Genetic Elements Identified in 10 Salmonella IncA/C Plasmids Conserved Across 18 Bacterial Genera

Lucas Harrison[1], Cong Li[1], Gregory Tyson[1], Shaohua Zhao[1], Patrick McDermott[1], Errol Strain[1]

US Food and Drug Administration, Center for Veterinary Medicine, Office of Research, Division of Animal and Food Microbiology

Abstract

Introduction: Plasmids are mobile genetic elements and a driving factor in the spread of antimicrobial resistance. Tracking plasmid transmission among bacterial strains is key to identifying sources of antimicrobial resistance genes and developing strategies to combat spread of antimicrobial resistance. Specifically, IncA/C plasmids commonly carry genes that confer a multiple drug resistant (MDR) phenotype and have played a significant role in increasing the populations of MDR organisms in the environment. Methods to classify plasmids using sequencing data alignments often fail because plasmids undergo frequent recombination and deletion events. To address this, we developed a method to identify conserved sequence fragments comprised of multiple genomic elements among heterogeneous plasmid populations and demonstrated its use in characterizing IncA/C plasmids. **Methods:** Conserved loci of coding and intergenic regions from 151 closed plasmid sequences representing 26 plasmid types and 28 *Salmonella* serotypes were identified using Roary and Piggy. Conserved loci clusters were identified and the loci cluster corresponding to the 10 closed IncA/C plasmids was validated against an external PLSDB database containing 18,457 plasmid sequences. Locations of coding and intergenic regions in IncA/C plasmids were used to identify contiguous genomic sequences in the plasmid population. Co-localization of coding and intergenic regions among contiguous genomic sequences was evaluated using Pearson's correlation coefficient. Results of the correlation matrix were clustered in R and conserved fragments defined at a mean r-value threshold of >0.95. **Results:** Our analysis revealed 28 coding and 19 intergenic loci conserved in the pangenome of our closed IncA/C plasmids. This set of loci identified all 343 IncA/C plasmids from 18 genera in PLSDB. The 47 loci were located among 13 conserved plasmid DNA fragments. Of these, 11 contiguous fragments (271–6,452bp) had a prevalence of >80% in the overall plasmid population, contained 1-11 loci each and defined the core elements of IncA/C plasmids. **Conclusions:** Our analysis shows IncA/C plasmids share common genomic elements distributed in conserved arrangements among 11 locations on the plasmid. Identifying and characterizing these conserved regions will aid in identifying plasmid recombination events and facilitate developing a subtyping scheme to track the dissemination of antimicrobial resistance genes among bacterial populations.

Introduction

Plasmids are small genetic structures that are transferred between bacteria. Plasmids that encode antibiotic resistance genes can provide an antibiotic resistant phenotype to the host. Because of this, it is important to have a plasmid typing system to classify plasmids with different genetic features.

The traditional approach to plasmid typing organizes plasmids by their ability to replicate in the presence of other plasmids. Plasmid replication efficiency directly affects the number of copies of a plasmid in a cell, and this plasmid copy number is stable within a bacterial strain. As a result, if the copy number of a plasmid decreases when a new plasmid is acquired by the bacterium, the two plasmids are classified as being members of the same incompatibility group, or "Inc" group. The genetic basis for the Inc group is the replicon sequence, and this replicon is the target used for short sequence plasmid typing. The challenge of this typing method is its reliance on a single locus to characterize plasmids prone to recombination.

With the increased availability of long read sequence technologies, closed plasmid sequences are more accessible for research. These sequences allow for characterization of both the gene coding regions and the intergenic regions that make up the full plasmid sequence. In this project, we leveraged the benefits of long read sequencing to evaluate the sequence diversity of a plasmid population and identify genomic elements conserved among IncA/C plasmid subpopulations that serve as potential targets for plasmid typing.

Materials and Methods

Ninety-seven *Salmonella* isolates recovered from food samples were sequenced on a PacBio platform. SEQuel assemblies generated closed chromosomal and plasmid sequences, resulting in a library of 151 complete plasmid sequences which were then annotated with PROKKA. Pangenome of coding and intergenic regions of this plasmid library was determined using Roary and Piggy, respectively. A presence/absence matrix of the 151 plasmids and their corresponding plasmid loci was generated in R, and plasmids with similar loci profiles were clustered into groups. Plasmid Inc group was determined with PlasmidFinder, Loci corresponding to the 10 closed IncA/C plasmids were validated against an external PLSDB database containing 18,457 plasmid sequences. The subset of loci corresponding to known IncA/C plasmids in the PLSDB database was selected as the core identifying loci of the IncA/C plasmid group. To determine whether these loci were sporadically placed along the plasmid or if they were inherited in co-localized sequences, loci coordinates were used to stitch together neighboring coding and intergenic loci in R. This co-localization analysis was performed for every plasmid in the PLSDB database with IncA/C loci, and the prevalence of colocalized fragments in the plasmid population was evaluated using Pearson's correlation coefficient.

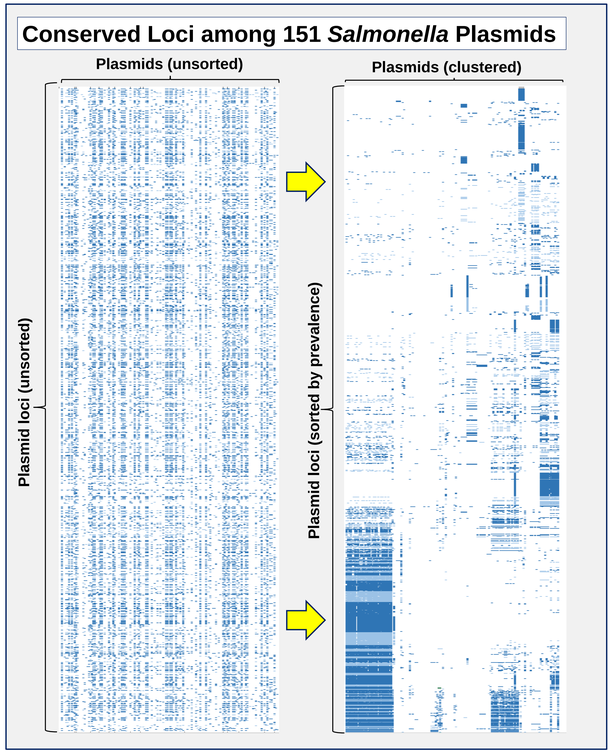


Figure 1. The pangenome of 151 Salmonella plasmids represented above in a presence/absence matrix. Coding regions are shown as dark blue and intergenic regions as light blue.

Results and Discussion

Sequence analysis revealed coding and intergenic regions were conserved in distinct patterns among the plasmid population (**Figure 1**). Following the determination of Inc type with the PlasmidFinder tool, the patterns of loci inheritance from Figure 1 corresponded to plasmid Inc types (**Figure 2**). Closer analysis revealed 28 coding and 19 intergenic loci conserved in our IncA/C plasmid pangenome. These loci matched all 343 IncA/C plasmids in 18 genera in the PLSDB database (**Figure 3**). These loci were located among 13 plasmid DNA fragments (**Figure 4**). Of these, 11 contiguous fragments (271–6,452bp, >80% prevalence) contained 1-11 loci each and defined the core elements of IncA/C plasmids (**Figure 4**).

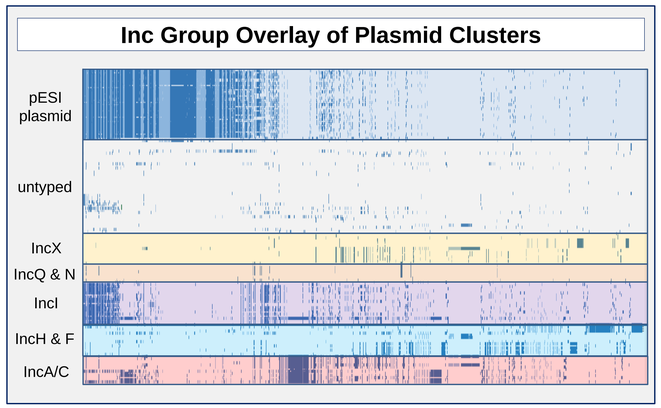


Figure 2. Plasmid loci profiles from Figure 1 were compared to traditional plasmid Inc groups. Inc group overlay reveals that many plasmid loci are conserved within Inc group. The pESI plasmid type is associated with a large plasmid common within an emergent *Salmonella* infantis strain.

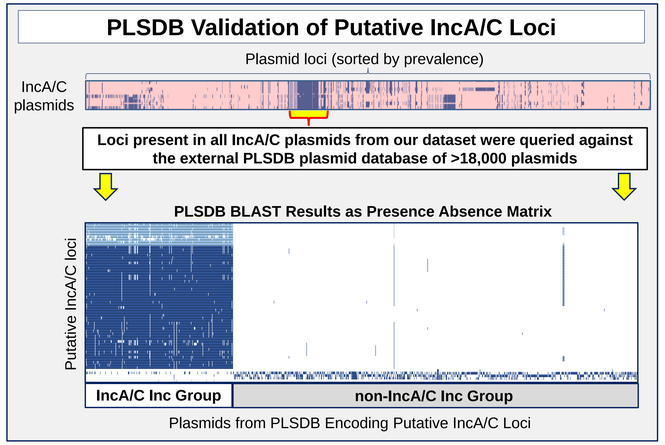
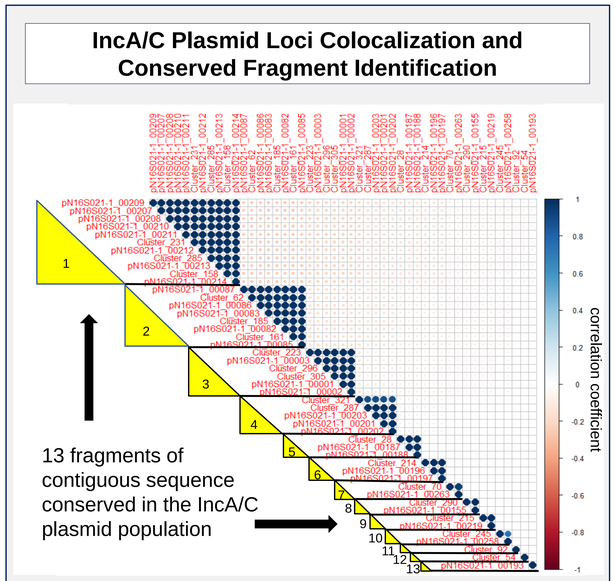


Figure 3. Results of querying IncA/C loci against the PLSDB database represented as a presence/absence matrix, sorted by plasmid type. The subset of coding (dark blue) & intergenic regions (light blue) define conserved IncA/C loci.

Figure 4. Colocalization of plasmid loci from Figure 3 reveals 13 fragments of contiguous plasmid loci (yellow triangles). The strength of correlation between two loci is represented by the color of the circle at each loci pairing, and the statistical significance is represented by the size of the circle. Gene products of plasmid loci and fragment prevalence in the IncA/C population are listed below. Fragments with grey headers have a prevalence <80%.



Fragment 1 - Prevalence: 89%	Fragment 3 Prevalence: 93%	Fragment 6 - Prevalence: 91%
Intergenic loci (3x) hypothetical protein (3x) SNase-like nuclease DNA-binding protein plasmid-partition protein Cro-like protein HigB2	Intergenic loci (3x) hypothetical protein Nucleoid occlusion protein methyltransferase	Intergenic locus hypothetical protein (2x)
	Fragment 4 - Prevalence: 81%	Fragment 7 - Prevalence: 88%
	Intergenic locus Clp-like protease disulfide isomerase lipoprotein	Intergenic locus hypothetical protein
Fragment 2 - Prevalence: 85%		Fragment 8 - Prevalence: 86%
Intergenic loci (3x) Pilus - TraF Helicase Partitioning protein DNA binding protein hypothetical protein	Fragment 5 - Prevalence: 96%	Intergenic locus hypothetical protein
	Intergenic locus hypothetical protein rod-shape determination-protein	Fragment 9 - Prevalence: 47%
		Intergenic locus dihydropteroate synthase
		Fragment 10 - Prevalence: 14%
		Intergenic locus hypothetical protein
Fragment 11 - Prevalence: 88%	Fragment 12 - Prevalence: 95%	Fragment 13 - Prevalence: 94%
Intergenic locus	Intergenic locus	hypothetical protein

Conclusion

In this project, we have shown that there are multiple common genetic elements present in the IncA/C plasmid population conserved in discrete arrangements among 11 locations on each plasmid. We have also shown that while these elements are present on most IncA/C plasmids, they are not present on every plasmid. Therefore, these may serve as markers in identifying plasmid recombination events and facilitate developing a subtyping scheme to track the dissemination of antimicrobial resistance genes among bacterial populations.