Vol. 4(1), pp. 1-13, July-December 2019
DOI: 10.5897/ISABB-JBB2019.0017
Article Number: 9A9E86662195
ISSN 2163-9758
Copyright © 2019
Author(s) retain the copyright of this article
http://www.academicjournals.org//ISABB-JBB

ACADEMIC JOURNALS
expand your knowledge

ISABB Journal of Biotechnology and Bioinformatics

Full Length Research Paper

In silico detection and correlative analysis of antibiotic resistance plasmid-incompatibility (Inc/ rep) groups from different environments

Monamodi Kesamang and Teddie O. Rahube*

Department of Biological Sciences and Biotechnology, Faculty of Science, Botswana International University of Science and Technology, Palapye, Botswana.

Received 13 September, 2019; Accepted 14 October, 2019

Plasmids are extra-chromosomal mobile DNA elements found mostly 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. The occurrence and distribution by plasmid incompatibility (Inc/rep) groups is not well understood. The study sought to determine the extent of diversity and distribution of different Inc/rep groups that are harbouring ARGs in three different environments broadly classified as natural, host-associated and managed. 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. From the metadata, resistance plasmids constitute 1404 (25%) of known natural plasmids; 729 (54%) of which have multidrug resistance (more than 1 resistance gene per plasmid). The majority of antibiotic resistance plasmids were found in host associated environments particularly in human samples from clinical settings. It was also clearly elucidated by the data that the diversity of natural plasmids surpasses the efficiency of polymerase chain reaction (PCR) based replicon typing (PBRT) plasmid identification scheme, which managed to identify 460 (35%) of the resistance plasmids in the dataset. Some major limitation were identified in this study, we recommend that sequence data submission to public databases be improved for future bioinformatics studies that aim to look particularly at evolution of plasmids encoding antibiotic resistance.

Key words: Plasmids, antibiotic resistance genes, polymerase chain reaction (PCR) based replicon typing (PBRT), incompatibility, environment, bioinformatics.

INTRODUCTION

The global community has been struggling with deterring the continuous emergence of bacteria that have acquired

*Corresponding author. E-mail: rahubet@biust.ac.bw.

Author(s) agree that this article remain permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

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 excessive figures largely from patients that die from untreatable illnesses owing to failing antibiotic regimes (CDC, 2013; WHO, 2014). It is now clearly evident from numerous studies that plasmids play a major role in the spread of antibiotic resistance genes (ARGs) in the environment that eventually make their way to clinical settings (Pop-Vicas et al., 2008; Davies and Davies, 2010; Rahube and Yost, 2010; Kim et al., 2011; Marti et al., 2013). Jointly field studies conducted at different parts of the globe have obtained abundant data about the diversity of a wide range of different classes of plasmids with ARGs that occur in different sources (Martinez, 2012). 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 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; Jia et al., 2017). 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. Classification of plasmids in its own right has undergone much development since earlier classifications into conjugation ability by Couturier and colleagues (Couturier et al., 1988). The Plasmid based replicon typing (PBRT) technique was developed in 2005 using PCR primers for the incompatibility (Inc) groups F, FIA, FIB, FIC, FIIA, HI1, HI2, I1-Ig, L/M, N, P, W, T, A/C, K, B/O, X, and Y replicons (Carattoli et al., 2005). The technique has since its inception been subsequently amended for specific replicons (García-Fernández et al., 2009; Villa et al., 2010; Johnson et al., 2012). Different micro-environments (either anthropogenically influenced or pristine) have the capacity to handle varying amounts of microbial loads depending on factors like nutrient and water availability, amongst others (Fierer and Lennon 2011; Gibbons and Gilbert, 2015). Abundance of such beneficial factors has the capability to increase immensely microbial load and diversity such as seen in waste water treatment plants (WWTPs). Such conditions serve as a sanctuary for AR plasmids, within such a distribution of plasmids some may have a fitness advantage over others due to replication proficiency and copy number variation along with the acquisition of a beneficial gene; resistance (Hall, 2004; Aminov, 2009; Svara and Rankin, 2011). This concordance can also be attributed to the permissive properties of the host (Heuer and Smalla, 2012). An eventual source of epidemiologically relevant AR plasmids and associate hosts that are subject for dissemination are then selected for in a manner explained above from different environments (Rahube and Yost, 2010).

The overall aim of this study is to determine the extent of diversity and distribution of different Inc/rep groups that are harbouring ARGs in three different environments broadly classified as natural, host-associated and managed. Furthermore, depending on how abundant such plasmids were from any given environment category, we would infer evolutionary diversity of replicons found exclusively in that ecological grouping. Indeed, we had intentions of eventually extrapolating implications of anthropogenic activities in such environmental niches. For this proposition; we have thus used a large-scale bioinformatics approach based on analysing publicly available, completely sequenced plasmids available in the NCBI repository.

MATERIALS AND METHODS

Data collection and screening

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/). 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. Variations of the sampling approach are explained elsewhere (Pal et al., 2015). Plasmid sequences that have been submitted to the NCBI databases can either be complete or incomplete depending on the purpose for which the maiden research was done. Complete sequences were selected for by an iterative screening procedure against NCBI submission records.

Data availability

The sequence dataset used in this study was retrieved from the NCBI Refseq ftp repository index of; ftp://ftp.ncbi.nlm.nih.gov/genomes. The dataset is also attached as supplementary material (S2).

Determination of AR plasmids

The comprehensive antibiotic resistance database (CARD) was used to select for ARGs from all bacterial plasmid sequences described as complete and circular. For this analysis, all sequences were uploaded subsequently depending on the intensity of the resistance genes per group into an analysis function in the CARD Resistance gene identifier (VERSION 2) online, available at; (http://arpcard.mcmaster.ca/?q=CARD/tools/RGI). The resistance gene identifier function determines the open reading frames (ORFs) 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.

In-silico sequence analysis and plasmid detection

The DNAstar software genomics suite was used for alignments, the MegAlign-pro function (DNAstar® Lasergene 12, Madison, USA). The Mauve alignment function was used to determine Local collinear block of PBRT amplicons as amplified by their respective primer pairs. Due to the fact that results were expected

Table 1. Description of the environment categories.

Natural environments	Host-associated Environments	Managed environments	Unknown environments
All open pristine environments that are not affected by anthropogenic activities e.g. natural soils and sediments, natural aquatic sources etc.	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.	All inanimate environments in which human activities are practiced e.g. waste water treatment plants (WWTP), acid mine drainage, households interior surfaces etc.	All environments with no documented isolation source.

to remain as consistent as possible to those determined from wet lab PBRT analysis, the alignment stringency was set at a default seed weight of between 13 and 15, with the former (13) allowing detection of an optimal diversity threshold (Darling et al., 2010).

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 (Martinez, 2012) (Table 1). 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 reference literature. 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 exercise, each individual plasmid was cross referenced with its NCBI nucleotide database (Refseq and 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; as explained by Pal and colleagues (Pal et al., 2015). In instances where no specific source could be determined for a given plasmid in both interrogation events 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; denoted unknown environment.

RESULTS AND DISCUSSION

Characterisation of antibiotic resistance in plasmid dataset

Initially, 6,090 plasmids that belonged to 429 genera of organisms from all the 3 domains of life including 56 plasmids of metagenomic samples designated to uncultured bacteria were extracted as the metadata. An initial screening procedure was made to remove plasmids from hosts of Eukaryota and Achaean origin, these constituted 195 (3%) of the initial population (Figure S1 in the supplementary materials). A subsequent screening procedure removed all incomplete sequences; inclusive undesignated sequences, that is partial, contigs and/or scaffolds etc. These constituted 299 (5.4%) of the successive total. The resultant dataset was characterised for resistance. The plasmid metadata was sorted by resistant against non- resistant (Figure 1A). From the CARD analysis, 1404 (25%) of plasmids were resistant and 4276 (75%) plasmids were found to be non-resistant. Inferring multidrug resistance as a function of number of ARG ORF per plasmid (p/p), 616 (45%) of the plasmid population were found to carry only as single resistance genotype while 729 (55%) exibit multidrug resistance phenotypes (Figure 1B). ARG plasmids allocated by phyla of their host bacteria, the results revealed that majority of plasmids were hosted by the major phyla; proteobacteria, firmicutes, actinobacteria and bacteriodetes (Figure 1C).

The ranges of the size of plasmids (n=1404) were segmented into 4 segments (depending on the density of plasmids per given size); 1 – 20 Kbp (green), 20 – 70 Kbp (blue), 70 – 170 Kbp (red) and 170 – 2600 Kbp (yellow) (Figure 2A). The chords (hierarchical edge bundles/HEB) corresponding in colour code with plasmids of a given range/quartile were correlated with the number of ARG ORF per plasmid (denoted as ARG 1, ARG 2, etc.). It was predicted that small sized plasmids will have few ARGs compared to large sized plasmids. However, the results reveal that even relatively small plasmids (>20Kbp) may carry multiple ARGs, larger plasmids also can carry few ARGs. The bar graph shows the relative distribution of AR plasmids, the number of ARG and the quantity of MDR plasmids within the population across a range of sizes (See Figure 2B). The link between the chord diagram quartile ranges and bar graph plasmid populations is shown by the light grey array field/lines. All the plasmid size ranges are therefore capable of multi-drug resistance.

Replicon typing of AR plasmids

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 PBRT technique managed to identify only 460 (35%)

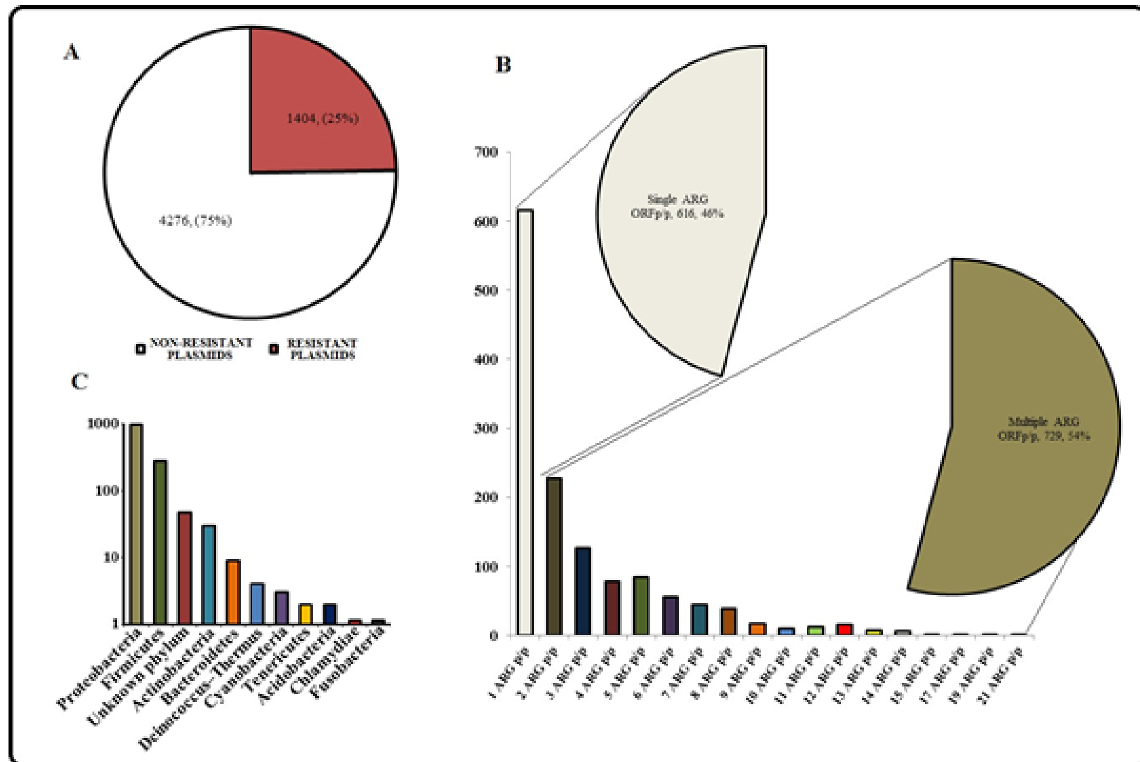

Figure 1. Characterisation of the AR Plasmid dataset.

of resistance plasmids belonging to Enterobacteriaceae, the remaining 885 (65%) could not be typed (non-typeable) (Figure 3A). From the typeable selection (n=460), 336 (73%) of the AR plasmids had a single replicon and only 124 (27%) had more than one replicon (Figure 3B). Furthermore, from the multiple replicon group of plasmids (n=124), majority of the plasmids had 3 replicons, followed by 2 and 4 replicon plasmids (Figure 3C).

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 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 the plasmid multi-locus sequence typing (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.

Incompatibility/replicon sorting of typeable AR plasmids

The link of typeable AR plasmids against those with prevalence of multiple replicons (denoted as replicon loci) is shown in Figure 4. Chord diagram (Figure 4A) shows associations between replicon type and the number of replicon loci typed within a cluster of plasmids of a given replicon type. Bar graph (Figure 4B) shows the number of typeable plasmids associated with each Incompatibility/replicon group. The link between the chord diagram Incompatibility/replicon groups and those in the bar graph are shown by the light grey array field/lines. There were 460 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 83 (17%), Inc A&C 58 (12%), Inc N 44 (9%) and Inc I1 40 (8%) (Figure 4).

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. The Inc F plasmids can further be divided into several subgroups. The Inc F(rep B) replicon occurred in 74 plasmids,

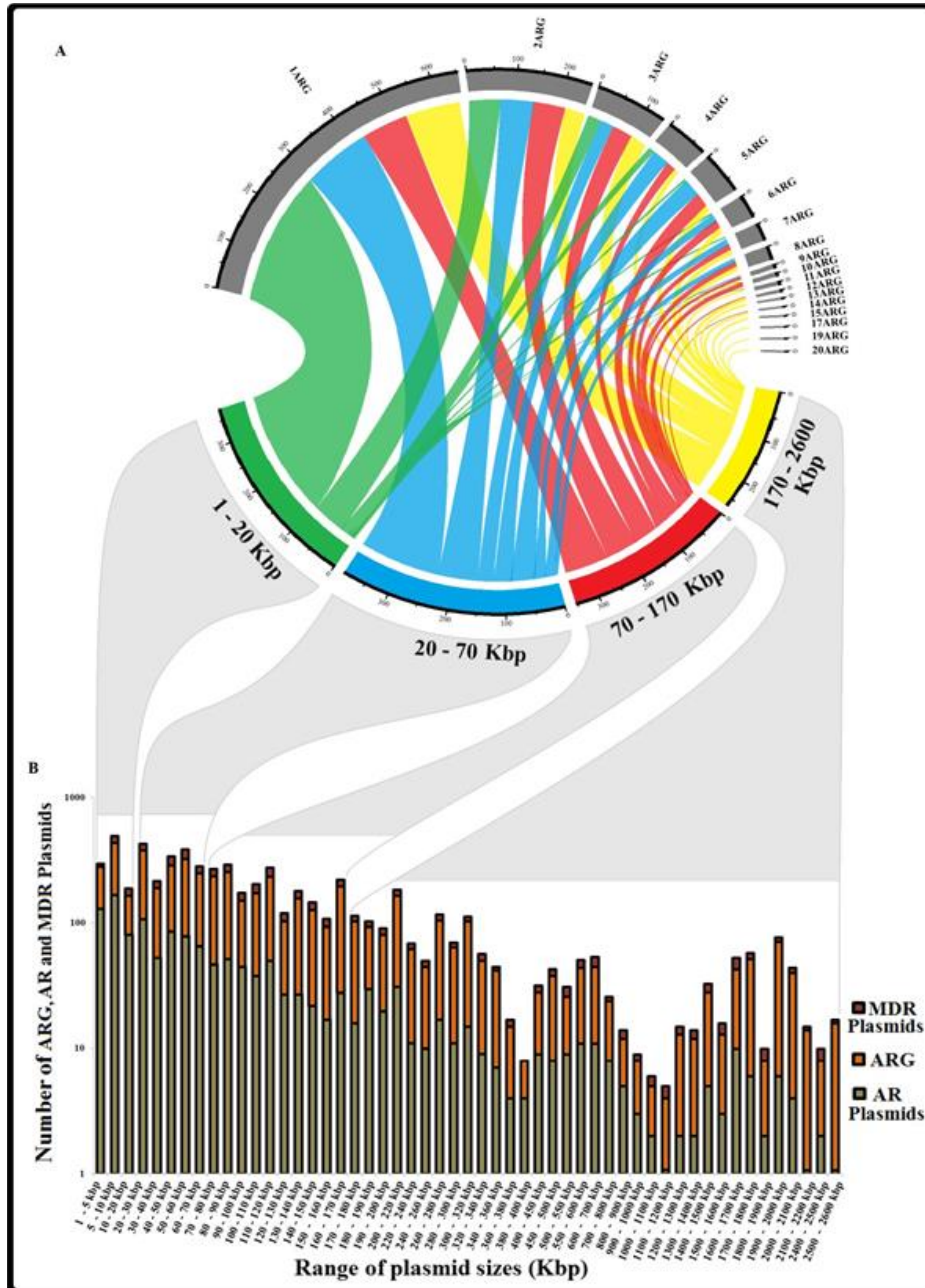

Figure 2. Association of plasmid size against prevalence of multidrug resistance as determined by number of ARG ORF per plasmid

followed by Inc FII(rep A) in 51 plasmids and Inc FII in 50 plasmids (data not shown). 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 et al. (2012) 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.,

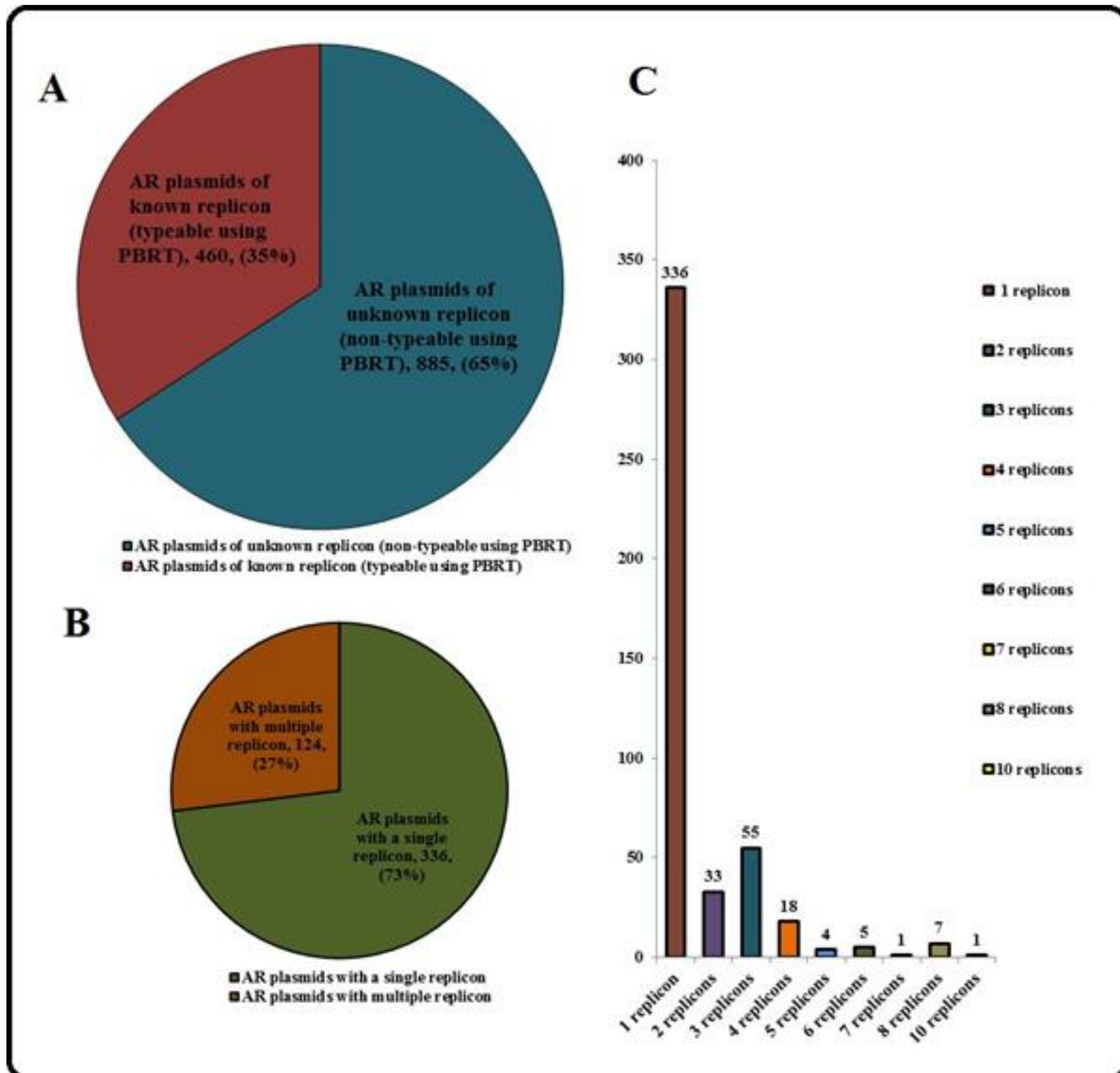

Figure 3. Identification and classification of antibiotic resistance plasmids with multiple replicons.

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 (data not shown) indicting that they share a common ancestral plasmid lineage.

Correlation of the different classes of ARGs against plasmids

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 (Figure 5).Chord diagram (Figure 5A) associates each ARG from each class to a respective Inc/rep group, the size of the chord is corresponsive to the number of ARG of each class. Bar graph (Figure 5B) indicates the abundance of plasmid acquired ARGs found in each class.

As already shown, the number of resistance plasmids

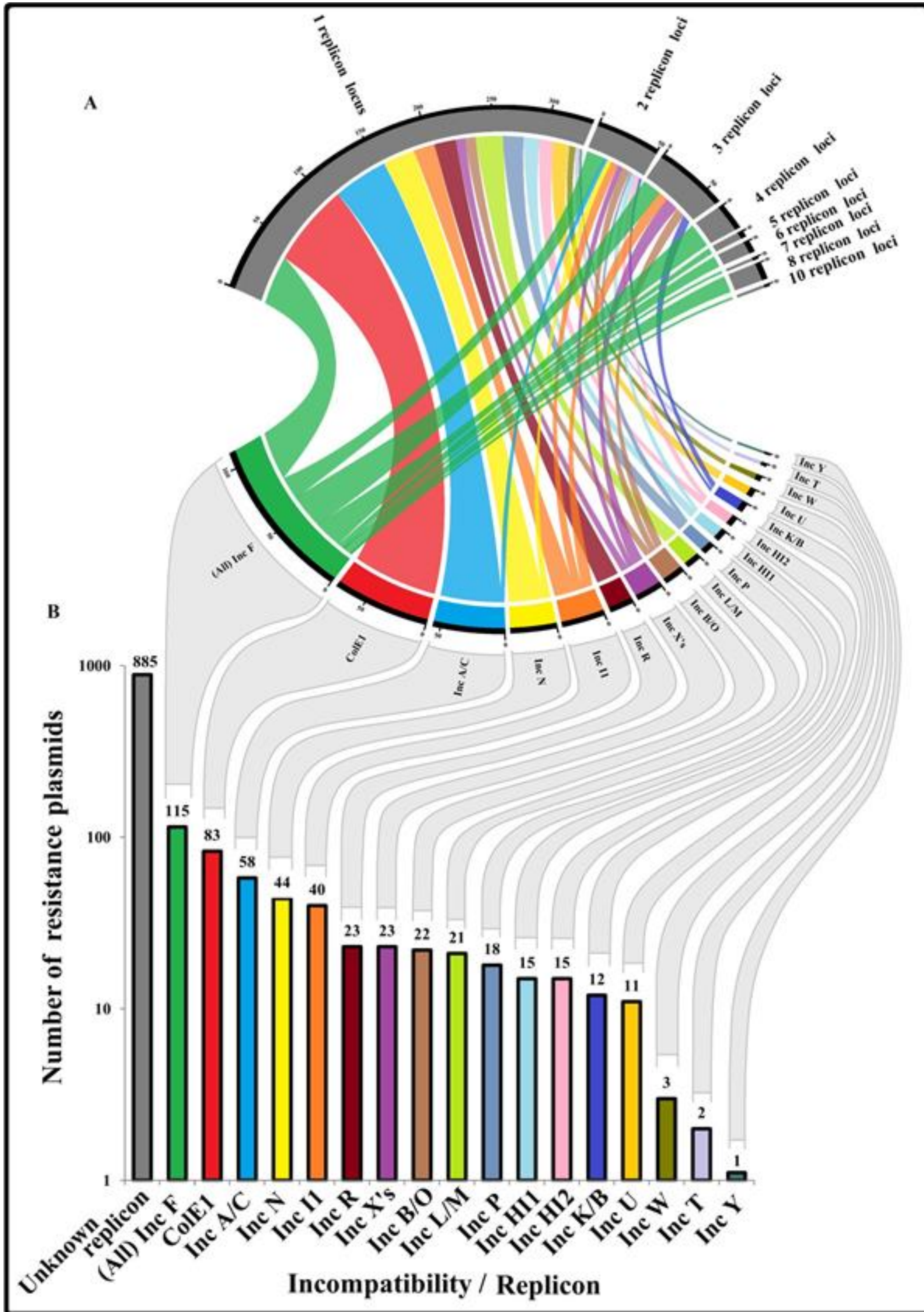

Figure 4. Associations of multiple replicon loci against plasmids of a given replicon type. Natural plasmids are known to have multiple replicons (Couturier et al., 1988).

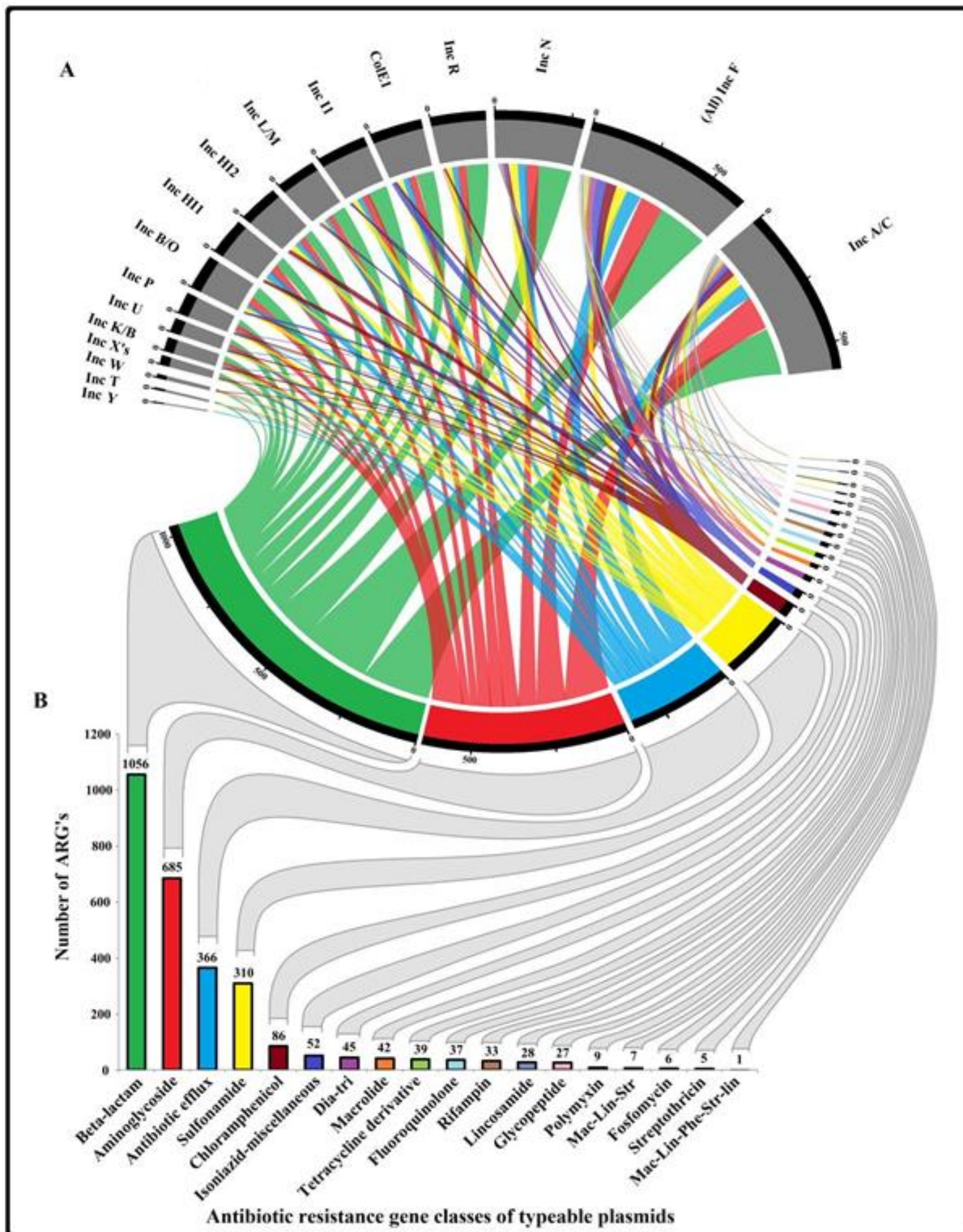

Figure 5. Correlation of the different classes of ARGs against plasmids of various replicon groups to elucidate general prevalence and possible preference.

that are typeable into different Inc/rep groups using the PBRT technique in relation to the total number of resistance plasmids was relatively low. Due to this, the quantifiable resistance genes determined by CARD are

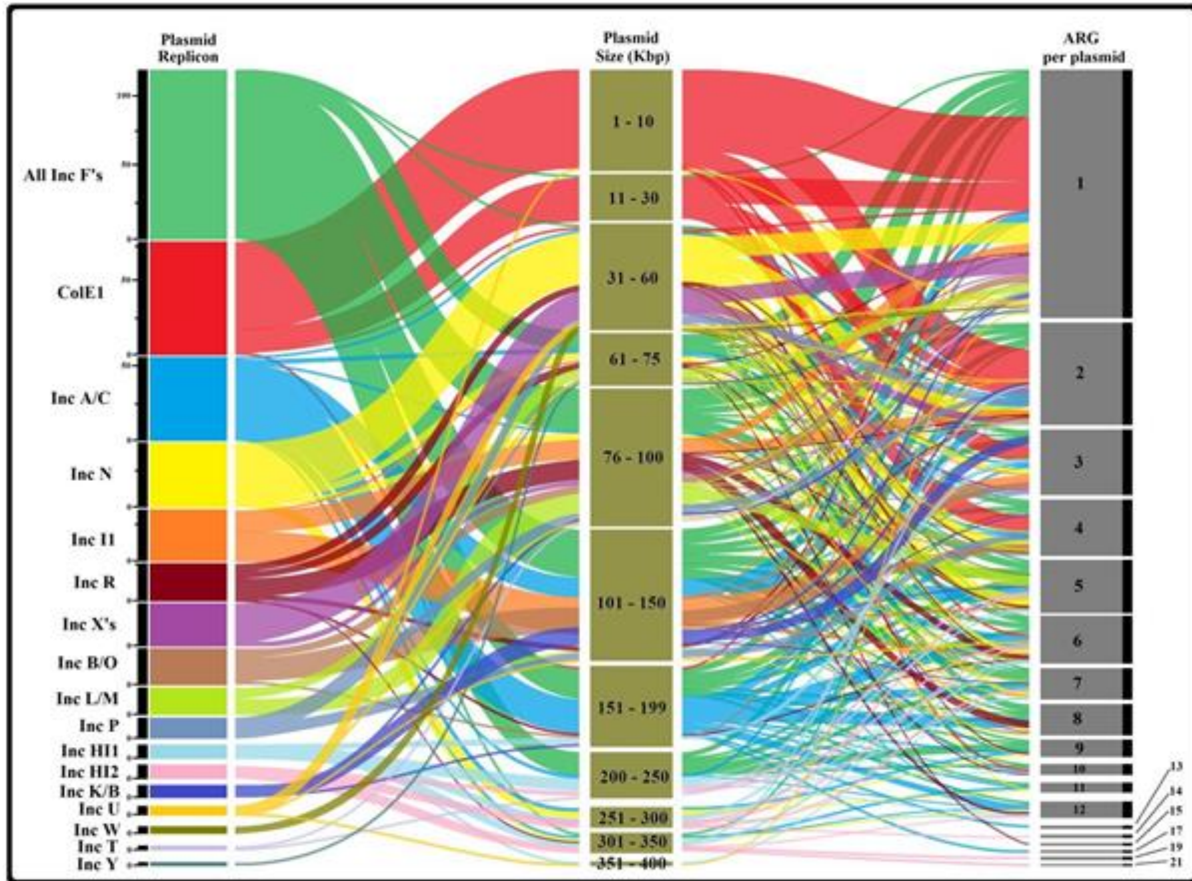

Figure 6. The sankey flow diagram indicates how the number of plasmids correlates with size (range indicated) and the corresponding ARG load (ARG per plasmid) for each replicon (colour code associated with replicon type).

not represented in their entirety in the population of plasmids that could be allocated into respective Inc/rep groups. Despite this 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 ARG 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).

The ARG 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). This observation is represented visually in the Sankey flow diagram (Figure 6). ColE1 (red hierarchical edge bundles) within a plasmid range size of 1-10 Kbp correlate with having 1 ARG per plasmid and likewise Inc HI2 (pink hierarchical edge bundles) within a plasmid range of 301 -350 Kbp correlate with having between 12 to 21 ARG per plasmid. All other Inc/replicon groups range in between having the same trend.

Association of replicon specific AR plasmids with different environments

Plasmids of known Inc/rep region were allocated to reference environment or isolation sources as shown in Table 2. Approximately 81% plasmids were found associated with a host, 4.4% of the plasmids were associated with managed environments, 2.0% plasmids

Table 2. 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’s	0	24	0	3
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(2.0%)	404(81%)	22(4.4%)	60(12.1%)

were associated with natural environments, and lastly 11.8% plasmids have unknown environments. One of the main goals of this investigation 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 (Figure 7) visually represents this correlation efficiently. As a general outlook the most striking feature of the results is that majority of plasmids were isolated from host associated environments. Further investigation into which host associated environments revealed that these were patients samples in clinical settings. 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.

Conclusion

The plasmid sequence data used in this project were collected from the NCBI nucleotide reference database at a single point in time (15 November, 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 ARGs in general because the deviation caused by the introduction of new completely sequenced plasmid 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 6,090 sequences (see supplementary info; S2) (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 such as MOB typing resistance plasmids by Orlek et al. (2017).

Based on the approach utilized in this study for investigating the ecology and diversity of plasmids derived from sequence data, the majority of antibiotic resistance plasmids were found in host associated environments particularly in human patients from clinical settings. It should be noted that the majority of these plasmids were sequenced using a genomics approach which relies on primary isolation and culture of host organism. The results may not suggest conclusive or notable variability on the proportions of plasmids of a specific incompatibility group 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

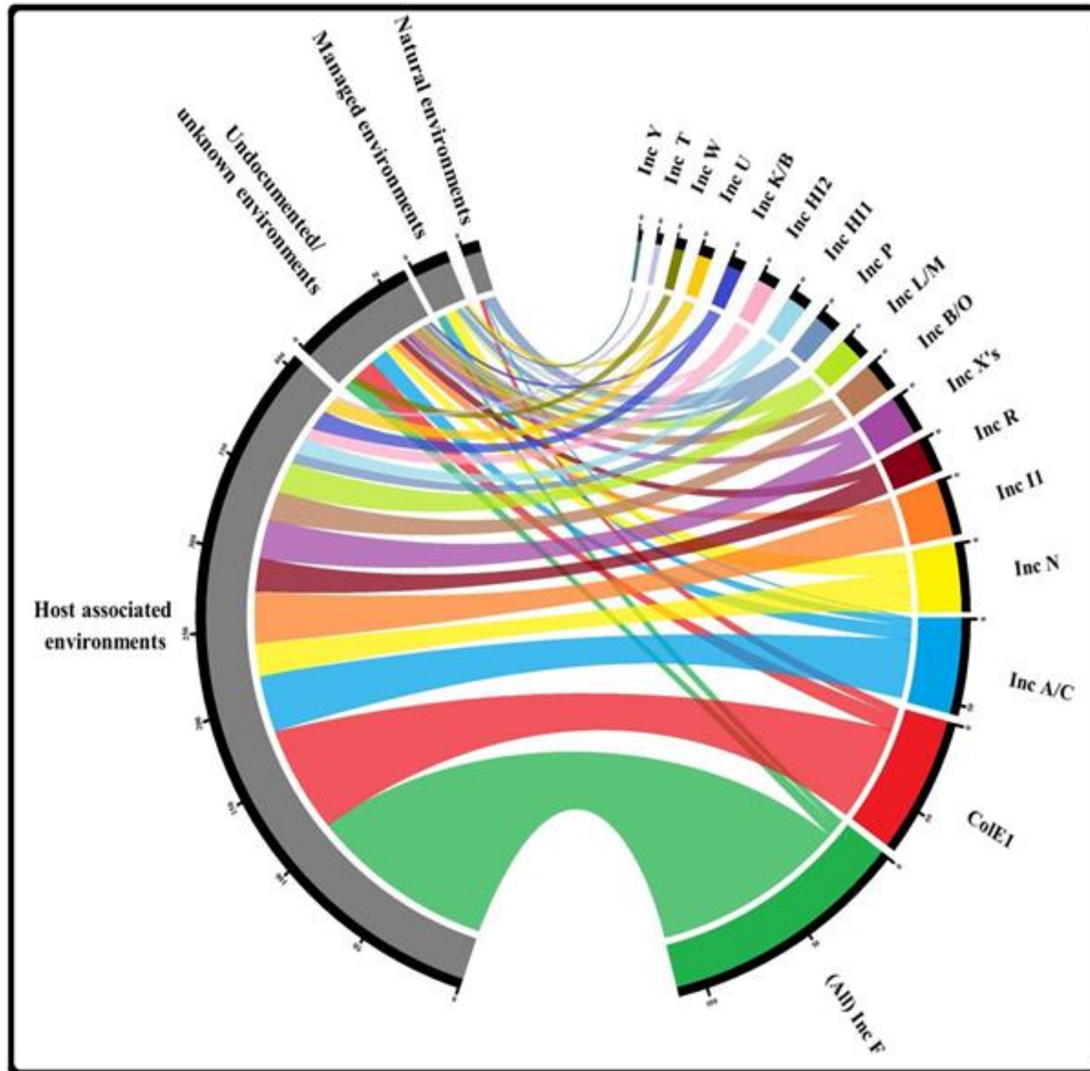

Figure 7. Chord diagram correlating plasmid incompatibility with their respective host's environmental isolation source; Natural environments, Managed environments and Host associated environments.

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 particularly at evolution of plasmids encoding antibiotic resistance.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGEMENTS

This study was funded by Botswana International University of Science and Technology (BIUST), research initiation grant [R00004] awarded to Teddie O. Rahube; TOR (Principal Investigator). Monamodi Kesamang; MK (MSc student) was also supported by BIUST through Teaching Assistantships.

REFERENCES

Alanis AJ (2005). Resistance to antibiotics: Are we in the post-antibiotic era?. Archives of Medical Research 36(6):697-705.

Aminov RI (2009). The role of antibiotics and antibiotic resistance in nature. Environmental Microbiology 11(12):2970-2988.

Carattoli A, Bertini A, Villa L, Falbo V, Hopkins KL, Threlfall EJ, (2005). Identification of plasmids by PCR-based replicon typing. Journal of Microbiological Methods 63(3): 219-228.

Carattoli A, (2009). Resistance Plasmid Families in *Enterobacteriaceae*. Antimicrobial Agents and Chemotherapy 53(6):2227-2238.

Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, MøllerAarestrup F, Hasman H (2014). In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrobial Agents and Chemotherapy 58(7):

3895-3903.
Centre for Disease Control and Prevention (CDC) (2013). Antibiotic resistance threats in the United StatesCentres for Disease Control and Prevention, US Department of Health and Human Services. Available @:<http://www.cdc.gov/drugresistance/threat-report-2013/pdf/ar-threats-2013-508.pdf#page=13>.
Couturier M, Bex F, Bergquist PL, Maas WK (1988). Identification and classification of bacterial plasmids. Microbiological Reviews 52(3):375.
Darling AE, Mau B, Perna NT (2010). progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLoS ONE 5(6):e11147.
Davies J, Davies D (2010). Origins and evolution of antibiotic resistance. Microbiology and Molecular Biology Reviews 74(3): 417-433.
Fierer N, Lennon JT (2011). The generation and maintenance of diversity in microbial communities. American Journal of Botany 98(3):439-448.
García-Fernández A, Fortini D, Veldman K, Mevius D, Carattoli A (2009). Characterization of plasmids harbouring qnrS1, qnrB2 and qnrB19 genes in Salmonella. Journal of Antimicrobial Chemotherapy 63(2):274-281
Gibbons SM, Gilbert JA (2015). Microbial diversity–exploration of natural ecosystems and microbiomes. Current Opinion in Genetics & Development 35:66-72.
Hall BG (2004). Predicting the evolution of antibiotic resistance genes. Nature Reviews Microbiology 2(5):430.
Heuer H, Smalla K (2012). Plasmids foster diversification and adaptation of bacterial populations in soil. FEMS Microbiology Reviews 36(6):1083-1104.
Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, Johnson TA, Brinkman FS, Wright GD, McArthur AG (2017). CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Research 4:45.
Johnson TJ, Bielak EM, Fortini D, Hansen LH, Hasman H, Debroy C, Nolan LK, Carattoli A (2012). Expansion of the IncX plasmid family for improved identification and typing of novel plasmids in drug-resistant Enterobacteriaceae. Plasmid 68(1):43-50.
Kline BC (1985). A review of mini-F plasmid maintenance. Plasmid 14(1):1-6.
Kim HR, Hwang SS, Kim EC, Lee SM, Yang SC, Yoo CG Kim YW, Han SK, Shim YS, Yim JJ(2011) Risk factors for multidrug-resistant bacterial infection among patients with tuberculosis. Journal of Hospital Infection 77(2):134-137.
Liu B, Pop M (2009). ARDB-Antibiotic resistance genes database. Nucleic Acids Research 37(suppl_1):D443-D447.
Marti E, Jofre J, Balcazar JL (2013). Prevalence of antibiotic resistance genes and bacterial community composition in a river influenced by a wastewater treatment plant. PLoS ONE 8(10):e78906.
Martinez-Villegas PA (2012). Exploratory Metagenomic Analysis of Antibiotic Resistance Genes in Bacterial Communities: A statistical approach for classification of the bacterial resistome. Available @; <http://publications.lib.chalmers.se/records/fulltext/160023.pdf>.
McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'brien JS, Pawlowski AC, Piddock LJ, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD (2013). The comprehensive antibiotic resistance database. Antimicrobial Agents and Chemotherapy 57(7):3348-3357.
Orlek A, Phan H, Sheppard AE, Doumith M, Ellington M, Peto T, Crook D, Walker AS, Woodford N, Anjum MF, Stoesser N (2017). Ordering the mob: Insights into replicon and MOB typing schemes from analysis of a curated dataset of publicly available plasmids. Plasmid 91:42-52.
Pal C, Bengtsson-Palme J, Kristiansson E, Larsson DG (2015). Co-occurrence of resistance genes to antibiotics, biocides and metals reveals novel insights into their co-selection potential. BMC Genomics 16(1):964.
Pop-Vicas A, Strom J, D'Agata EMC (2008). Multidrug-resistant gram-negative bacilli among patients who require chronic hemodialysis. Clinical Journal of the American Society of Nephrology 3(3):752-758.
Rahube TO, Yost CK (2010). Antibiotic resistance plasmids in wastewater treatment plants and their possible dissemination into the environment. African Journal of Biotechnology 9(54):9183-9190.
Shintani M, Sanchez ZK, Kimbara K (2015). Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. Frontiers in Microbiology 6:242.
Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EP, De la Cruz F (2010). Mobility of plasmids. Microbiology and Molecular Biology Reviews 74(3):434-452.
Svara F, Rankin DJ (2011). The evolution of plasmid-carried antibiotic resistance. BMC Evolutionary Biology 11(1):130.
Villa L, García-Fernández A, Fortini D, Carattoli A (2010). Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. Journal of Antimicrobial Chemotherapy 65(12):2518-2529.
World Health Organisation (WHO) (2014). World Health Organisation: Antimicrobial resistance: global report on surveillance 2014. Available @: < http://apps.who.int/iris/bitstream/10665/112642/1/9789241564748_eng.pdf?ua=1 >

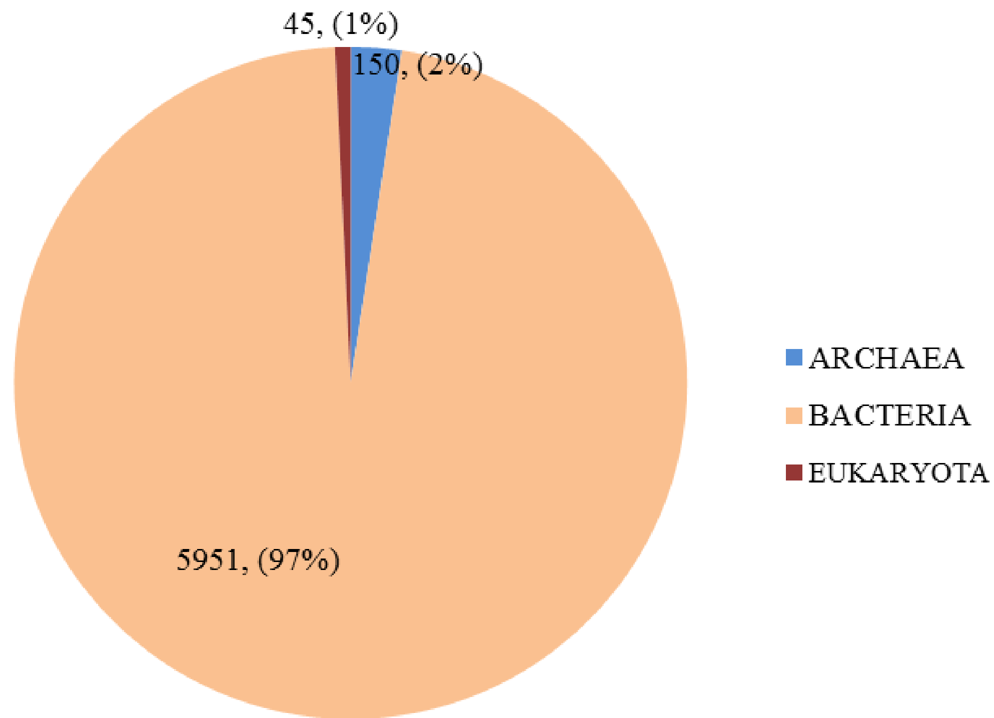

Figure S1. Plasmids sorted by domain of host organism.