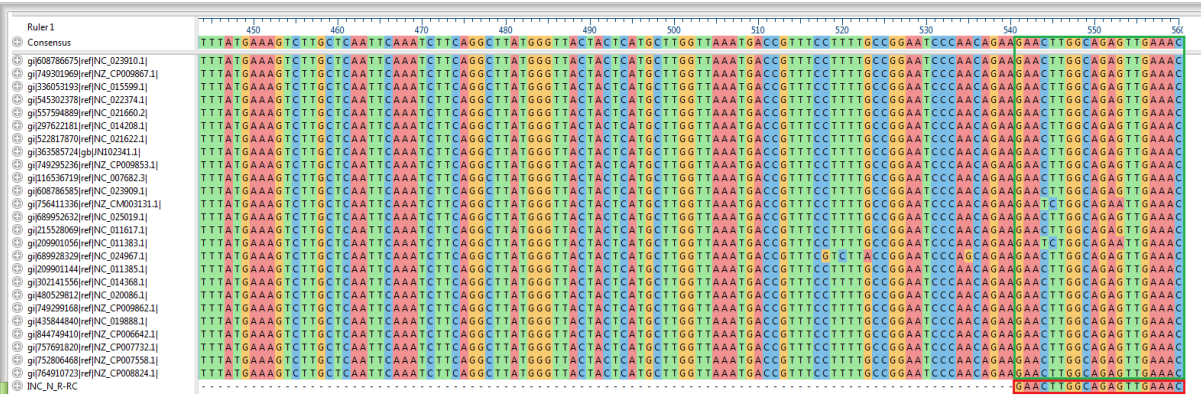
1759 Appendix B54. Inc Y^a Reverse primer MSA against 1 amplicons

1760

1761 Appendix B55. Inc P^a Forward primer MSA against 18 amplicons

1762

1763



1764

1765 Appendix B56. Inc P^a Reverse primer MSA against 18 amplicons



1766

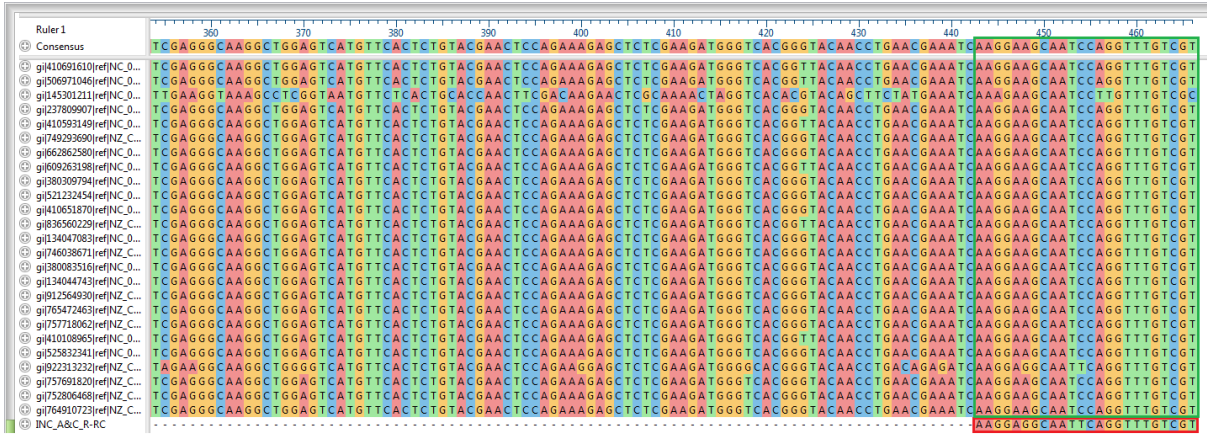
1767

1768 Appendix B57. Inc A & C^a Forward primer MSA against 57 amplicons



1769

1770 Appendix B58. Inc A & C^a Reverse primer MSA against 57 amplicons



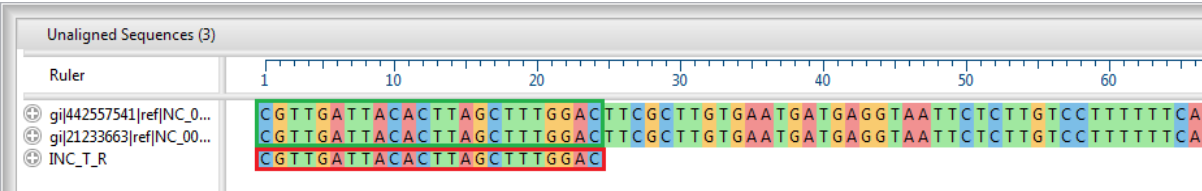
1771

1772

1773

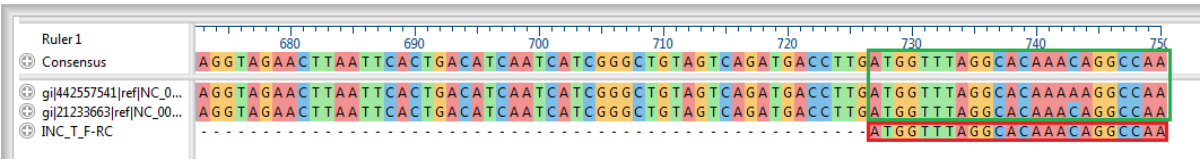
1774 Appendix B59. Inc T^a Forward primer MSA against 2 amplicons

1775



1776 Appendix B60. Inc T^a Reverse primer MSA against 2 amplicons

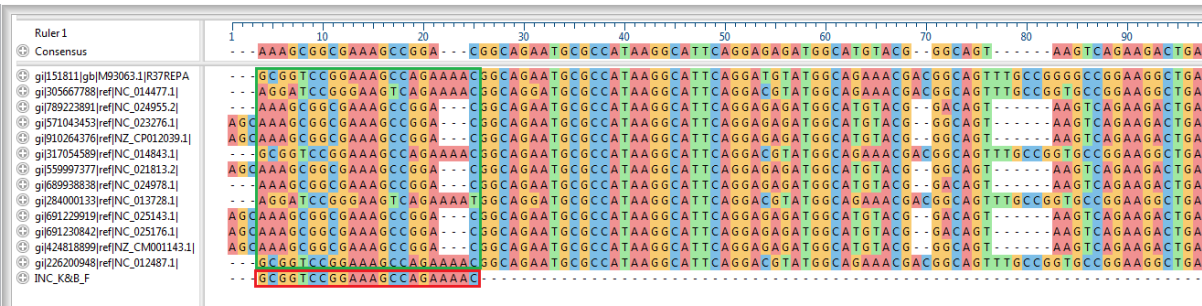
1777



1778

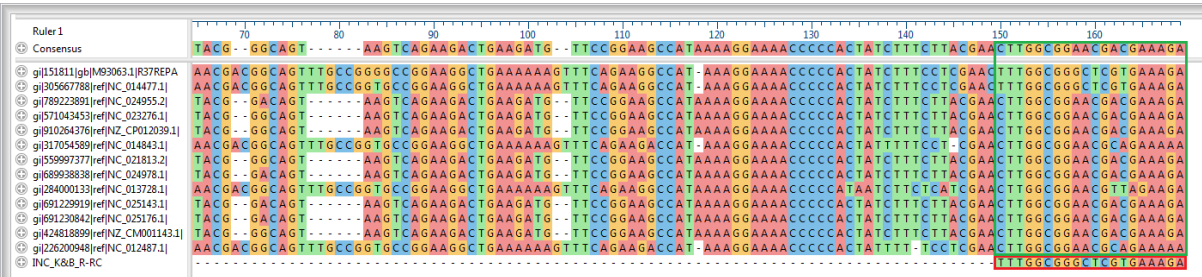
1779 Appendix B61. Inc K & B^a Forward primer MSA against 12 amplicons

1780



1781 Appendix B62. Inc K & B^a Reverse primer MSA against 12 amplicons

1782



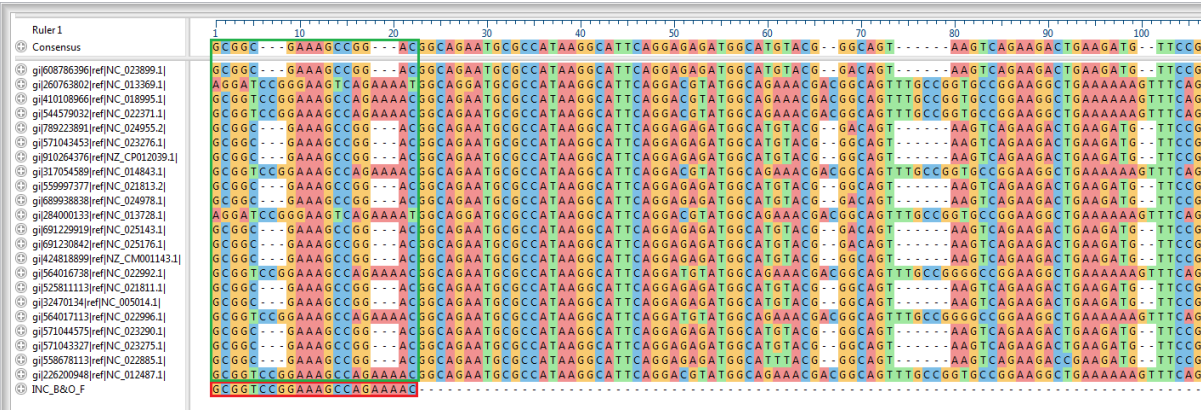
1783

1784

1785

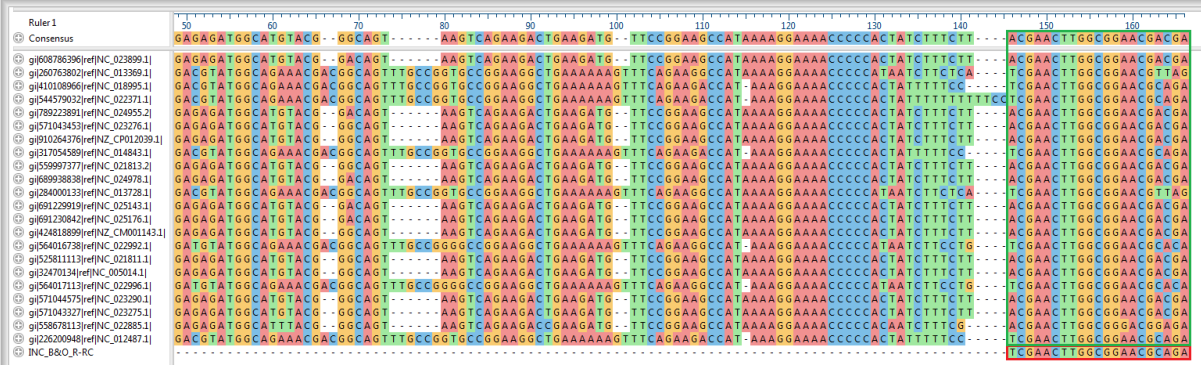
1786

1787 Appendix B63. Inc B & O^a Forward primer MSA against 22 amplicons



1788

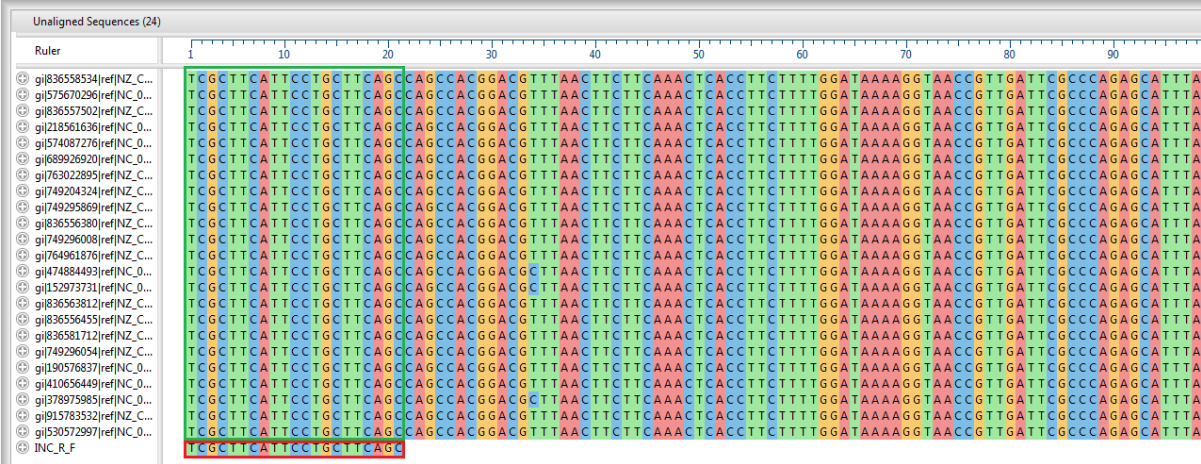
1789 Appendix B64. Inc B & O^a Reverse primer MSA against 22 amplicons



1790

1791

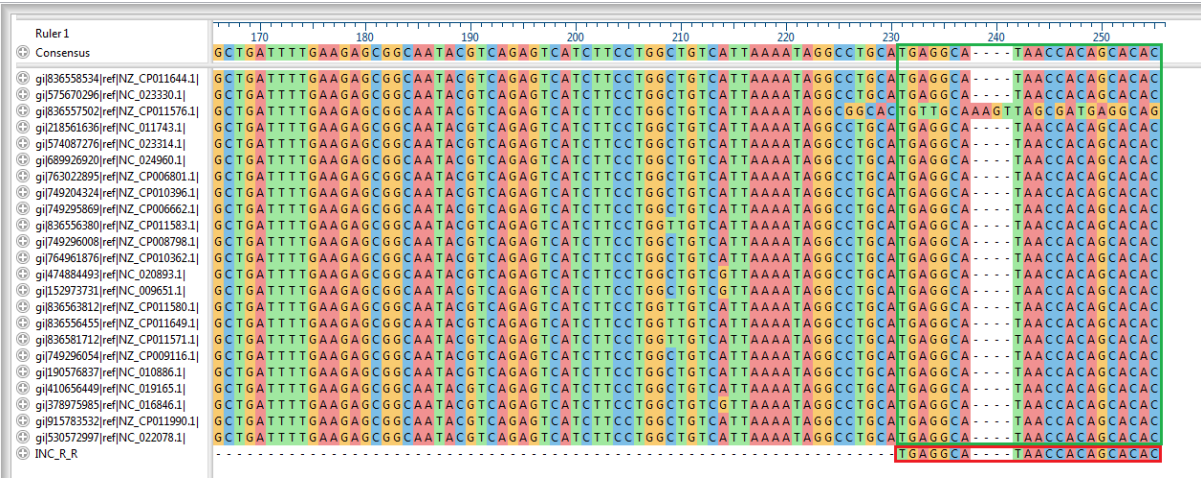
1792 Appendix B65. Inc R^C Forward primer MSA against 23 amplicons



1793

1794

1795 Appendix B66. Inc R^c Reverse primer MSA against 23 amplicons



1796

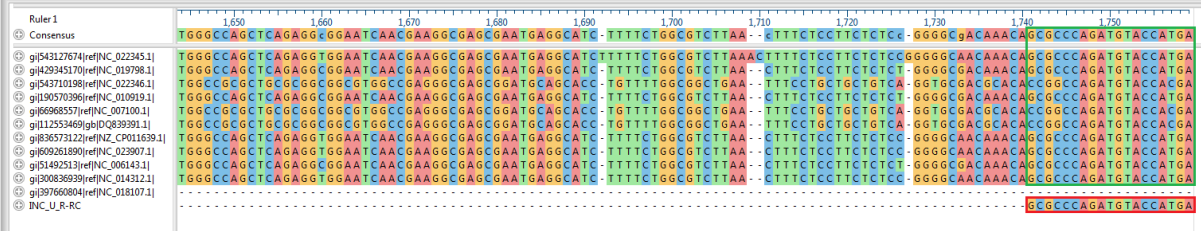
1797

1798 Appendix B67. Inc U^c Forward primer MSA against 11 amplicons



1799

1800 Appendix B68 .Inc U^c Reverse primer MSA against 11 amplicons



1801

1802

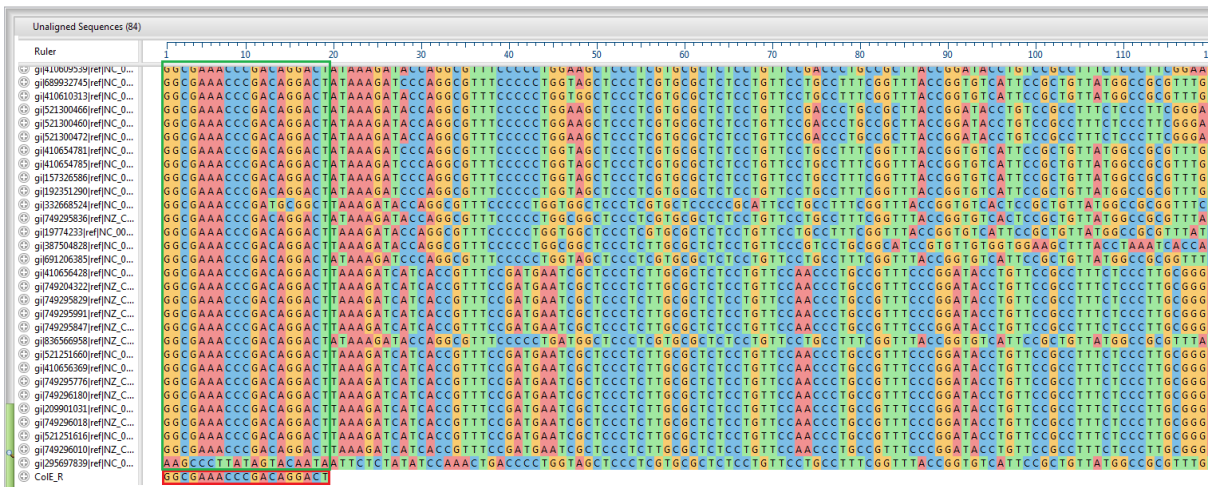
1803

1804

1805

1806

1807 Appendix B69. ColE^c Forward primer MSA against 83 amplicons



1808

1809 Appendix B70. ColE^c Reverse primer MSA against 83 amplicons



1810

6.1.3. APPENDIX C: Additional/Supplementary Figures

Appendix C1. Plasmids sorted by multidrug resistance in a population of circular resistance plasmids

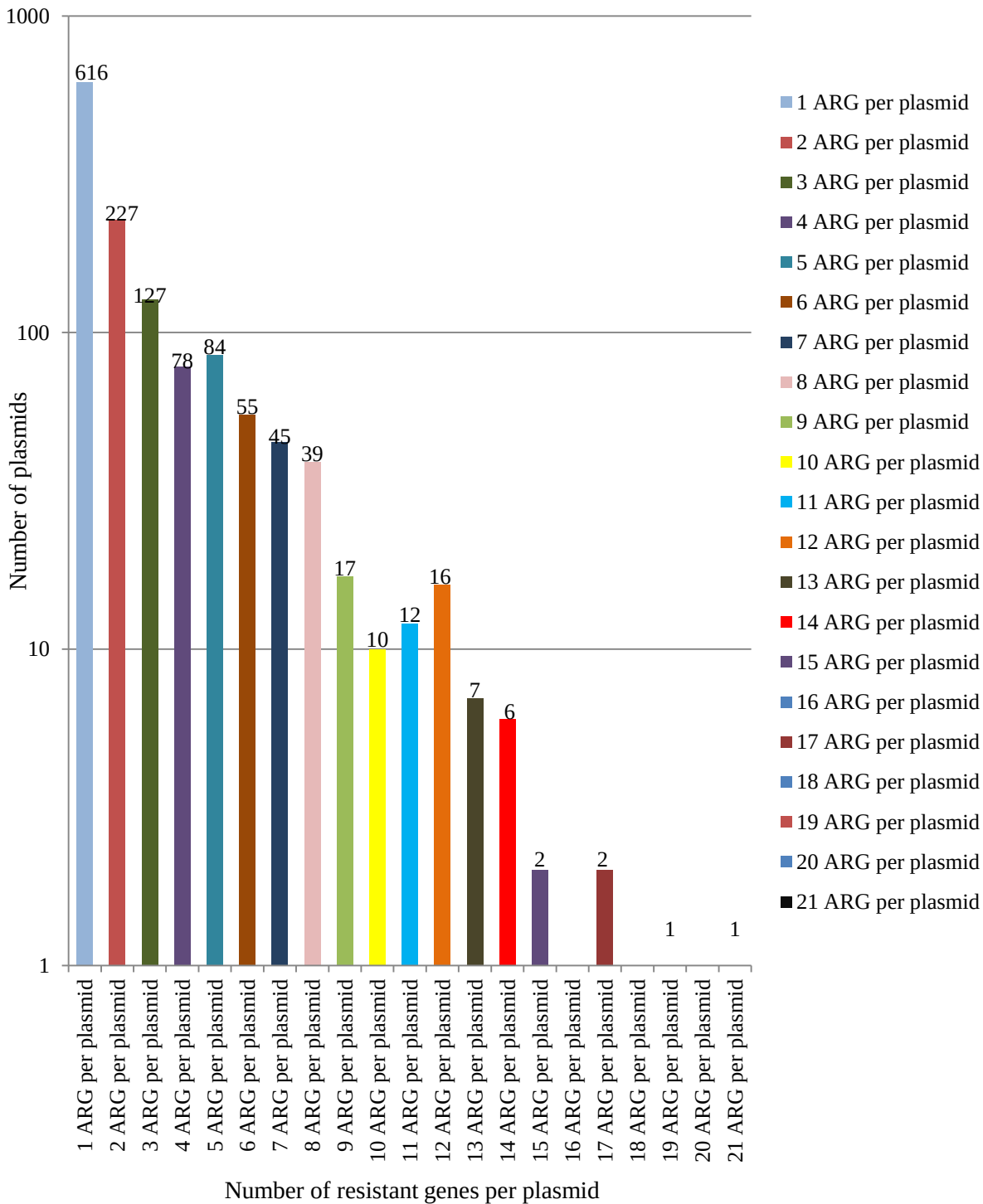
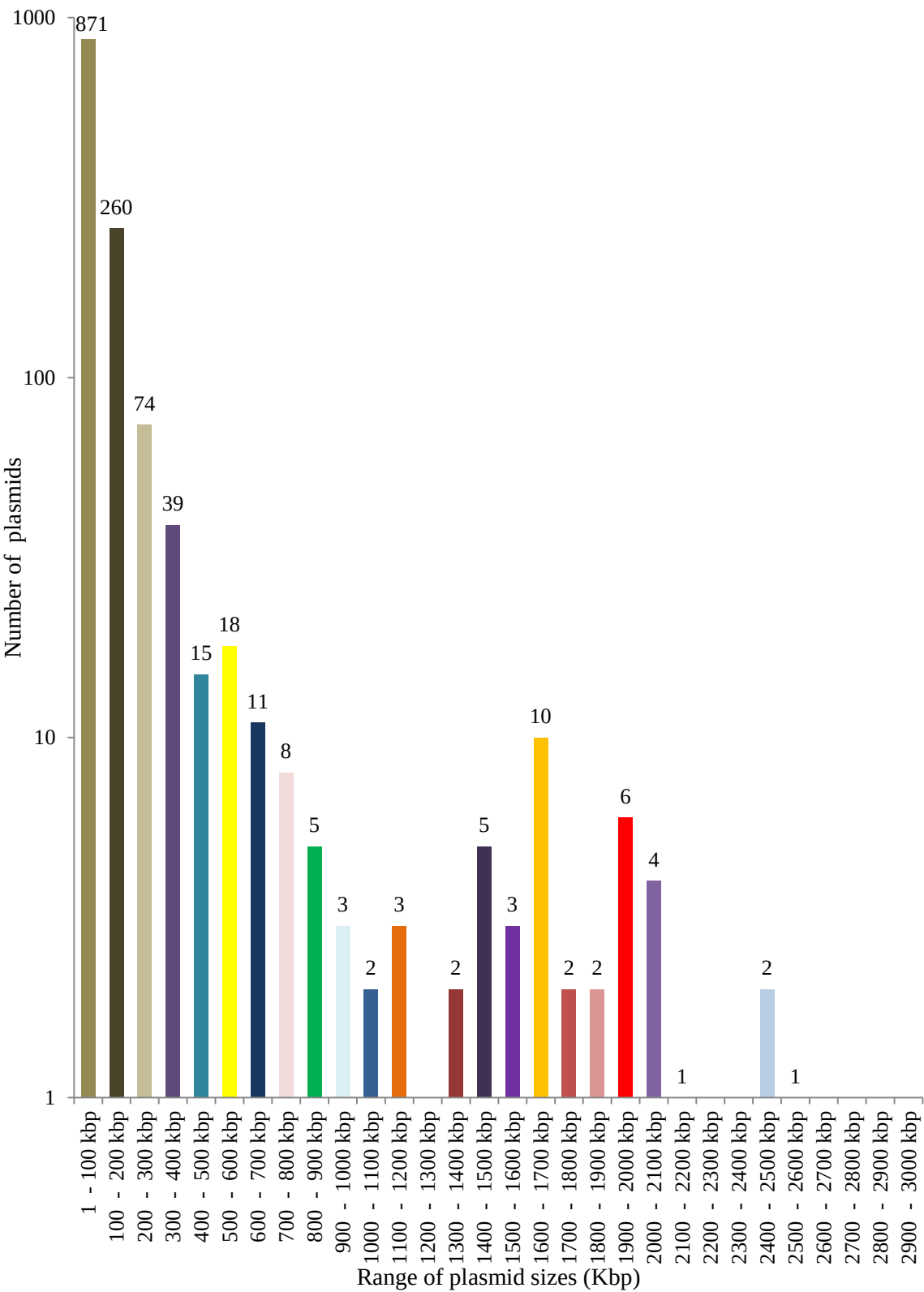


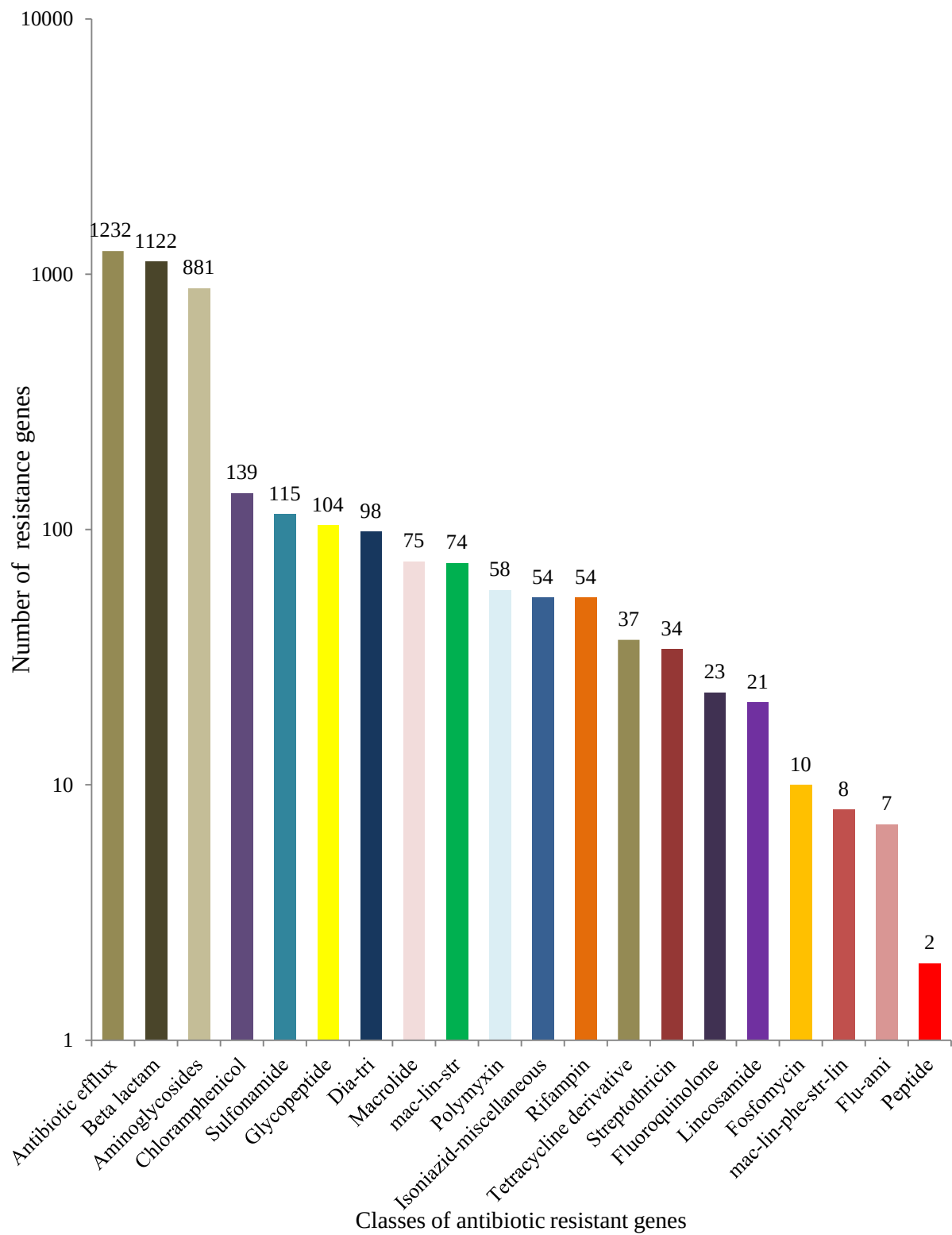
Figure 4.6b. Plasmids sorted by multidrug resistance in a population of circular resistance plasmids

1818 Appendix C2. Distribution of resistance plasmids according to the size in molecular weight
1819 (Kbp).



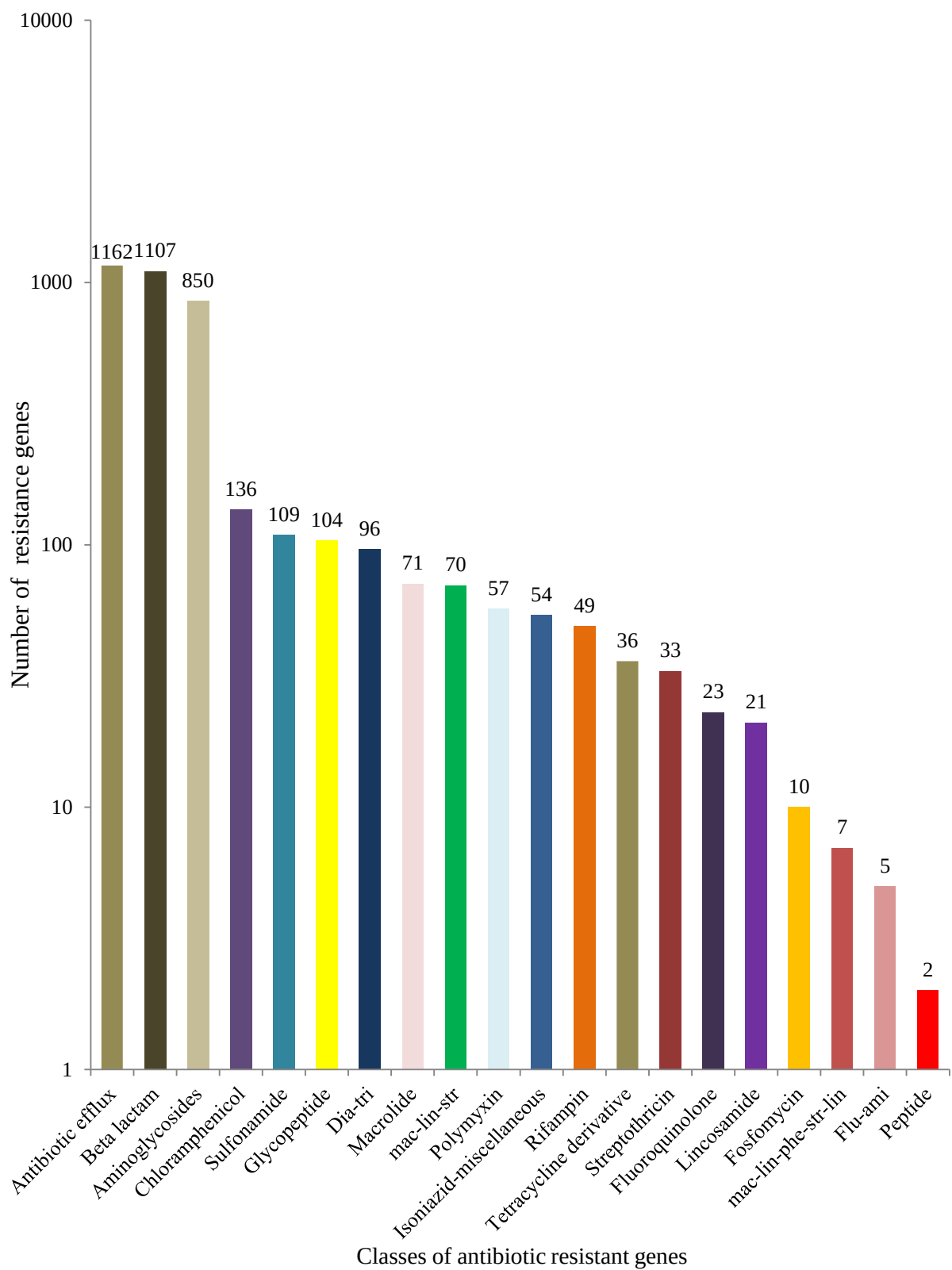
1820
1821 Figure 4.7b. Distribution of resistance plasmids according to the size in molecular weight
1822 (Kbp).

1823 Appendix C3. Classes and relative quantity of antibiotic resistant genes found in all resistant
1824 plasmids



1825
1826 Figure 5.0b. Classes and relative quantity of antibiotic resistant genes found in all resistant
1827 plasmids

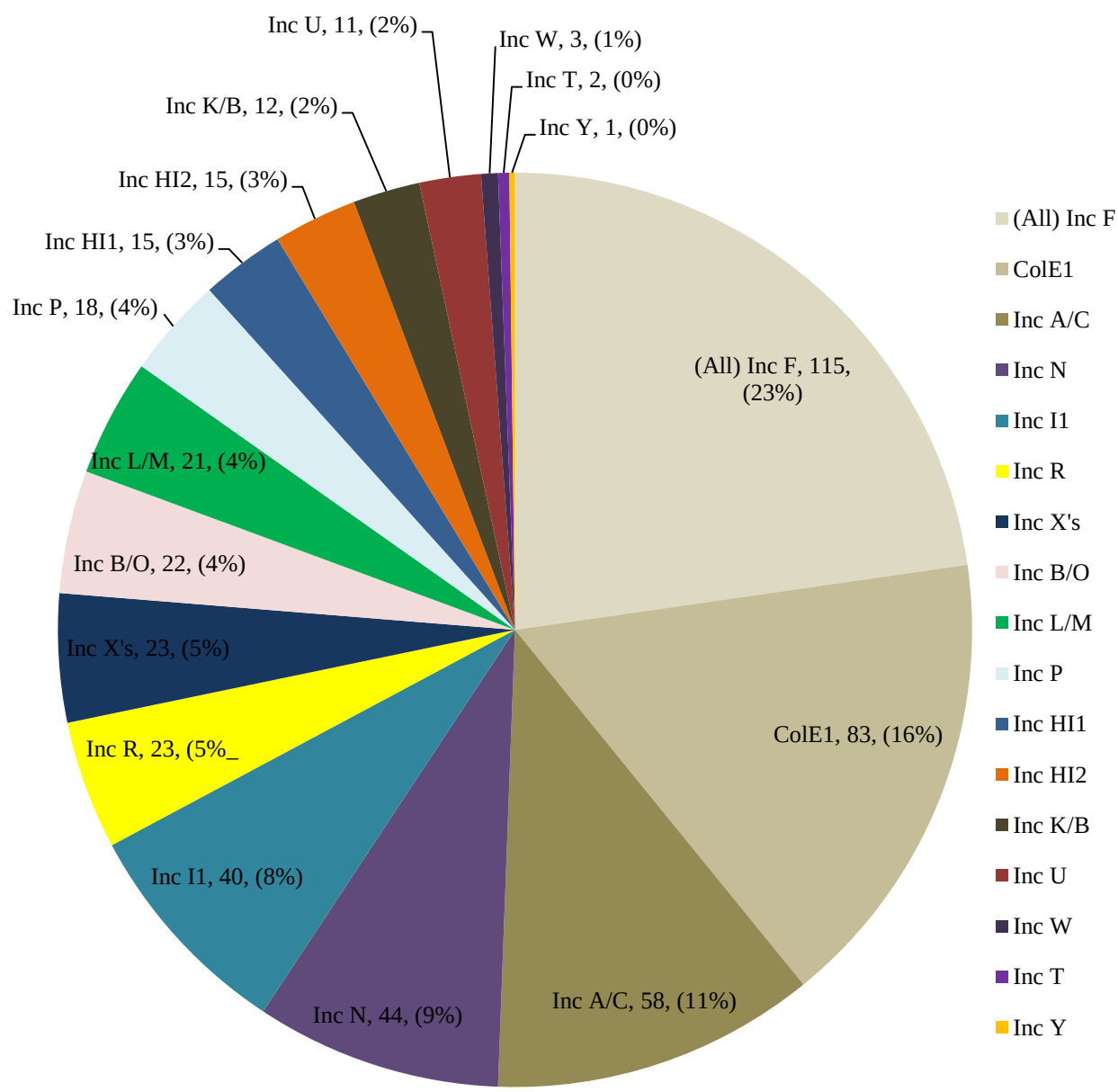
1828 Appendix C4. Classes and relative quantity of antibiotic resistant genes found in circular
1829 resistant plasmids



1830
1831 Figure 5.1b. Classes and relative quantity of antibiotic resistant genes found in circular
1832 resistant plasmids

1833

1834 Appendix C5. Antibiotic resistant plasmids (typeable using PBRT) sorted by Inc/rep
1835 region



1836

1837 Figure 5.2c. Antibiotic resistant plasmids (typeable using PBRT) sorted by Inc/rep region

1838

Appendix C6. Quantification of ARG's identified in all resistance plasmids into ARG Classes.

ARG's	ARG 1	ARG 2	ARG 3	ARG 4	ARG 5	ARG 6	ARG 7	ARG 8	ARG 9	ARG 10	ARG 11	ARG 12	ARG 13	ARG 14	ARG 15	ARG 17	ARG 19	ARG 21	Quantity (Total)
Aminoglycosides	43	82	123	98	81	88	76	94	44	23	32	50	6	10	6	13	6	6	881
Antibiotic efflux	293	178	109	69	140	76	58	64	19	16	60	40	49	37	14	5	2	3	1232
Beta-lactam	129	158	78	111	102	97	91	109	55	36	38	51	25	20	5	11	3	3	1122
Chloramphenicol	24	8	7	10	11	14	16	13	3	6	9	6	3	4	0	1	2	2	139
Dia-tri	11	9	8	8	7	13	6	8	5	1	6	7	2	4	0	1	1	1	98
Flu-ami	4	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	7
Fluoroquinolone	14	1	1	0	0	2	2	0	1	0	2	0	0	0	0	0	0	0	23
Fosfomycin	0	1	3	2	0	2	1	0	0	0	0	0	0	0	0	1	0	0	10
Glycopeptide	1	0	3	3	20	24	25	10	11	6	0	1	0	0	0	0	0	0	104
Isoniazid- miscellaneous	23	3	4	2	5	4	3	4	0	1	3	1	0	0	0	1	0	0	54
Lincosamide	10	0	1	2	1	2	2	1	0	2	0	0	0	0	0	0	0	0	21
Mac-lin-phe-str- lin	3	1	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	8
Mac-lin-str	45	7	3	4	2	1	3	4	1	2	1	1	0	0	0	0	0	0	74
Macrolide	1	6	4	3	20	10	6	6	3	3	1	6	0	3	1	0	1	1	75
Peptide	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Polymyxin	18	4	18	1	3	0	5	2	0	0	0	0	0	2	3	0	1	1	58
Rifampin	4	0	2	2	6	3	4	13	4	2	4	5	3	0	1	1	0	0	54
Streptothricin	0	4	5	3	17	1	0	1	1	1	1	0	0	0	0	0	0	0	34
Sulfonamide	2	12	17	6	6	16	10	6	6	1	6	13	3	4	0	0	3	4	115
Tetracycline derivative	21	7	4	1	2	1	0	1	0	0	0	0	0	0	0	0	0	0	37

Appendix C7. Quantification of ARG's identified in circular resistance plasmids into ARG Classes.

ARG's	ARG 1	ARG 2	ARG 3	ARG 4	ARG 5	ARG 6	ARG 7	ARG 8	ARG 9	ARG 10	ARG 11	ARG 12	ARG 13	ARG 14	ARG 15	ARG 17	ARG 19	ARG 21	Quantity (Total)
Aminoglycosides	41	76	122	91	80	81	73	90	44	23	32	50	6	10	6	13	6	6	850
Antibiotic efflux	275	165	104	68	136	76	58	53	19	16	42	40	49	37	14	5	2	3	1162
Beta-lactam	127	154	78	109	102	96	88	106	55	36	38	51	25	20	5	11	3	3	1107
Chloramphenicol	23	7	7	10	11	14	15	13	3	6	9	6	3	4	0	1	2	2	136
Dia-tri	11	8	8	8	7	13	6	7	5	1	6	7	2	4	0	1	1	1	96
Flu-ami	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
Fluoroquinolone	14	1	1	0	0	2	2	0	1	0	2	0	0	0	0	0	0	0	23
Fosfomycin	0	1	3	2	0	2	1	0	0	0	0	0	0	0	0	1	0	0	10
Glycopeptide	1	0	3	3	20	24	25	10	11	6	0	1	0	0	0	0	0	0	104
Isoniazid- miscellaneous	23	3	4	2	5	4	3	4	0	1	3	1	0	0	0	1	0	0	54
Lincosamide	10	0	1	2	1	2	2	1	0	2	0	0	0	0	0	0	0	0	21
Mac-lin-phe-str-lin	3	1	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	7
Mac-lin-str	42	7	3	3	2	1	3	4	1	2	1	1	0	0	0	0	0	0	70
Macrolide	1	6	3	3	20	9	6	4	3	3	1	6	0	3	1	0	1	1	71
Peptide	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Polymyxin	18	4	18	1	3	0	5	1	0	0	0	0	0	2	3	0	1	1	57
Rifampin	3	0	2	2	6	3	4	11	4	2	2	5	3	0	1	1	0	0	49
Streptothricin	0	3	5	3	17	1	0	1	1	1	1	0	0	0	0	0	0	0	33
Sulfonamide	2	10	16	6	6	13	10	6	6	1	6	13	3	4	0	0	3	4	109
Tetracycline derivative	20	7	4	1	2	1	0	1	0	0	0	0	0	0	0	0	0	0	36

Appendix C8. Comprehensive AR frequency table of all typeable plasmids in various Inc/rep groups

Inc/rep group	Plasmid population	*Number of ARG ORF (CARD)	Number of ARG classes	# Frequency
Inc W	3	9	4	9/3 = (3)
Inc F's	114	451	14	451/114 = (3.95)
Inc HI1	15	93	10	93/15 = (6.2)
Inc HI2	10	102	10	102/10 = (10.2)
Inc I1	40	121	10	121/40 = (3.025)
Inc X	4	5	3	5/4 = 1.25
Inc X1 & X2	19	27	7	27/19 = (1.42)
Inc X3	22	37	8	37/22 = (1.68)
Inc X4	2	5	3	5/2 = (2.5)
Inc L & M	21	79	11	79/21 = (3.76)
Inc N	44	189	14	189/44 = (4.3)
Inc Y	1	1	1	1/1 = (1)
Inc P	18	56	10	56/18 = (3.1)
Inc A & C	58	405	14	405/58 = (6.98)
Inc T	2	3	4	3/4 = (0.75)
Inc K & B	12	39	9	39/12 = (3.25)
Inc B & O	22	74	10	74/22 = (3.36)
Inc R	23	159	11	159/23 = (6.91)
Inc U	11	48	14	48/11 = (4.36)
ColE1	83	150	12	150/83 = (1.81)