2019

Assessing the prevalence of multidrug resistant Salmonella enterica from stream sediment and poultry litter in the Shenandoah Valley of Virginia

Noah Greenman
James Madison University

Assessing the Prevalence of Multidrug Resistant *Salmonella enterica* from Stream Sediment and Poultry Litter in the Shenandoah Valley of Virginia

Noah A. Greenman

A thesis submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Master of Science

Department of Biology

May 2019

FACULTY COMMITTEE:

Committee Chair: James Herrick, Ph.D.

Committee Members:

Ray Enke, Ph.D.

Louie Wurch, Ph.D.

Acknowledgements

Dr. James Herrick has been my academic advisor for the past 2 years. Under his tutelage I have learned so much more than I thought previously possible. It has been an honor and a privilege to learn from such a phenomenal researcher with so much passion for his research and the students under him. I will be forever grateful for his mentorship through the completion of this thesis and for shaping me into the scientist I am today.

I also thank my committee members Dr. Louie Wurch and Dr. Ray Enke who have offered invaluable feedback and helped to create and execute a project that I can be proud of. I would also like to thank Dr. Steven Cresawn for his help in assisting with the sequencing aspects of my thesis. I am grateful to past students Curtis Kapsak and Kevin Libuit, both of whom were instrumental in the setup of both wet lab and sequencing techniques, being great mentors, and who are friends and colleagues to this day. I am grateful for Sophie Jurgensen and Charles Holmes who first got this project on its feet and provided early data this project used.

Table of Contents

List of Tables

List of Figures

Abstract

Multidrug resistant *Salmonella enterica* present in stream sediment and poultry litter represent a critical health concern. A small number of *S. enterica* serotypes are responsible for most lab-confirmed infections in the US each year. To assess the prevalence of these significant strains, we isolated 88 *S. enterica* from stream sediment and poultry litter. Sequence data for all isolates were generated using an Illumina® sequencing platform, with long-reads for some isolates from the Oxford Nanopore MinION™ used in a hybrid genome assembly approach. Isolates were typed according to their serotype and multi-locus sequence type using SeqSero/SISTR and Enterobase respectively. Antibiotic resistance genes were annotated using ABRicate and Prokka. Thirty-one isolates possessed one or more antibiotic resistance genes, with resistance genes located exclusively on plasmids identified by MOB_Suite in 26 of those isolates. Eight of the 26 isolates with plasmids containing antibiotic resistance genes displayed phenotypic resistance to multiple antibiotics. Multiple plasmids were found to contain *tra* and/or *pil* gene cassettes, implicating them as conjugative in solid and/or liquid mediums. Septic *Salmonella* infections require antibiotic intervention, and the existence of multiple antibiotic resistance genes on transmissible plasmids in *Salmonella* isolated from streams and litter may indicate that a significant reservoir for transmissible resistance occurs in these environments. Infections with multidrug resistant *Salmonella* be difficult to treat, and plasmid-borne resistance may be transmissible to other, potentially even more pathogenic bacteria in these environments.

Introduction

The bacterium *Salmonella enterica* is one of the leading causes of foodborne illness in the world causing an estimated 40,000 cases annually in the U.S alone, with many more going unreported (Fábrega and Vila, 2013). In 2013, the total cost of *Salmonella*-related foodborne illnesses in the U.S was approximately $3.66 billion (Hoffmann S, 2017). While in healthy adults *Salmonella* is usually self-limiting, a major concern is when a person requires antibiotics to treat more resilient infections. The misuse by humans of antibiotics in healthcare, agricultural, and industrial settings has led to the evolution of antibiotic resistant *S. enterica*. Although we know of this trend, our understanding of where this selective evolution occurs and how it is occurring is still limited.

Salmonella enterica

Classification and morphology

Salmonella enterica is a gram-negative, facultatively anaerobic bacillus that is in the family *Enterobacteriaceae* (Baron, 1996; Coburn *et al.*, 2007). *S. enterica* is one of two recognized species of *Salmonella*, the other being *S. bongori* (Acheson *et al.*, 2001; Andino and Hanning, 2015). *S. enterica* is further subdivided into six distinct subspecies: enterica, salamae, arizonae, diarizonae, houtenae, and indica (Fierer and Guiney, 2001). Subspecies are further subdivided into what are called serotypes or serovars. Over 2500 serotypes have been identified; each defined by the antigens it presents (Coburn *et al.*, 2007). Differences in the antigens occur within the flagellar proteins (H antigen), the sugars that make up part

of the outer cell membrane (O antigen), and the lipopolysaccharides of the cell wall (Baron, 1996; Shipp and Rowe, 1980). Yet another level of classifying *S. enterica* is the definitive phage type (DT). As the name implies, subspecies and serotypes can be further classified by the phages that infect them (Anderson *et al.,* 1977). This typing scheme adds another layer of complexity to identifying *S. enterica*.

Important serotypes

Despite the ever-growing number of *S. enterica* serotypes, only a fraction of these have been implicated in causing disease in people. While most serotypes can cause enterocolitis/diarrhea, a few are the causative agents of enteric typhoid fever. Three of the major serotypes of this group are Typhi, Paratyphi, and Sendai (Coburn *et a.,* 2007; Maloy and Edwards, 1999). When large studies on *S. enterica* outbreaks are done, four serotypes often make up over half of the cases (Brown *et al.,* 2017; Jackson *et al.,* 2013; Jones *et al.,* 2008; Maloy and Edwards, 1999). Those serotypes are Typhimurium, Enteritidis, Newport, and Heidelberg. Which serotype causes the most outbreaks changes over time, but these four serotypes are the ones that most often cause outbreaks (Jones *et al.,* 2008). These serotypes have a variety of reservoirs and have been implicated in outbreaks caused by distinctly different sources (Jackson *et al.,* 2013). These factors help explain the high prevalence outbreaks and the diversity of sources that cause them.

S. enterica serotype Typhimurium has numerous factors that confer increased sustainability, virulence, and surivability. The diversity of environments and hosts

Typhimurium has adapted to is the reason it is such a successful pathogen and why it is implicated in numerous outbreaks. While other serotypes come from mainly one source, Typhimurium also has been linked to numerous sources (Jackson *et al.,* 2013). These include chicken, beef, and pork, as well as plants like lettuce (Dechet *et al.*, 2006; Horby *et al.,* 2003; Jackson *et al.,* 2013). Typhimurium has a variety of hosts, including humans, cattle, pigs, birds, fruits and vegetables (Jackson *et al.,* 2013; Rabsch *et al.,* 2002). Typhimurium has a few specialized features. *Salmonella enterica* contain clusters of genes called Salmonella Pathogenicity Islands (SPIs). These islands will be addressed more later, but one key aspect is that Typhimurium has 5 important SPIs that confer increased virulence while most other strains possess 2 (Fábrega and Vila, 2013). One strain of Typhimurium, DT104, is of major concern due to multiple drug resistance genes it possesses and the sources it has come from, including cattle, poultry, and other livestock (Boyd *et al.,* 2001; Dechet *et al.,* 2006; Horby *et al.,* 2003; Mølbak *et al.,* 1999). Typhimurium also has greater survivability when the intestine is inflamed due to its resistance to the secreted endogenous human antimicrobial peptide Lipocalin-2 (Raffatellu *et al.,* 2009). While the beneficial microbes die, Typhimurium survives and gains easy access to more epithelial cells it can now infect (Raffatellu *et al.,* 2009).

When compared to Typhimurium, Enteritidis has similarities and differences. An interesting trend throughout reported outbreaks is that Enteritidis and Typhimurium tend to cause the largest percentage of outbreaks during different time periods (Jackson *et al.,* 2013; Rabsch *et al.,* 2002). Enteritidis has a narrower host range when compared to Typhimurium. Enteritidis is most often found in poultry sources like chicken, turkey, and eggs, and it is able to persist in its avian hosts for long time periods (Jackson *et al.,* 2013;

Braden, 2006). One study reported that Enteritidis was isolated from the guts of houseflies (*Musca domestica*), providing an example of how it could be distributed to other environments and maintained in ones treated with antimicrobial compounds (Holt *et al.,* 2005). Enteritidis also possesses distinct virulence factors, as well as factors commonly found in all major serotypes including antibiotic resistance (Mølbak *et al.,* 1999; Cheung *et al.,* 2005). One critical virulence factor is the BapA protein. This protein is necessary for biofilm formation and host colonization (Latasa *et al.,* 2005). Biofilms provide protection from immune responses so Enteritidis can survive and the infection can persist. Enteritidis has been reported to be more invasive than most other serotypes. This leads to more cases of bacteremia rather than just strictly enterocolitis, and also highlights Enteritidis's survivability (Phiri *et al.,* 2008).

The last two significant serotypes, Newport and Heidelberg, are not as prevalent as Typhimurium and Enteritidis, but they still cause a large number of outbreaks worldwide (Egorova *et al.,* 2008; Hoffmann *et al.,* 2014; Jackson *et al.,* 2013). Newport is like Typhimurium in the broad range of hosts it is associated with, but food sources most often implicated are those derived from cattle (CDC, 2002). Other major outbreaks of Newport have also been linked to fresh produce (Van Beneden, 1999; Sivapalasingam *et al.,* 2003). Like Enteritidis, serotype Heidelberg is most often linked to poultry sources (Hoffmann *et al.,* 2014; Jackson *et al.,* 2013). Like the other significant serotypes, Newport, and Heidelberg possess several virulence factors. The factor differentiating them from each other, and from other important serotypes are their resistance profiles. Both Newport and Heidelberg isolates have been found with multi-drug resistance. In recent years, serotype Newport isolates have been found to be resistant to ceftriaxone, but Heidelberg tends to

still be sensitive to this antimicrobial agent (Egorova *et al.,* 2008). Heidelberg isolates are more commonly found to be resistant to ceftiofur (Dutil *et al.,* 2010; Zhao *et al.,* 2008). Like Typhimurium and Enteritidis, Newport and Heidelberg will also often be resistant to other antimicrobial agents like tetracycline and streptomycin (Hoffmann *et al.,* 2014; Zhao *et al.,* 2003; Zhao *et al.,* 2008).

Key genetic features

Salmonella enterica possess numerous important genetic features that allow for increased virulence, pathogenicity, and survival within the host. These features include *Salmonella* pathogenicity islands, *Salmonella* plasmid virulence genes, islets, single genes, and plasmids. *Salmonella* pathogenicity islands, or SPIs, are DNA segments of the chromosome which carry key virulence genes (Fierer and Guiney, 2001). The best characterized SPIs are SPI-1 and SPI-2, with more information on SPI-3, 4, and 5 starting to come out. As of 2007, 15 SPIs had been identified, however very little is known outside of the first 5 (Cook *et al.,* 2007). SPI-1 is found at centisome 63 and codes for a Type III secretion system (Galán, 1999). This secretion system is critical for initiation of host cell invasion and adherence, also having a minor role in effector molecule secretion (Coburn *et al.,* 2007). The importance of a functional SPI-1 for virulence has been established by studying environmental *Salmonella* that do not have a functioning SPI-1 (Ginocchio *et al.,* 1997). *Salmonella* with a non-functional SPI-1 are unable to initiate host cell invasion, so it is probable that all infectious serotypes of *Salmonella* contain a functional SPI-1. SPI-2 genes primarily code for another Type III secretion system that is responsible for secreting

the majority of effector molecules that control the internal processes of host cells (Figueira and Holden, 2012). There are nearly 30 effectors that have been characterized, but only 2-3 genes encoding some of these effectors have been identified (Figueira and Holden, 2012). The function of the genes in SPI-2 is essentially to sequester nutrients and other necessary molecules from the host cell while also avoiding host cell defense responses (Coburn *et al.,* 2007; Figueira and Holden, 2012). Certain SPI-2 Type III secretion systems may also work in reverse to secrete other effectors from other genes in order to knock down immune signaling (Figueira and Holden, 2012). *S. enterica* serotype Typhimurium's SPI-2 allows it to avoid NADPH oxidase when it infects macrophages protecting it from death (Coburn *et al.,* 2007). SPIs 3, 4, and 5 are not as well studied, but what functions have been discerned and hypothesized are all linked to increasing survivability and increasing pathogenicity. SPI-3 and SPI-4 have similar roles in aiding *Salmonella's* survival within macrophages as well as adhering to host cells like SPI-1 (Fábrega and Vila, 2013). SPI-3 is found at the *selC* locus and contains genes necessary for replication within macrophages (Blanc-Potard *et al.,* 1997). SPI-4 has been implicated in allowing *S. enterica* species to adapt to the altered, hostile environment of macrophages. Such adaptations are likely necessary to repair damage by the macrophage, adapt to an environment with a different pH, and improve nutrient sequestration (Bäumler *et al.,* 1994). SPI-4 may also encode cytotoxic effector molecules design to induce apoptosis in immune cells and thus survive the onslaught of antimicrobial agents (Wong *et al.*, 1998). SPI-5 is responsible for increased pathogenicity. The effectors encoded within this gene cluster mediate inflammatory responses and fluid secretions by translocation into the intestinal lumen (Marcus *et al.,* 2000; Wood *et al.,* 2000). Because *S. enterica* species have resistances to host antimicrobial

agents like Lipocalin-2, it is possible that this mediation of inflammation and fluid secretion may be a strategy to compete with other microbes. These effectors encoded by other SPIs likely use the Type III secretion systems of SPI-1 and SPI-2 to be secreted to their target environments because no secretion system has been identified among other SPIs (Marcus *et al.,* 2000).

Another genetic feature that all clinically relevant serotypes share are *Salmonella* plasmid virulence genes, now called *spv* (Fábrega and Vila, 2013; Fierer and Guiney, 2001). The five genes of the *spv* locus are *spvR, A, B, C,* and *spvD* (Fierer and Guiney, 2001). So far, the function of *spvR, spvB,* and *spvC* have been determined (Fábrega and Vila, 2013). *SpvR* is the most critical gene as it serves as the transcriptional activator for operon *spvRABCD*. Studies of *Salmonella* isolates with mutations in this gene have found that infections rarely persist, but initiation of infection still occurs (Gulig *et al.,* 1993; Fierer and Guiney, 2001; Libby *et al.,* 2000; Pesold *et al.,* 2002). This highlights the importance of *spv* genes for persistence of infection, but not initiation. *Salmonella* may possess other plasmids containing other virulence genes such as antibiotic resistance genes. This topic will be explored later on when antibiotic resistance is addressed.

Other small genetic loci referred to as islets as well as individual genes do play a role in *Salmonella* virulence (Fierer and Guiney, 2001). These roles tend to be specific to the serotype that contains such genes (den Bakker *et al.,* 2011). Some islets contain resistance genes to antibiotics and other islets contain a gene or genes that assist with host cell invasion (Di Conza *et al.,* 2002; Fierer and Guiney, 2001). Such islets and individual genes are often hard to characterize generally because individual populations of serotypes

may evolve to have different virulence factors depending on the environment they exist in (den Bakker *et al.,* 2011; Fierer and Guiney, 2001).

Pathogenesis

Infection by *Salmonella enterica* follows a consistent progression regardless of the strain. Infection with any *Salmonella* strain begins by ingestion of the organism (Fábrega and Vila, 2013; Ohl and Miller, 2001). The first hurdle *Salmonella* must surmount is the acidity of the stomach. One method for tolerating the high acidity is an Acid Tolerance Response (ATR) which allows *Salmonella* to maintain the pH of its internal environment in the presence of an extremely different extracellular pH (Fábrega and Vila, 2013; Foster and Hall, 1991; Ohl and Miller, 2001). Upon reaching the small intestine, a cell must then wait to contact the intestinal epithelia in order to adhere (Fábrega and Vila, 2013). *Salmonella* can infect numerous cell types, but the preference is to target M cells of Peyer's patches (Takeuchi *et al.,* 1967). Shortly after adhering, the *Salmonella* then uses a Type III secretion system to induce micropinocytosis (Coburn *et al.,* 2007; Francis *et al.,* 1993). After the *Salmonella* has entered the host cell it forms a *Salmonella*-containing vacuole (SCV) (Fábrega and Vila, 2013). To prevent the induction of lysosomes into the SCV *Salmonella* uses effector molecules (Rathman *et al.,* 1997). While *Salmonella* cells continue to replicate within the SCV other effectors are secreted outside the host cell to recruit other phagocytes for further dissemination of *Salmonella* through the body (Deiwick *et al.,* 2006; Fábrega and Vila, 2013). For replication to occur it is essential for the maturing SCV to migrate to a location where nutrient acquisition can be facilitated

easily (Deiwick *et al.*, 2006; Fábrega and Vila, 2013). Upon maturation the new *Salmonella* return to the intestinal epithelium by lysing the host cell where they are engulfed by phagocytes such as macrophages, neutrophils, and dendritic cells (Fábrega and Vila, 2013). As with other cell types, the uptake of *Salmonella* by a macrophage occurs via induced micropinocytosis (Alpuche-Aranda *et al.,* 1994). The intracellular environment of a macrophage is more hostile than that of intestinal epithelial cells, so *Salmonella* has evolved mechanisms to neutralize the environment and make it more suited for replication (Alpuche-Aranda *et al.,* 1994; Bäumler *et al.,* 1994; Blanc-Potard *et al.,* 1997). From here, the *Salmonella* can persist systemically and may infect other tissues (Alpuche-Aranda *et al.,* 1994; Fábrega and Vila, 2013; Ohl and Miller, 2001). The mechanisms by which different serotypes infect and survive within a host varies. Furthermore, there is no definitive explanation for why different serotypes have certain distinct pathogenic features, nor is there an explanation for how so few serotypes cause the majority of diseases (Fierer and Guiney, 2001; Jones *et al.,* 2008). Some serotypes like Enteritidis have resistances to bodily antimicrobial agents through biofilm formation (Latasa *et al.,* 2005). Typhimurium is able to resist the antimicrobial peptide Lipocalin-2 that is secreted by the immune system in response to infections (Raffatellu *et al.,* 2009). Different serotypes are also linked to varying degrees of disease severity. Serotypes Typhimurium and Enteritidis are noted as being more invasive than other known serotypes and are implicated in more cases of bacteremia (Phiri *et al.,* 2008). Such distinct phenotypic features highlight the underlying complexity of each individual serotype's pathogenic capabilities and the need to explore them further.

Antibiotic resistance in the environment

Antibiotic Resistance

Antibiotic resistance is not a new problem but is one that continues to grow. Resistance is a process that happens rapidly, but often disappears slowly (Levy and Bonnie, 2004). Currently, we have yet to find or create an antibiotic that bacteria are unable to develop resistance towards (Ainsa, 2002; Frye and Jackson, 2013; Levy and Bonnie, 2004). Strains of drug-resistant *Salmonella* have been identified as far back as the 1950's and new studies continue to further describe multi-drug resistant strains (Levy, 2001; Mather *et al.,* 2013). With resistance increasing in *Salmonella* and our options decreasing we are rapidly running out of time. In order to understand and start to tackle antibiotic resistance we must understand what antibiotics *Salmonella* is resistant to, how it acquires resistance and where resistance develops.

Salmonella infections are most often treated with antibiotics from a select few antibiotic families, each designed to target a key physiological process. Cephalosporins and carbapenems inhibit cell wall synthesis (Levy and Bonnie, 2004). Tetracyclines and aminoglycosides inhibit protein synthesis (Amin *et al.,* 1996, Davis, 1987). Fluoroquinolones are the only family that inhibits DNA synthesis (Wolfon *et al.,* 1985). Sulfonamides and trimethoprim inhibit folic acid synthesis (Briganti *et al.,* 1996). Rifampin is the only antibiotic family that inhibits RNA synthesis (Wehrli, 1983). Because of the selectivity and limited number of treatment options for treating *Salmonella* infections, the presence of resistance to any of the listed antibiotics is of great concern. Unfortunately, *Salmonellae* have been found with resistance to almost every

major antibiotic family used to treat it (Akiyama *et al.,* 2013; Baucheron *et al.,* 2002; Diarra *et al.,* 2014; Leekitcharoenphon *et al.,* 2016; Mąka and Popowska, 2016; Noda *et al.,* 2015; Su *et al.,* 2012).

Horizontal gene transfer is the mechanism by which *Salmonella* and most bacteria acquire resistance genes (Jain *et al.*, 1999). While spontaneous mutations can result in resistance, it is becoming clear that mobile elements play a significant role in the spread of antibiotic resistance (Nakamura *et al.,* 2004; Thomas and Nielsen, 2005). There are a number of ways *Salmonella* and other bacteria acquire genes via horizontal gene transfer. The methods of gene movement focused on are plasmids, integrons, transposons, and genomic islands. Plasmids are circular pieces of DNA that can be transferred between bacteria (Helinski, 2004). When a plasmid enters a bacterium, it does not need to integrate into the chromosome for a bacterium to access its genes. However, some plasmid-borne genes are part of mobile elements that allow them to jump into and out of the bacterial chromosome. Transposons allow individual genes to move from a plasmid to the chromosome, from a chromosome to a plasmid, and from one plasmid to another (Carattoli, 2003). Transposons code for the transposase protein that allows these genes to move into a genome. Collections of similar genes, called gene cassettes, can be found within integrons (Carattoli, 2003). Integrons code for a protein called an integrase which is essential for the excision and insertion of gene cassettes (Tosini *et al.,* 1998). Both integrons and transposons can originate from a plasmid or a bacterial genome, but those in a chromosome typically need to be mobilized into plasmids for actual movement (Tosini *et al.,* 1998). Genomic islands are large, potentially mobile pieces of DNA that can encode numerous genes of varying biological function. The most important in

Salmonella are the SPIs which can be mobilized and transferred between other *Salmonella* and are critical for pathogenicity (Levings *et al.,* 2005). Lastly, bacteriophages are also important contributors to the development of both resistance and virulence in *Salmonella*. P22-like phages have been found to transfer genes conferring resistance to sulfanomides, tetracycline, chloramphenicol, and other antibiotic resistance genes to *S. enterica* serotype Typhimurium DT104 (Schmieger and Schicklmair, 1999). Certain effector molecules secreted by the Type III secretion systems of *Salmonella* originated from phages (Ehrbar and Hardt, 2010; Ho *et al.*, 2002).

In *Salmonella*, plasmids are one of the primary vehicles for moving important genes such as those encoding virulence factors and antibiotic resistance. Some plasmids, like those found by Gay et al., contain only one resistance gene (Gay *et al.,* 2006). *S. enterica* serotype Typhimurium DT104 has been found to contain multi-drug resistance-gene-containing plasmids (Briggs *et al.,* 1999; Tosini *et al.,* 1998). Originally these resistance genes were not thought to be mobile, but now gene cassettes are being identified on major plasmids like IncFI, IncL/M, and IncFII (McCollister *et al.,* 2016; Tosini *et al.,* 1998). Recent studies have identified *Salmonella* isolates from food and human sources containing a plasmid-mediated *mcr-1* resistance gene (Doumith *et al.,* 2016; Lekunberri *et al.,* 2017). This gene confers resistance to colistin, an antibiotic of last resort (Lekunberri *et al.,* 2017). Its presence in *S. enterica*, especially in important serotypes like Typhimurium, further emphasizes the ever-growing problem of the spread of antibiotic resistance (Doumith *et al.,* 2016; Lekunberri *et al.,* 2017).

Environments linked to the persistence of antibiotic resistance genes

The environments that horizontal gene transfer takes place in are diverse (Baquero *et al.,* 2008). There are a range of environments that are of interest concerning the increasing prevalence of antibiotic resistance. Many studies report on antibiotic resistance development in clinical environments such as hospitals, however little attention has been paid to nonclinical environments (Martínez, 2008). Streams, fields fertilized with manure, poultry litter, and other external natural environments are proposed source of most antibiotic resistance development, citing rapid dissemination of resistance (Finley *et al.,* 2013; Nesme *et al.,* 2014). Along with development of antibiotic resistance, both natural and clinical environments are known as antibiotic resistance reservoirs. Understanding how water sources, soil, animals, and sewage treatment plants serve to maintain antibiotic resistance genes (ARGs) and allow for exchange of ARGs is critical for potentially disrupting the emergence of more multidrug resistant bacteria.

Water environments of particular interest are sewage treatment plants (STPs), effluents, groundwater and surface water as they can be linked back to people (Hirsch *et al.,* 1999). These water-based environments are hotbeds for antibiotic transfer and the maintenance of antibiotic resistant populations (Cabello, 2006). These sources often feed back into water sources such as drinking water and other distribution systems (Xi *et al.,* 2009; Xu *et al.,* 2015). STPs are effectively points of collection for low levels of antibiotics. STPs receive the runoff from rivers and other water sources and then deliver it back supposedly treated. STPs are not perfect, as seen by the persistence of low levels of antibiotics in STP effluent (Naquin *et al.,* 2015; Yang *et al.,* 2014). Bacteria present in

STPs are often heavy metal resistant and these heavy metal resistance genes are often accompanied by antibiotic resistance genes (Calomiris *et al.*, 1984; Rajbanshi, 2009). This linkage is also seen between antibiotics as well. It is more common for multiple antibiotic resistance genes to be present together than apart (Herrick *et al.*, 2014). As with STPs, other water-based environments are implicated in being reservoirs for antibiotic resistance genes (Amos *et al.*, 2014; Ash *et al.*, 2002; Cabello, 2006; Hirsch *et al.*, 1999; Goñi-Urriza *et al.*, 2000). Studies identifying the presence of *Salmonella* and antibiotic resistance genes in streams typically involve analyzing the stream water and/or stream sediment. Streams are not typically chosen randomly, but rather are selected due to their proximity to agricultural activity, urban effluent, or other source of contamination (Haack et al., 2015; Pei et al., 2006).

Soil is also a common reservoir for antibiotic resistance genes. Depending on the source of contamination, different types of antibiotic resistance compared to water sources may be present (Esiobu *et al.*, 2002). Antibiotic contamination of soil is usually related to agricultural or industrial sources like farms and factories. It is a common practice to use low levels of antibiotics to treat livestock prophylactically and promote growth (Mellon et al., 2001). As with humans, use of antibiotics selects for antibiotic resistant bacteria, which enter the environment through feces, improper disposal of antibiotics, and other sources (Mellon et al., 2001; Zhu *et al.*, 2013). These bacteria then enter the soil and other environments like water sources from manure, litter, and other sources (Apata, 2009; Ghosh *et al.*, 2007; Gilliver *et al.*, 1999; Heuer *et al.*, 2011).

Animals themselves are often reservoirs for antibiotic resistance genes as well. This includes farm animals and wild animals. Wild animals are important as they provide

ways for antibiotic resistance genes to spread to other environments. Rodents have been found to carry *Salmonella* (Gilliver *et al.,* 1999). Commensal *E. coli* are a primary example of how antibiotic resistance genes can persist in the bodies of humans and animals and be transferred to *S. enterica* and other pathogenic bacteria (Bailey *et al.,* 2010; Van *et al.,* 2012). Diarrassouba *et al.* report on how commensal *E. coli* in broiler chickens carry ARGs shared by *Salmonellae* also identified in the same broiler chickens (Diarrassouba *et al.,* 2007) This co-occurrence has also been seen in cattle and humans as well (DeFrancesco *et al.,* 2003 and Winokur *et al.,* 2001).

Techniques for studying *Salmonella* and antibiotic resistance

There are many tools and techniques that allow researchers to study antibiotic resistance and its relationship with *Salmonella*. Certain techniques and tools are standards that practically all researchers use, but how data is utilized and what analyses are performed can vary. With the rise of whole-genome sequencing (WGS) and more advanced sequencing technologies large data sets are increasing rapidly in both number and size. These large data sets can be utilized in numerous ways, providing useful data depending on the analytical techniques used. The focus of this review will be on techniques used to collect and analyze data on *Salmonella* and antibiotic resistance.

Isolation

Salmonella isolates are primarily obtained by enrichment after collecting samples directly from the environment. Sampling techniques vary depending on the source. Samples collected *in situ* are placed directly into containers such as falcon tubes or other containers that are typically sterile (Kingston, 1981). Samples may include poultry litter (a combination of chicken feces, bedding, feathers, and spilled feed), manure, water, water sediment, and others (University of Kentucky School of Agriculture, 2014). Once samples are obtained, a pre-enrichment in buffered peptone water (BPW) is performed to recover as many bacteria as possible including sub-lethally injured isolates (Edel and Kampelmacher, 1969). Culturing is typically done first with selective media such as tetrathionate broth, selenite-cysteine broth, or Rappaport-Vassiliadis broth (Jorgensen *et al.*, 2002; Kinde *et al.*, 2004; Stone *et al.*, 1994). Selective broths are useful for isolating and improving recovery of only *Salmonella*, but they prevent any quantification of original levels of *Salmonella* present in the sample. Isolation of individual colonies is done on selective media, including Brilliant Green with Novobiocin (BGN), xylose-lysine-tergitol 4 (XLT4), xylose-lysine-decarboxylase (XLD), bismuth sulfite (BS) agar, and CHROMagar™ (Edel and Kampelmacher, 1969, Jorgensen *et al.*, 2002; Kinde *et al.*, 2004; Perez *et al.*, 2003; Stone *et al.*, 1994).

Genomic analysis techniques

When examining techniques that provide substantial amounts of data, next-generation sequencing (NGS) is the leading method. NGS is referred to as high-throughput, meaning it can perform simultaneous parallel sequencing reactions on a massive scale. The sensitivity of these techniques allows for the calling of bases in a sequence with minimal errors, thus providing accurate data for analysis. NGS comes in two forms: Short-read sequencing (SRS) and long-read sequencing (LRS) (Goodwin *et al.*, 2016). Currently SRS is primarily done using sequencing by synthesis (Metzker *et al.*, 2010). Illumina® technologies including the MiniSeq™ are DNA-polymerase-dependent and sequencing comes from bases being identified during the synthesis process (Metzker *et al.*, 2010). The primary method for LRS is known as single-molecule real-time sequencing (Laver *et al.*, 2015). Single-molecule approaches use either a fixed DNA polymerase or a special camera to capture labelled bases as they are incorporated or a protein pore has single-stranded DNA fed through it and the bases are detected on that strand (Clarke *et al.*, 2009; Eid *et al.*, 2009). This protein pore-based method is used by the Oxford Nanopore Technologies MinION™ sequencer. This feature, along with a non-PCR based form of sequencing, differentiates the MinION™ from Illumina® sequencers and so the MinION™ is referred to as a 3rd generation sequencing technology and not a NGS technology like Illumina® sequencers. All NGS techniques have their advantages and disadvantages concerning cost, error rate, quantity and quality of data produced (Glenn, 2011). Illumina® short-read sequencers provide highly accurate data compared to the MinION™. However, the MinION™ does not require PCR amplification which lessens the risk of contamination affecting sequence data (Laver *et al.*, 2015). Recently it has

been observed that using short and long read data in a hybrid assembly yields more accurate assemblies than just utilizing one set of data (Goodwin *et al.,* 2015). A benefit of NGS in general is the rise of Whole Genome Sequencing (WGS). NGS provides sufficiently accurate and comprehensive data allowing WGS to be applied in many areas of study, including *Salmonella* and antibiotic resistance. An example of this is being able to compare whole-sequenced genomes of *Salmonella* to each other and the reference genome of non-typhoidal *Salmonella* (McClelland, 2001).

Applying traditional and new techniques to *Salmonella* and antibiotic resistance

Serotyping *S. enterica* subspecies is one of the most common methods for identifying an unknown *Salmonella* isolate. Traditional techniques are being used in conjunction with NGS to identify isolates rapidly and accurately. The gold standard of serotyping involves agglutination tests (Shipp and Rowe, 1980). Agglutination tests use antisera mixed with pure *Salmonella* isolates to identify the serotype based on whether cells agglutinate; in other words, the cells collect together (Shipp and Rowe, 1980). Agglutination tests are both time consuming and labor intensive, leading scientists to pursue WGS for rapid and effective serotyping (Wain *et al.,* 2013; Yachison *et al.,* 2017). Whole genome sequence data is utilized by *in* silico PCR techniques to determine the serotype of *Salmonella* isolates. *In silico* techniques like SeqSero and the *Salmonella In Silico* Typing Resource (SISTR) both use databases of sequenced *Salmonella* genomes, taking the H and O antigen sequences and using them for comparison to unknown *Salmonella* isolates (Yoshida *et al.,* 2016; Zhang *et al.,* 2015). Multiplex PCR is another

lab-based form of serotyping. It differs from traditional agglutination testing in that primers for multiple serotypes can be tested at once versus agglutination testing has only one antibody against one serotype per well (Alvarez *et al.,* 2004). Using databases allows for more rapid identification when compared to traditional or modern lab-based serotyping methods. Of the three approaches, *in silico* techniques are the most promising complements to standard agglutination tests.

Multi-locus sequence typing (MLST) is another typing method that could replace or complement serotyping. Along with identification, MLST is also used for assessing phylogenetic relationships. Standard MLST uses the sequences of seven housekeeping genes for comparing *Salmonella* isolates (Achtman *et al.,* 2012). Variations of MLST used to study *Salmonella* involve changing what loci are used to identify an organism. Ribosomal MLST (rMLST) compares and identifies isolates based on variations in ribosome-encoding genes, core genome MLST (cgMLST) uses a set of genes that make the core genome, and whole genome MLST uses the whole genome (Haley *et al.,* 2016; Mohammed *et al.,* 2017; Toro *et al.,* 2016). These various types of MLST allow for greater selectivity when identifying and differentiating *Salmonellae* isolates (Achtman *et al.,* 2012).

WGS produces large amounts of data for analysis tools such as ABRicate and MOB_Suite to identify ARGs and plasmids. These tools are often openly accessible and can be either downloaded and used through command line or accessed on hosted websites such as Galaxy (Afgan *et al.,* 2018) and GalaxyTrakr (https://galaxytrakr.org). Both tools utilize a similar approach to analyze sequence data for different features. ABRicate (Seemann, 2016) utilizes curated databases of known ARGs to compare potential ARGs

from an assembled genome to. ABRicate can even access multiple databases such as NCBI's ARG database and the ARG-ANNOT database (Gupta *et al.,* 2014). More recently, bioinformatics tools have come out to aid in identification of plasmids from NGS data. MOB_Suite (Robertson and Nash, 2018) uses a curated database of complete plasmid sequences to use for reference as the tool parses sequence data for potential plasmid sequences. One of the key benefits of bioinformatic tools is they can be updated and improved, allowing for continued development of discoveries from large datasets including NGS data.

WGS has the potential to aid in surveying and tracking *S. enterica* outbreaks and evolution worldwide. WGS provides in depth comparisons of *Salmonella* isolated from patients during outbreaks, allowing for a faster response (Leekitcharoenphon *et al.,* 2014). Epidemiologically, WGS allows researchers and health-related institutions to track *S. enterica* outbreaks (Deng *et al.,* 2012). Scientists can even track specific serotypes or track global distribution of serotypes (Hendriksen *et al.,* 2011). Surveillance of *S. enterica* through analysis of WGS data is currently being employed by the CDC and other organizations (Leekitcharoenphon *et al.,* 2014).

Despite the technological advances, the cornucopia of data available and the ever-increasing focus on slowing the spread of antibiotic resistance, *Salmonella*-related illness continues to be a worldwide issue. *S. enterica* is well understood pathogenically speaking, but certain holes are still present in our knowledge of the organism. Certain virulence factors remain unknown, and the role of important effector molecules for disease progression are still not well understood. The relationship between contamination of the environment with low levels of antibiotics and the rapid development of antibiotic

resistance is well-established (Haack et al., 2015; Mellon et al, 2001), but the role of mobile genetic elements such as plasmids has yet to be fully understood.

To address the gaps in our knowledge, my thesis focused on assessing the prevalence of antibiotic-resistant strains of *S. enterica* in stream sediment and poultry litter from the Shenandoah River Valley. Specifically, I focused on identification of *S. enterica* isolates using bioinformatics, with an interest in clinically significant serotypes. I also examined these isolates for the presence of one or multiple antibiotic resistance genes and focused on what role plasmids potentially have in the dissemination of these genes. These objectives were undertaken to heighten our understanding of what *Salmonella* persist in these environments, as well as study the potential dynamics at work that perpetuate development of antibiotic resistance in *S. enterica.*

Methods

Environmental sample collection

Poultry litter was acquired on 02/01/2017 from various commercial and non-commercial sources. Sterilized plastic containers were filled with litter and transported back to the lab. Litter was stored at room temperature (20-25°) until sub-sampled.

Stream sediment was collected between 10/02/2016 and 09/30/2018 from stream sources around the Shenandoah Valley of Virginia including Muddy Creek Rt. 33 (Latitude: 38.467152, Longitude: -78.974999), Cook's Creek Rt. 11 (Latitude: 38.37270572, Longitude: -78.93450069), Cook's Creek Rt. 704 (Latitude: 38.39030156, Longitude: -78.94758535), Pleasant Run Rt. 989 (Latitude: 38.35448124, Longitude: -78.91992456), Cook's Creek Park Rt. 732 (Latitude: 38.41984001, Longitude: -78.9394984), Black's Run (Latitude: 38.424085, Longitude: -78.882169), Wenger's Mill Farm, and Cook's Creek Arboretum (Latitude: 38.386341, -78.950744). Sample collection at all sites was done in triplicate to ensure a sufficient quantity of sediment was available for bacterial isolation and plasmid capture. Stream water characteristics, including salinity, temperature, pressure, and conductivity, were collected using a Sonde™ probe (YSI Incorporated, OH, USA). All metadata was recorded using the mobile application Epicollect5 (https://five.epicollect.net/). Stream sediment was collected by inverting a 50 mL Falcon® tube with a gloved hand and inserting tube straight down into sediment devoid of plant matter and gravel. The tube was then inverted in a scooping motion and excess water was poured out. Each tube was filled with

approximately 75 g of sediment. Subsequent samples from the same source were collected upstream if possible. Sediment samples were stored on ice during transport to the lab where they were stored at 4°C.

Isolation of environmental *Salmonella enterica* from stream sediment/poultry litter

A modified version of the US FDA BAM protocol was used (Figure 1) (Andrews *et al.*, 2014, Jurgensen SK, 2017). Pre-enrichment from sediment and poultry litter samples began within 24 hours of sample collection. For each site, approximately 50 g of stream sediment or litter was weighed out and transferred to sterile 250 mL Erlenmeyer flasks in duplicate. One hundred milliliters of buffered peptone water (10 g peptone, 5 g NaCl, 7 g Na_2HPO_{4-}, and 3 g $KH_{2-}PO_{-4}$) were added to each flask and then mixed by swirling. The pre-enrichments were incubated with shaking at 35°C for 16-22 hours.

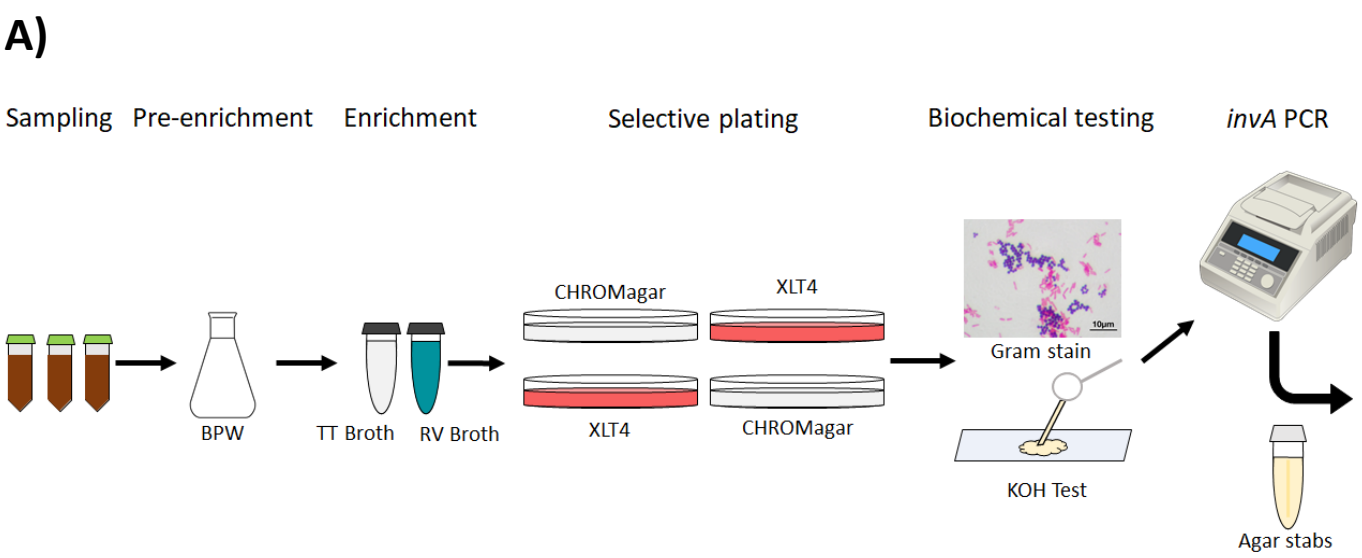


Figure 1: A) Protocol used to isolate *Salmonella enterica* from poultry litter and stream sediment. Agar stabs were shipped to either the FDA or the VA DCLS for sequencing. BPW- Buffered Peptone Water, TT- Tetrathionate, RV- Rappaport-Vassiliadis, BS- Bismuth Sulfite, XLT4- Xylose-Lysine-Tergitol™

Pre-enrichments were removed from the incubator and vortexed to re-suspend sample solids. Solids were allowed to settle, and 1 ml of the supernatant was added to the enrichment tubes. Two types of enrichment media, Tetrothionate (TT) broth and Rappaport- Vasilliadis (RV) broth were used for selective enrichment of *S. enterica* (Stone *et al.*, 1994, Jorgensen *et al.*, 2002, and Kinde *et al.*, 2004), with each pre-enrichment inoculated into both enrichment media. One liter of TT broth was prepared by boiling TT broth base (5 g polypeptone, 1 g bile salts, 10 g $CaCO_3$ and 30 g --$Na_2S_2O_3$). The solution was cooled to a minimum of 45°C and 20 mL of potassium iodide solution (5 g KI, 6 g Iodine resublimated) were added. Ten milliliters of TT broth were added to sterile screw-cap test tubes. RV medium was made by adding 100 mL of magnesium chloride solution (400 g $MgCl_2 \cdot 6H_2O$) and 10 mL of malachite green oxalate solution (0.4 g malachite green oxalate) to one liter of broth base (5 g tryptone, 8

g NaCl, and 1.6 g KH_2PO_4). Ten milliliters of RV broth were added to screw-cap test tubes and autoclaved for 15 minutes at 115°C. Each tube was inoculated with 1 mL of pre-enrichment. All tubes were incubated shaking at 42°C for 5 days.

A 100 µL aliquot from each enrichment was spread-plated onto Bismuth Sulfite (BS) or CHROMagar™ *Salmonella* agar (referred to as CHROMagar hereafter) and Xylose Lysine Tergitol-4 (XLT4) agar plates. After 05/15/18, BS agar was replaced by CHROMAgar™ as *Salmonella* was found to be more easily discernible on the CHROMagar™. Plates were incubated at 35°C for 15-22 hours. Colony color and morphology were used to identify putative *S. enterica*. On BS agar, *S. enterica* colonies appear dark and round with a metallic sheen (Andrews *et al.*, 2011). On XLT4, *S. enterica* colonies have a convex shape and are either totally black or have a black dot at the center of the colony. On CHROMagar™ *Salmonella* have a mauve coloration. Any colonies that displayed *Salmonella*-like morphology were streaked onto the other, complimentary agar. If *Salmonella*-like morphology was present on both the BS/CHROMagar™ and XLT4 media, then those colonies were re-streaked onto tryptic soy agar (TSA) plate for purification and further testing.

To verify the identities of putative *Salmonella*, a KOH test, Gram stain, and *invA* PCR were performed. A single colony from a TSA plate was collected with a sterilized inoculating loop and transferred to a microscope slide containing a drop of 3% KOH. Formation of a sticky DNA 'string' (KOH+) was indicative of a Gram-negative bacterium. Isolates that were not KOH+ were discarded. Isolates which passed the KOH test were gram-stained to verify the results of the KOH test. Confirmation of *Salmonella* was done by performing a modified *invA* PCR (Burgess *et al.*, 2015). An inoculating

needle was used to acquire a minute number of cells from a single colony. Cells were added to 5µL of ddH_2O in a sterile 0.2mL PCR tube and placed into a thermocycler to be lysed at 95°C for 5 minutes. A master mix was prepared using 12.5 µL of 2X AmpliTaq Gold® (0.625 U AmpliTaq Gold DNA polymerase, 30 mM Tris/HCl, pH 8.05, 100 mM KCl, 400 µM each dNTP, and 5 mM $MgCl_2$), 1 µL of both 139 and 141 primers at a concentration of 10 µM each primer, and 5 µL of ddH_2O for a final volume of 25 µL per tube. The *invA* 139 primer sequence used was 5'-GTGAAATTATCGCCACGTTCGGGCAA-3' and the *invA* 141 primer sequence used was 5'-TCATCGCACCGTCAAAGGAACC-3'. The expected band size from *invA* PCR is 285bp. The *invA* PCR thermocycler program used were those described by Malorny *et al.*, 2003; tubes were incubated for 1 minute at 95°C, followed by 36 cycles of 95°C for 30 seconds, 64°C for 30 seconds, 72°C for 30 seconds, and a final extension of 72°C for 4 minutes. Gel electrophoresis was performed on the PCR amplicons to check for successful amplification and correct band size. Two microliters of 5X loading dye were mixed with 8 µL of PCR product. The mixture was loaded into a 1.5% DNA agarose gel. The gel was run at 5 V/cm for ca.120 minutes. After the run, gels were submerged in 0.5% GelRed (Biotium Inc., Fremont, CA) for 20-30 minutes for staining and de-stained with ddH_2O for approximately 5 minutes. Visualization of bands was done using a UV transilluminator.

For long term preservation of cultures, isolates that produced the correct target band size were grown up in TSB for 16-22 hours at 37°C shaking at 180-220 rpm. One milliliter of broth culture was combined with 1 mL of sterile glycerol in a 2mL cryogenic freezer tube and stored at -80°C.

Extraction of Plasmid DNA

Extraction of plasmid DNA was performed on HJ-04R transconjugants to verify the presence of a plasmid. Plasmid DNA was isolated using a plasmid miniprep protocol developed in our laboratory (Libuit, 2016). Selected isolates were grown in TSB at 37°C, 180-220 rpm for 16-22 hours. *E. coli* containing the pEG1 plasmid was used as a positive control (Gehr, 2012). One and a half milliliters of cell culture were transferred to microcentrifuge tubes and centrifuged at 10,000 x g for 5 minutes at either 4°C or room temp. Supernatant was removed by aspiration and pelleted cells were resuspended in 100 µL of filter-sterilized resuspension buffer (50 mM dextrose, 10 mM EDTA, 10 mM Tris-Cl, at pH 8). One hundred microliters of fresh NaOH/SDS solution (0.2M NaOH, 1% SDS) were added and the tubes were mixed by inverting 5 times. The tubes stood at room temperature for 5 minutes. One hundred and fifty microliters of filter-sterilized 7.5M ammonium acetate ($C_2H_7NO_2$) were added, followed by 150 µL of chloroform. This was mixed by inversion 5 or more times. The tubes were chilled for 10 minutes on ice and then centrifuged at 10,000 x g for 10 minutes at room temperature. After centrifugation, as much of the aqueous phase as possible was added to a new 1.5 mL microcentrifuge tube that contained 200 µL 30% polyethylene glycol 8000 (PEG)/1.5M NaCl. The new tube was inverted 3 to 4 times and chilled on ice for 10 to 15 minutes. Tubes were then centrifuged at 10,000 x g for 10 minutes. Supernatant was removed by aspiration and 1 mL of 70% ethanol was added to the tube with the pellet. The tubes were centrifuged at 10,000 x g for 5 minutes. Supernatant was removed without disturbing the pellet and

samples were allowed to air dry for 10 minutes. After drying, the pellet was resuspended in 100 µL of sterile, ddH_2O water and stored at 4°C for at least 24 hours to allow DNA to dissolve. Confirmation of plasmid presence and removal of chromosomal DNA carry-over was done using PlasmidSafe™ ATP-dependant DNase (Epicentre Technologies, Madison, WI) enzyme as described by the manufacturer and by Libuit (2016). In 1.5 mL microcentrifuge tubes, 42 µL of plasmid DNA, 2 µL of 25mM ATP, 5 µL of 10X reaction buffer (vortexed vigorously prior to addition) and 1 µL PlasmidSafe™ were added. Tubes were incubated at 37°C for 1 hour and then incubated again at 70°C for 30 minutes to inactivate the DNase. The tubes were briefly centrifuged to collect liquid to the bottom of the tube. Samples were visualized on a 1% DNA agarose gel. The gel was prepared and submerged in 1X TAE buffer. A λ/HinDIII ladder was used as a marker. Prior to loading, 16 µL from each tube were mixed with 4 µL of 5X loading dye. The gel was run at 70 V/cm for 60 minutes. After running, the gel was submerged in a 0.5% GelRed solution (Biotium Inc., Fremont, CA) for 30 minutes for staining of DNA. Excess stain was removed by submerging the gel in ddH_2O for 5 minutes and bands were visualized under UV light. Samples were stored at -20°C until use.

Genomic DNA extraction and concentration

Genomic DNA was extracted from *Salmonella* isolates for the purpose of long read sequence data generation. Cells were grown for 16-20 hours in 4 mL of tryptic soy broth. To extract genomic DNA, a Qiagen™ DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) was used following the standard preparation for gram-negative

bacteria from the DNeasy Blood & Tissue handbook with the following modifications. Cell density was not assessed prior to extraction. Cells were incubated at 56°C with proteinase K for up to an hour to prevent DNA degradation. The DNA was allowed to dissolve at 4°C for 24 hours before being stored at -20°C.

Sequencing genomic DNA on the Oxford Nanopore Technologies® MinION™

Prior to sequencing, DNA purity was determined using the Synergy H1 Multi-Mode Reader (BioTek Instruments, VT, USA). DNA of sufficient purity has an $OD_{260/280}$ ratio between 1.8 and 2.0. Quantification of DNA was done using the Qubit 2.0 fluorometer and the Qubit dsDNA broad range kit (Thermo Fisher Scientific, MA, USA). For sequencing on the MinION™ (Oxford Nanopore Technologies, Oxford, UK) a DNA concentration of at least 53.3 ng/µL (400 ng of DNA in 7.5 µL water) was required. If the concentration of DNA was too low for sequencing, DNA was concentrated using Microcon® centrifugal filters (Merck Millipore Ltd, MA, USA). One hundred microliters of genomic DNA (gDNA) were added to the sample reservoir of the Microcon® tube. The tubes were spun at 500 x g for 20 minutes. After spinning, the sample reservoir was inverted into a new Microcon® tube and these tubes were spun at 1000 x g for 3 minutes. DNA concentration was then re-evaluated as previously described.

Prior to sequence preparation, the MinION™ flow cell was quality controlled as described in the protocol. Quality control and sequencing were performed using the most up to date version of MinKNOW available at the time of sequencing. Flow cells must

have at least 800 active pores to be considered usable. Once the number of pores was determined, the flow cell was kept at 4°C until sequence preparation was complete.

DNA barcoding and library preparation for the MinION™ was done using the rapid barcoding kit (RBK-004) and rapid barcoding kit sequencing protocol (version RNK_9054_v2_revA_23Jan2018, las modified 07/03/2018) (Oxford Nanopore Technologies, Oxford, UK). First, 7.5 µL 400 ng of each template DNA and 2.5 µL of fragmentation mix were added to 0.2 mL PCR tubes. A unique fragmentation mix was given to each sample, labeled RB01 through RB12. These tubes were spun down and incubated at 30°C for 1 minute, then 80°C for 1 minute using a BIO-RAD C100 Touch™ thermal cycler (BIO-RAD Laboratories Inc, CA, USA). If 4 or more gDNA samples were being barcoded and sequenced, an optional AMPure XP bead cleanup (Beckman Coulter, CA, USA) was performed. This was done to ensure that ample, equal amounts of gDNA from each sample were sequenced. The protocol was followed as written with a modification to the final elution step. Sample DNA was pooled and equal amounts of AMPure XP beads were added. Tubes were incubated at room temperature for 5 minutes, after which samples were pelleted using a magnetic rack. The beads were washed with 200 µL fresh 70% ethanol and the ethanol was then removed. Tubes were spun down and the pellet was allowed to reform to allow for collection of any remaining ethanol. DNA was eluted in 11 µL of Tris-HCl pH 7.5-8.0 instead of 10 µL to ensure at least 10 µL of eluate was acquired without carry-over of any beads. After incubating for 2 minutes, the beads were pelleted again and 10 µL of eluate was removed into a 1.5 mL Eppendorf DNA LoBind tube (Eppendorf, Hamburg, Germany). One microliter of rapid adapter was added to the eluate and incubated at room temperature for 5 minutes.

The flow cell was primed with 800 µL of priming mix (30 µL flush tether mixed with one tube of flush buffer) and allowed to sit for 5 minutes. In a new LoBind tube, 34 µL of sequencing buffer were mixed with 25.5 µL of loading beads, 4.5 µL ddH_2O water, and the 11 µL of DNA library previously prepared. The library was mixed before adding it to the flow cell to ensure loading beads were amply re-suspended. Seventy-five microliters of the library were added to the flow cell in a dropwise fashion to ensure the library spread sufficiently over the pores. Once added, the sample port cover was returned, and the sequencing run was executed. Runs were allowed to proceed for 48 hours. If extra library remained, the remainder was added 24 hours after the run was initiated to maximize throughput.

MinION™ base calling and quality control

After sequencing, base calling of MinION™ raw data was done followed by adapter trimming. First, base calling was done using Albacore (Oxford Nanopore Technologies, 2017) and adapter sequences added during the library preparation were removed using Porechop (Wick RR, 2017). Raw reads were uploaded to Open Science Framework (OSF; Foster and Deardorff, 2017). Raw reads were also uploaded to Galaxy (Afgan *et al.*, 2018) for quality control and assembly. MinION™ reads were used in conjunction with Illumina® short reads to perform hybrid assemblies (described below).

Short read sequence data acquisition

Illumina® short read whole genome DNA sequencing was carried out by the US FDA Center for Food Safety and Applied Nutrition or the Virginia Department of Consolidated Laboratory Services (VA-DCLS). Sequencing at both institutions was done on an Illumina® sequencing platform using either a 500 cycle (2x250) kit or a 300 cycle (2x150) kit to generate short paired-end reads. Raw sequence data was submitted to BaseSpace and uploaded to the GalaxyTrakr (Afgan *et al.*, 2018) DCLS-JMU shared library space. All short read sequence data are available through the NCBI's sequence read archive.

Sequence data quality control and assembly

Because the overall quality of long reads from the MinION™ is low compared to Illumina® short reads, these reads were not submitted to FastQC, but were used as is after being run through Porechop. Quality assessment, trimming, and assembly using short reads was done through GalaxyTrakr. Raw short reads generated by the FDA or DCLS were first downloaded from the data library shared with the DCLS. If reads from the DCLS needed to be moved to Galaxy, that was done by downloading the reads to a local drive and re-uploaded to Galaxy. Quality assessment was done in GalaxyTrakr using FastQC v. 0.6.9 (Andrews S, 2010).

Short reads were processed using Trimmomatic v. 0.36.4 (Bolger *et al.,* 2014). For each set of reads, a standardized trimming protocol was used. Standard Trimmomatic operations used were, in this order, SLIDINGWINDOW with a window 4 bases long and an average quality required of 20, AVGQUAL where the average quality of each base required was 27, and MINLEN where the minimum length of a given sequence was the maximum length of a sequence possible minus 70% the maximum length of a sequence possible (i.e. if a sequence length was 250bp long, then the minimum length of a sequence kept had to be 75bp long (250-(250*0.7))). After trimming, quality was re-assessed using FastQC.

Whole genome assembly was done using short read sequence data, as well as doing a hybrid assembly using short and long read data when possible. For short read only assemblies, the short-read assembly tool SPAdes was used in GalaxyTrakr (Bankevich *et al.,* 2012). Default options were used with some exceptions. Single-cell options were not used. Read error correction was performed (option no), as was careful correction (option no). This was done to produce the most accurate assemblies possible. K-mer values were specified as follows: 21, 33, 55, 77, 99, and 127. No coverage cutoff was chosen (option off). All short-read only assemblies were done using paired-end reads in a forward-reverse orientation (fr). Forward and reverse reads were input as separate files in FASTQ format. Lists of trusted and untrusted contigs were not used. Outputs specified were an assembly in FASTA format and a graph in FASTG format. Hybrid assemblies were done using the most recent version of Unicycler (Wick *et al.,* 2017) available on Galaxy and default options were used. Paired-end short reads in FASTQ format, along with long-reads in FASTQ format, were input as separate files. Bridging

mode was set to normal to minimize the chance of mis-assembly, but still produce contigs of a decent length. Minimum contig length was set to 100bp and no linear sequences were expected in the assembly. All segments could be used as scaffolding anchors. All other options for SPAdes, rotation, pilon, graph cleaning, and long read alignment parameters were left unchanged. The default job resource parameters were used as well. Final outputs included a final assembly in FASTA format and a final assembly graph in TXT format.

To assess assembly quality, assemblies were run through the most recent version of QUAST (Gurevich *et al.*, 2013) available on Galaxy and GalaxyTrakr. Quality of assembly was judged based on the N50 value, contig length, and number of contigs. The following quality thresholds were used to judge assemblies: Mean Q greater than 30, an N50 greater than 200,000bp, and fewer than 200 contigs making up the assembly. In cases where an assembly yielded an N50 below 200,000 bp, different trimming procedures were done to yield the best assembly possible. If multiple assemblies were generated with the same N50, the assembly chosen for use in analyses was based on which produced the fewest contigs. Figure 2 shows a flow chart depicting the order of events for sequence quality control and analysis.

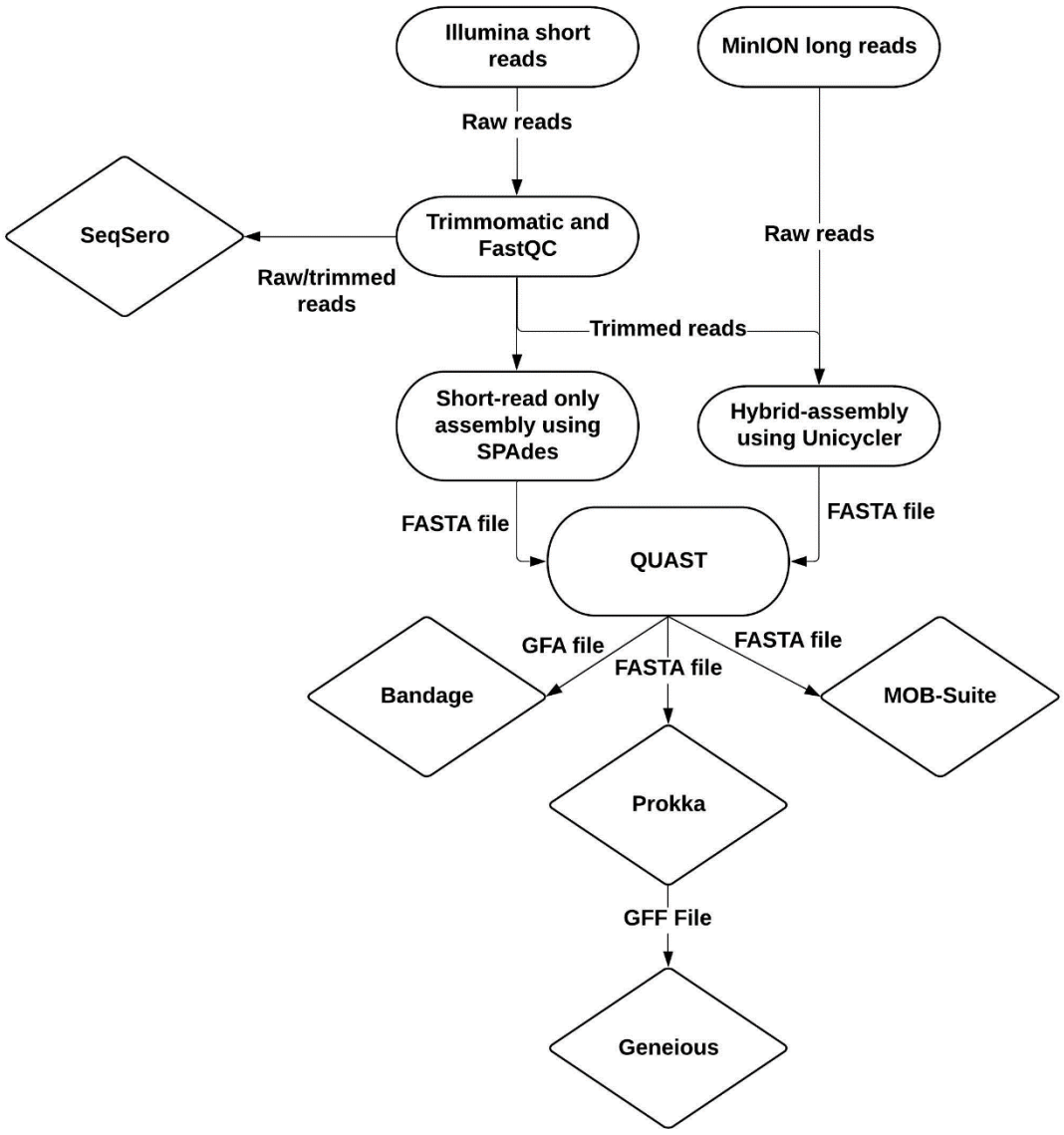


Figure 2: Sequence data analysis flowchart. Long-reads were first base-called using Albacore (Oxford Nanopore Technologies, 2017) and adapters were trimmed using Porechop

Determination of antimicrobial resistance genotype and phenotype

Determination of antibiotic minimum inhibitory concentration (MIC) was performed by Dr. Jonathan Frye of the USDA using Sensititre™ NARMS Gram Negative plates (Thermofisher Scientific, Waltham, MA). Our lab used Sensititre™ Gram Negative Non-Fermenters MIC plates (Thermofisher Scientific, Waltham, MA) to determine the MIC for additional antibiotics.

Detection of antimicrobial resistance genes was done using ABRicate v. 0.8.7 (Seemann, 2016) and the ARG-ANNOT database (Gupta *et al.,* 2014). ABRicate is an antimicrobial resistance gene identifier that can utilize different available databases such as the NCBI's database of known resistance genes. ABRicate was installed through Bioconda and run in Unix (Grüning *et al.,* 2018). Input files for ABRicate were FASTA assembly output files from either SPAdes or Unicycler. The database specified for use was the ARG-ANNOT database (--db argannot) and the minimum similarity percentage required was set to 80% (--minid 80). After identification by ABRicate, genes with a percent coverage below 90% were discounted. To supplement ARG annotation, Prokka (Seemann, 2016) was used for comparison.

In silico identification of plasmids in *S. enterica* isolates

In silico identification of plasmids was done primarily with MOB_Suite v. 1.4.8 and newer (Robertson and Nash, 2018). MOB_Suite uses curated databases of known, complete plasmid sequences and then uses two linked modules, MOB_recon and MOB_typer, to identify and type putative plasmid sequences within a bacterial genome. MOB_typer works similar to plasmidfinder (Carattoli *et al.*, 2014), but includes typing relaxases, oriT predictions, and conjugative ability predictions. When possible, to support the results of MOB_Suite, hybrid assembly graphs were examined for the presence of circularized contigs in Bandage indicative of a plasmid. The output of mob_typer includes the length of plasmid sequences identified; these were used for comparison to the length of individual, circularized contigs present in hybrid assembly graphs for the isolate the plasmid was found in. The outputs of MOB_recon are one or more FASTA sequences of plasmids identified within the input assembly file.

Confirming the presence of ARGs on plasmids was done using ABRicate. MOB_Suite returns all putative plasmid sequences as individual FASTA files, as well as a chromosome FASTA file for all sequences not predicted to be part of a plasmid. Isolates that possessed ARGs first had each plasmid checked for their presence using the settings described previously. If ARGs were identified on one or more plasmids, the chromosome alone was also run through ABRicate to confirm.

To visualize ARGs as well as other genes present on plasmids, annotation files were uploaded to Geneious (Geneious Prime 2019.0.4). Annotation files were generated using Prokka in Galaxy (Seemann, 2014). Identification of putative coding sequences not

defined by Prokka or ABRicate was done by uploading the coding sequence into BLASTx. When using BLASTx, the following parameters were used: Max target sequences was 100, expect threshold was set to 10, word size was set to 6, and the maximum number of matches in a query range was set to 0. For scoring parameters, a BLOSUM62 matrix was used, gap costs option was not changed (Existence: 11, Extension: 1), and the compositional adjustments options was not changed (conditional compositional score matrix adjustment). For filters and masking, low complexity regions were filtered out and no masks were applied.

In silico typing of S. enterica isolates

Serotyping was done primarily using the *Salmonella In Silico* Typing Resource (SISTR) (Yoshida *et al.*, 2016), and SeqSero (Zhang *et al.*, 2015). The most recent versions of SeqSero and SISTR were used through the GalaxyTrakr website. For SeqSero, untrimmed reads were submitted as inputs if the quality was sufficient as determined by FastQC. If untrimmed read quality was determined to be poor, then trimmed reads were submitted. Analysis mode was left on the default option (k-mer mode) and the mem algorithm for BWA mapping was used. For SISTR, the highest quality assembly was chosen as the input. Outputs were returned in a tabular format. The full cgMLST database was not used. Mash MinHash-based serovar prediction was performed, along with cgMLST-based serovar prediction and a QC of the results. The results verbosity option was not changed (basic results only) and a temporary analysis directory was not kept.

Multi locus sequence typing (MLST) was done using Enterobase v. 1.1.2 (Alikhan *et al*., 2018). Raw short read sequences were submitted to Enterobase along with metadata for each isolate. Metadata included isolate ID, collection date, source niche, source type, source details, longitude and latitude, serotype/serovar, and a lab contact. The isolate ID used was the ID assigned by our lab (Table 1.). Isolates submitted by the FDA-CFSAN were submitted under the source niche environment and source type soil/dust for both litter and sediment. All other isolates submitted from stream sediment had the source niche aquatic and source type water. Source details specified the specific source name of the isolate (e.g. Cook's Creek Rt. 11). Enterobase accepts raw or trimmed short reads and performs its own assembly. Enterobase also automatically determines the MLST as well as ribosomal MLST (rMLST), core genome MLST (cgMLST), and whole genome MLST (wgMLST). We used the MLST and cgMLST for characterizing isolates.

Generation of minimum spanning trees using GrapeTree via Enterobase

Minimum spanning trees (MSTrees) were generated using GrapeTree through Enterobase (Zhou *et al.*, 2018). The algorithm used to calculate genetic relatedness for all trees was a neighbor joining algorithm (Saitou and Mei, 1987). Neighbor joining was utilized because Zhou *et al.* found that this approach generated trees with the best accuracy and precision. For MSTrees generated to examine serotype genetic variation among source niches, a subset of isolates from each source niche was chosen. Of the 11 source niches available on Enterobase, the laboratory and ND niche were not used for our MSTrees due to the uniqueness of the laboratory niche; ND was treated as not having a

source niche. For each of the remaining 9 source niches, including aquatic, companion animal, environment, feed, food, human, livestock, poultry, and wild animal, a maximum of 25 isolates were selected. Isolates used in an MSTree for each serotype possessed a predicted serotype (according to SISTR via Enterobase) matching the listed serovar in the metadata. Isolates used also had a cgMLST as determined by Enterobase. Any isolates that did not possess a cgMLST value were rejected for usage. MSTrees were logarithmically scaled to produce more readable trees while maintaining genetic relatedness according to distance.

Results

Isolation of *Salmonella* from stream sediment and poultry litter

Putative *S. enterica* were identified by plating on two types of selective media. Bismuth sulfite (BS) agar and xylose-lysine-tergitol (XLT4) were used until May 2018, when BS agar was replaced by CHROMagar™ *Salmonella* agar (CHROMagar). *Salmonella* on BS agar forms small, shiny, metallic-looking colonies. *Salmonella* on XLT4 forms colonies that display either a black bulls-eye in the center or are totally black. *Salmonella* on CHROMagar form small, mauve colonies. Putative *Salmonella* were re-streaked to obtain pure colonies and underwent both KOH test and a Gram stain. Colonies that were pink (Gram -) rods under a microscope and had a string form when exposed to KOH (KOH +) underwent *invA* PCR to confirm the presence of the *invA* gene (Figure 3). These isolates were then shipped to either the Food and Drug Administration (FDA) or the Virginia Department of Consolidated Laboratory Services (VA DLCS) for short-read sequencing.

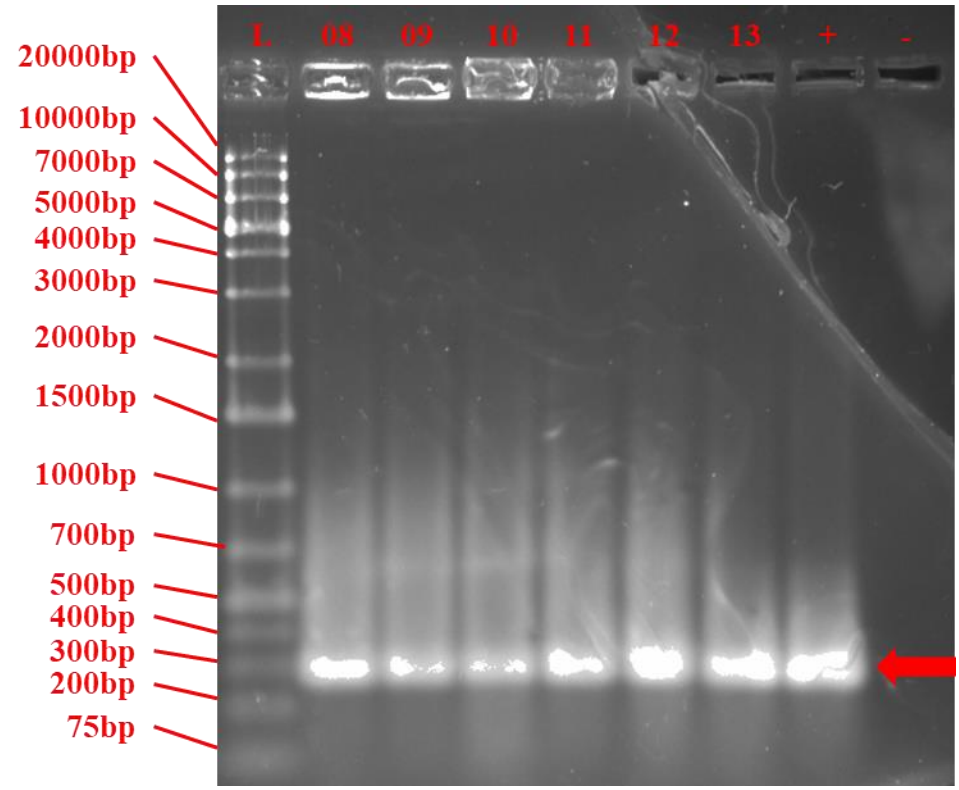


Figure 3: DNA agarose gel depicting positive results for an *invA* PCR for putative *S. enterica*. Numbers represent WMD isolates 08-13. Positive control (+) was HJ-24. Negative control (-) was ddH_2O. L- GeneRuler 1kb ladder (Thermofisher Scientific, Waltham, MA). Arrow indicates the expected band sizes for *invA* positive *S.enterica*

Genome assembly of *S. enterica* isolates

Short-read data were obtained from the NCBI sequence read archive (SRA) through the web platforms Galaxy and/or GalaxyTrakr. Short-read data were first quality checked using FastQC (Andrews, 2010). Short-read data were trimmed using Trimmomatic (Bolger *et al.*, 2014). A standardized trimming operation was used (see Methods) and then sequence quality was assessed again using FastQC. MultiQC was used to generate a cumulative report on the quality of short-read sequences post trimming

(Figure S1-S3; Ewels *et al.*, 2016). Short-read data was assembled using SPAdes (Bankevich *et a.*, 2012) and quality of assembly was assessed using QUAST (Gurevich *et al.*, 2013). Assemblies with an N50 of greater than 200,000 base pairs and fewer than 200 contigs were used for analyses.

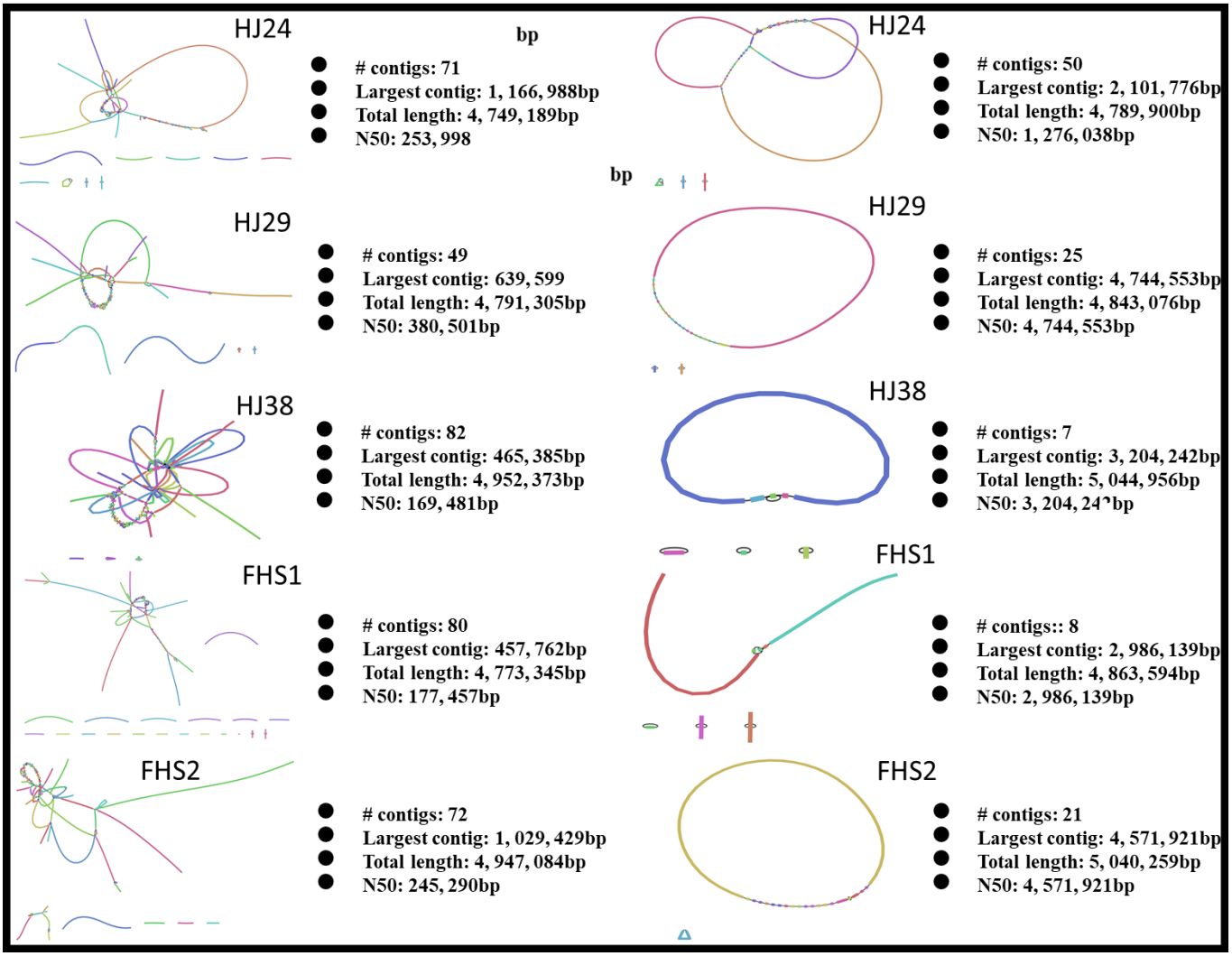


Figure 4: Comparison of assemblies using exclusively short-read data (left) and hybrid assemblies using short and long-read data (right). Images were taken from assembly graphs visualized using Bandage. Assembly metrics were generated using QUAST

Long-read data generated by the MinION™ were not subjected to FastQC prior to use due to the expected low quality of long-read data. Short-read data was supplemented with long-read data to generate more accurate assemblies using a hybrid approach. Hybrid assemblies created with Unicycler produced assemblies with fewer, longer contigs and a higher N50 value (Figure 4) (Wick *et al.,* 2017).

Characterization of *S. enterica* isolates

S. enterica isolates were characterized by multi locus sequence type (MLST), core genome MLST (cgMLST), and serotype. Serotype was determined using both SeqSero (Zhang *et al.,* 2015) and SISTR (Yoshida *et al.,* 2016). SeqSero accepts raw and trimmed short-read data, while the input for SISTR is an assembled genome. Matching results from both tools were used to determine the serotype of the 88 *S. enterica* isolates collected by the Herrick lab as shown in Table 1. Sequence typing was done via Enterobase (Alikhan *et al*., 2018). The core genome of *Salmonella* on Enterobase encompasses 3,002 genes that were present in at least 98% of the 3144 representative *S. enterica* genomes and intact in at least 94% of those same representative genomes (Alikhan *et al.,* 2018). *S. enterica* isolated by the Herrick lab group into clusters of like serotypes when using cgMLST (Figure 5-6).

Table 1: 88 *S. enterica* isolates collected over a 2-year period from stream sediment and poultry litter. *Salmonella* were isolated by previous students from the lab and from the *Bacterial Discovery* (BIO346) course at JMU. ✤: Serotypes that caused the greatest number of cases of salmonellosis in 2016 according to the National Salmonella Surveillance Report (CDC, 2016). MLST- Multi-locus sequence typing, cgMLST- core genome multi-locus sequence typing

Isolate	SRR #	Collection Date	Source[a]	Source type	Serotype	MLST	cgMLST
HJ-01	SRR5886281	10/2/2016	MC	Sediment	Give	654	97788
HJ-02	SRR5886286	10/2/2016	MC	Sediment	Give	654	80760
HJ-03	SRR5886299	10/2/2016	PR	Sediment	Uganda	684	80845
HJ-04	SRR5886298	10/2/2016	PR	Sediment	Uganda	684	70491
HJ-05	SRR5886283	10/16/2016	CC11	Sediment	Litchfield	214	80915
HJ-06	SRR5886290	10/16/2016	CC11	Sediment	Schwarzengrund	96	80849
HJ-07	SRR5886279	10/16/2016	PR	Sediment	Muenster	321	80796
HJ-08	SRR5886351	10/16/2016	PR	Sediment	Muenster	321	80796
HJ-09	SRR5886350	10/16/2016	CC704	Sediment	Mbandaka	413	80794
HJ-10	SRR5884063	12/5/2016	MC	Sediment	Anatum✤	64	80537
HJ-11	SRR5884068	12/5/2016	CC704	Sediment	Schwarzengrund	96	80539
HJ-12	SRR5884053	12/5/2016	CC704	Sediment	Senftenberg	14	80536
HJ-14	SRR5884070	1/15/2017	CC11	Sediment	Hadar	33	80527
HJ-15	SRR5884058	1/15/2017	CC11	Sediment	Hadar	33	80538
HJ-16	SRR5884066	2/1/2017	WM	Litter	Cerro	367	80542
HJ-17	SRR5884067	2/1/2017	WM	Litter	Typhimurium✤	19	80529
HJ-18	SRR5884056	2/1/2017	WM	Litter	Typhimurium✤	19	80535
HJ-19	SRR5884062	2/1/2017	WM	Litter	Typhimurium✤	19	80530
HJ-20	SRR5884079	2/1/2017	WM	Litter	Typhimurium✤	19	80534
HJ-21	SRR5884080	2/5/2017	CC11	Sediment	Typhimurium✤	19	80528
HJ-22	SRR5884077	2/26/2017	CC704	Sediment	Muenchen✤	112	80532
HJ-23	SRR5884081	2/26/2017	CC704	Sediment	Muenchen✤	112	80526
HJ-24	SRR6832866	9/10/2017	PR	Sediment	Montevideo✤	138	111799
HJ-25	SRR6832877	9/10/2017	PR	Sediment	Montevideo✤	138	169196
HJ-26	SRR6832873	9/10/2017	MC	Sediment	Senftenberg	14	111800
HJ-27	SRR6366729	10/22/2017	CC704	Sediment	Cerro	367	101373
HJ-28	SRR6367403	10/22/2017	CC704	Sediment	Cerro	367	101373
HJ-29	SRR6369106	10/22/2017	CC704	Sediment	Anatum✤	64	11895
HJ-30	SRR6367413	10/22/2017	CCP	Sediment	Braenderup✤	22	4601
HJ-31	SRR6367404	10/22/2017	CCP	Sediment	Braenderup✤	22	101420
HJ-32	SRR6367414	10/22/2017	CCP	Sediment	Braenderup✤	22	4601
HJ-33	SRR6367467	10/22/2017	MC	Sediment	Montevideo✤	138	101411

HJ-34	SRR6832876	1/21/2018	CCP	Sediment	Braenderup*	22	113399
HJ-35	SRR6832896	1/21/2018	CCP	Sediment	Braenderup*	22	4601
HJ-36	SRR6832925	1/21/2018	CCP	Sediment	Typhimurium*	19	113353
HJ-37	SRR6832911	1/21/2018	CCP	Sediment	Typhimurium*	19	111797
HJ-38	SRR6832904	1/21/2018	CCP	Sediment	Typhimurium*	19	111798
HJ-39	SRR6832916	1/21/2018	CCP	Sediment	Typhimurium*	19	111391
FHS-01	SRR6832910	1/23/2018	CC704	Sediment	Montevideo*	138	166542
FHS-02	SRR6832913	1/23/2018	CC704	Sediment	Infantis*	32	111896
FHS-04	SRR6832912	1/23/2018	CC704	Sediment	Montevideo*	138	152383
DG-01	SRR8104592	5/20/2018	PR	Sediment	Muenchen*	112	131452
DG-02	SRR8104587	5/20/2018	PR	Sediment	Montevideo*	138	131598
DG-03	SRR8104581	5/20/2018	PR	Sediment	Montevideo*	138	131599
DG-06	SRR7504390	5/20/2018	CC11	Sediment	Hadar	33	131627
DG-07	SRR7506701	5/20/2018	CC11	Sediment	Hadar	33	136579
DG-08	SRR7506695	5/20/2018	CC704	Sediment	Typhimurium*	19	136587
DG-09	SRR7506699	5/20/2018	CC704	Sediment	Typhimurium*	19	131813
DG-10	SRR7506697	5/20/2018	CC704	Sediment	Typhimurium*	19	153131
DG-11	SRR7504359	5/20/2018	CC704	Sediment	Typhimurium*	19	153155
DG-12	SRR7506710	5/20/2018	CC704	Sediment	Typhimurium*	19	153158
DG-13	SRR7499244	5/20/2018	CC704	Sediment	Typhimurium*	19	153152
DG-14	SRR7499253	5/20/2018	MC	Sediment	Meleagridis	463	135814
DG-15	SRR7499245	5/20/2018	MC	Sediment	Meleagridis	463	135813
DG-16	SRR7499272	5/20/2018	MC	Sediment	Cerro	367	135849
DG-17	SRR7499280	5/20/2018	MC	Sediment	Meleagridis	463	135804
DG-18	SRR7499278	5/20/2018	MC	Sediment	Meleagridis	463	135809
DG-19	SRR7889322	5/20/2018	CC11	Sediment	Schwarzengrund	96	144617
DG-20	SRR7878396	5/20/2018	CC704	Sediment	Alachua	1298	144057
DG-21	SRR7889352	5/20/2018	CCP	Sediment	Mbandaka	413	144638
DG-22	SRR7878395	5/20/2018	BR	Sediment	Mbandaka	413	144059
AP-01	SRR8179982	9/3/2018	CC11	Sediment	Bareilly*	2553	153210
AP-02	SRR8179943	9/3/2018	CC11	Sediment	Bareilly*	2553	153210
AP-03	SRR8179958	9/3/2018	CC11	Sediment	Bareilly*	2553	153210
AP-04	SRR8179911	9/3/2018	CC11	Sediment	Bareilly*	2553	153210
BES-01	SRR8179966	9/3/2018	CC704	Sediment	Reading	412	153206
BES-02	SRR8179907	9/3/2018	CC11	Sediment	Reading	412	153208

PPL-01	SRR8179980	9/3/2018	CC11	Sediment	Braenderup ✤	22	4601
PPL-02	SRR8179892	9/3/2018	CC11	Sediment	Braenderup ✤	22	4601
WEK-01	SRR8179901	9/3/2018	CC704	Sediment	Montevideo ✤	138	153130
WEK-02	SRR8179936	9/3/2018	CC11	Sediment	Mbandaka	413	153196
WEK-03	SRR8179927	9/3/2018	CC11	Sediment	Braenderup ✤	22	4601
WEK-04	SRR8179981	9/3/2018	CC11	Sediment	Mbandaka	413	153188
WMD-01	SRR8570270	9/30/2018	PR	Sediment	Muenchen ✤	112	164635
WMD-02	SRR8570265	9/30/2018	CC11	Sediment	Give	654	169281
WMD-03	SRR8570271	9/30/2018	MC	Sediment	Montevideo ✤	138	164641
WMD-04	SRR8570267	9/30/2018	MC	Sediment	Montevideo ✤	138	164640
WMD-05	SRR8570264	9/30/2018	CCA	Sediment	Muenster	321	164639
WMD-07	SRR8573695	9/30/2018	PR	Sediment	Montevideo ✤	138	164637
WMD-08	SRR8570269	9/30/2018	CCA	Sediment	Montevideo ✤	138	164636
WMD-09	SRR8570273	9/30/2018	MC	Sediment	Montevideo ✤	138	164643
WMD-10	SRR8570263	9/30/2018	CCA	Sediment	Cerro	367	164638
WMD-11	SRR8570268	9/30/2018	MC	Sediment	Typhimurium ✤	19	164642
WMD-12	SRR8570266	9/30/2018	BR	Sediment	Give	654	164644
WMD-13	SRR8570272	9/30/2018	CCA	Sediment	Cerro	367	164645

[a]: PR- Pleasant Run, CC11- Cook's Creek Rt. 11, CC704, Cook's Creek Rt. 704, CCP- Cook's Creek Park, CCA- Cook's Creek Arboretum, BR- Black's Run, MC- Muddy Creek, WM- Wenger's Mill Farm

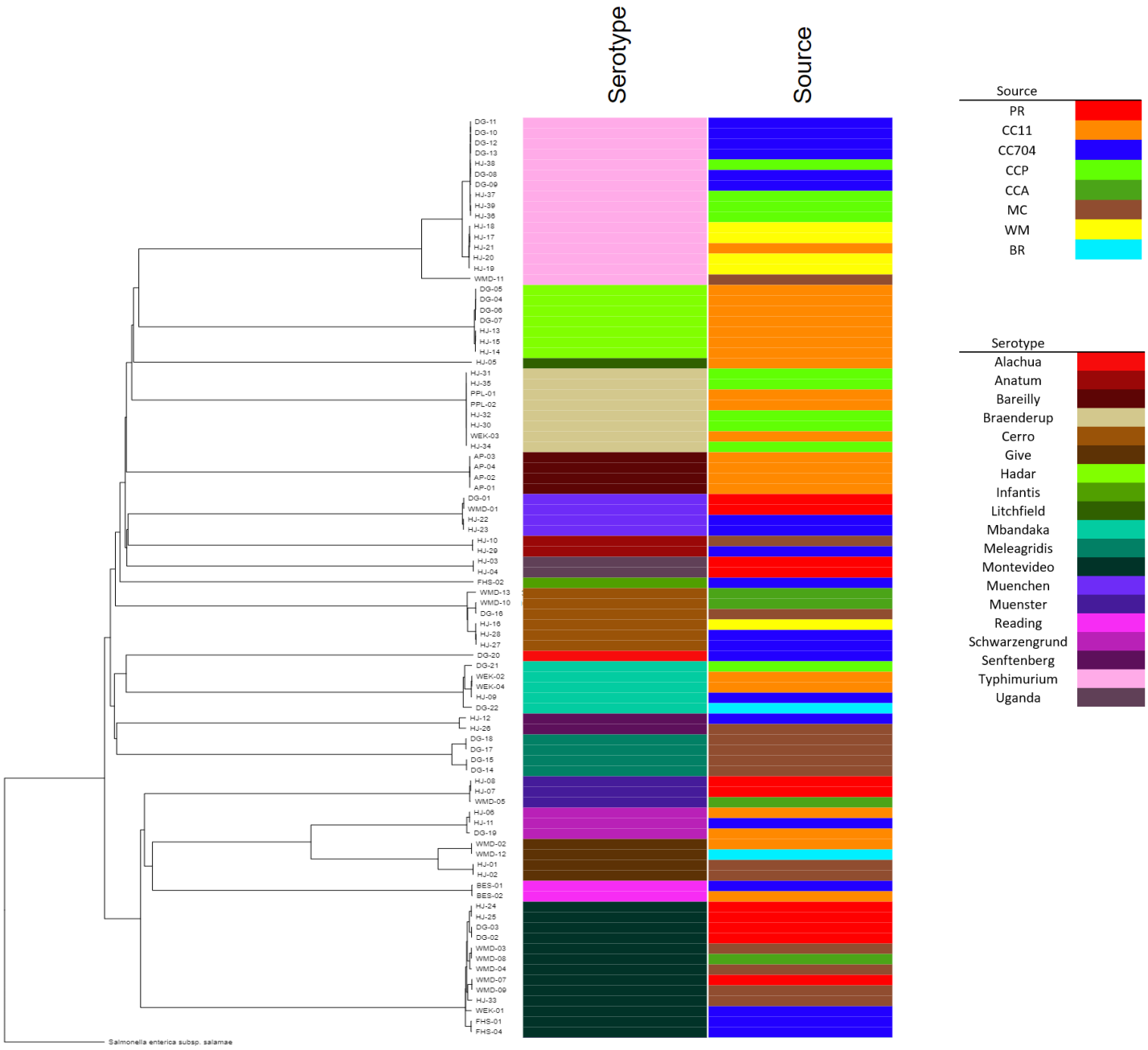


Figure 5: Phylogenetic tree showing genetic relatedness of 88 *S. enterica*. Tree was generated using GrapeTree (Zhou *et al.*, 2018) in Enterobase. A neighbor joining algorithm was used and the tree was exported from Enterobase in newick format. Tree was rooted using *S. enterica* subsp. salamae. Tree and metadata were uploaded to Phandango (Hadfield *et al.*, 2018) for visualization. PR- Pleasant Run, CC11- Cook's Creek Rt. 11, CC704- Cook's Creek Rt. 704, CCP- Cook's Creek Park, CCA- Cook's Creek Arboretum, MC- Muddy Creek, WM- Wenger's Mill Farm, BR- Black's Run

Examination of clinically significant serotypes Typhimurium and Infantis

To examine the genetic relatedness of the isolates, a GrapeTree was created using a neighbor joining algorithm (Zhou *et al.*, 2018). GrapeTrees perform hierarchical clustering based on cgMLST.

The CDC tracks and reports major serotypes that cause disease in their yearly *Salmonella* surveillance report. Serotypes are ranked according how many lab-confirmed cases of salmonellosis are caused by a given serotype, and the report highlights the top 20 each year. Isolates that match one of those top 20 are marked with a ✤ in Table 1. Significant disease-causing serotypes were isolated from both sediment and litter. These serotypes included Typhimurium, Braenderup, Muenchen, Barielly, Montevideo, Infantis, and Anatum. GrapeTrees were generated for each serotype containing select isolates from Enterobase and all isolates collected. To examine how serotypes from different source niches were related to one another and to Herrick Lab isolates, a subset of up to 25 isolates from each source niche in Enterobase were chosen and a minimum spanning tree (MSTree) was generated using a neighbor joining algorithm. Figures 7 and 8 show MSTrees for a subset of *S.* Typhimurium and *S.* Infantis respectively. Isolates for these trees were chosen randomly from the 9 "source niches" on Enterobase. Included in each tree are the strains isolated by the Herrick Lab. In both cases isolates identified by the Herrick Lab group more closely to *Salmonella* isolates from the poultry "niche" than from any other. This grouping of our isolates with those from the poultry "niche" was especially evident when comparing *S.* Typhimurium.

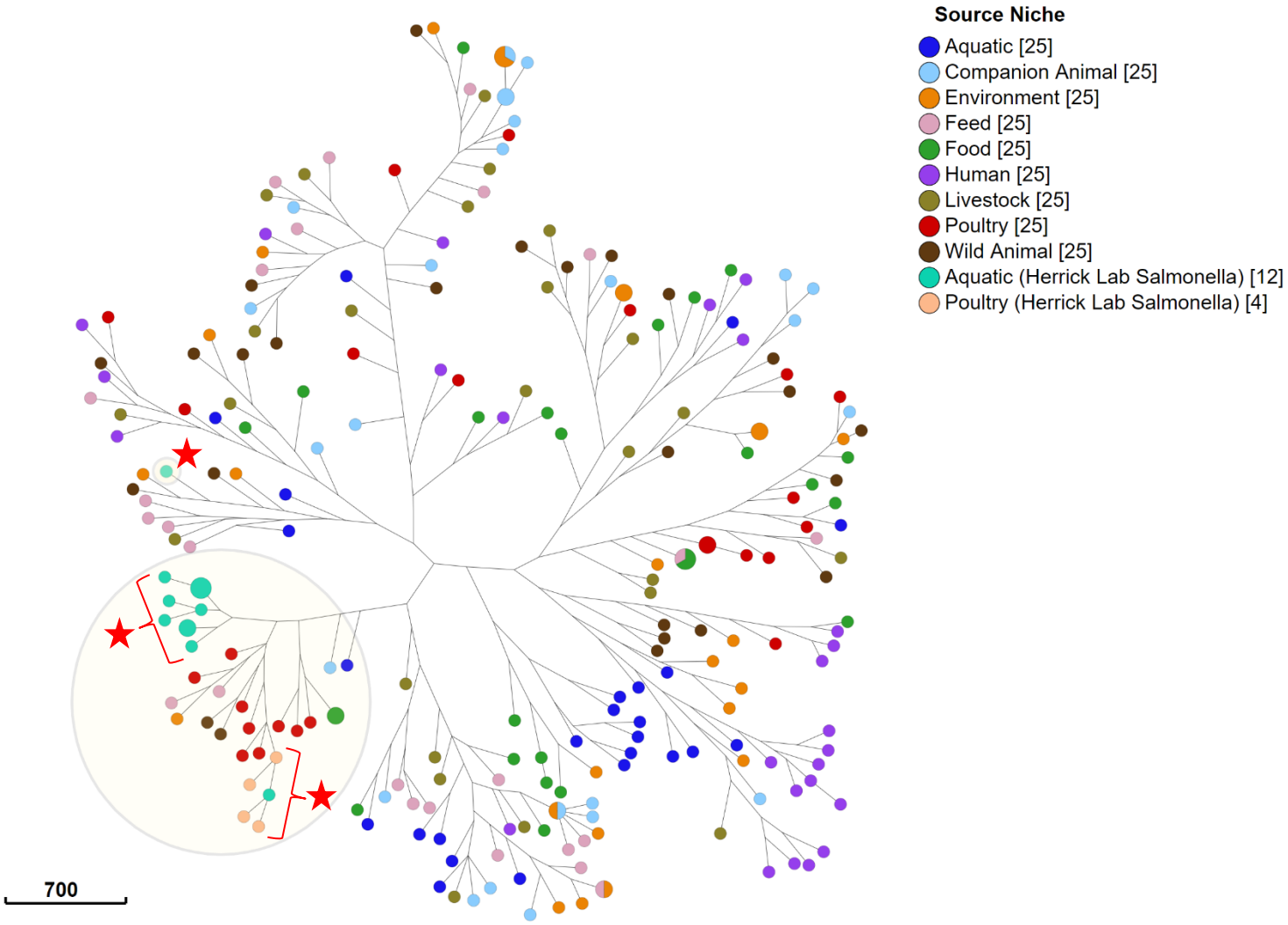


Figure 6: MSTree depicting randomly selected *S.* Typhimurium from 9 different source "niches" with Typhimurium isolated by our lab. A subset of Typhimurium from 9 source niches were selected at random. The tree was logarithmically scaled to improve visual interpretation Larger circles represent multiple *S. enterica* with indistinguishable cgMLSTs. The scale represents the number of allelic differences between isolates. Stars represent *S. enterica* isolated by the Herrick lab. *S. enterica* that are from the Herrick lab and/or share close genetic relatedness to our isolated *S. enterica* are highlighted in the colored circles

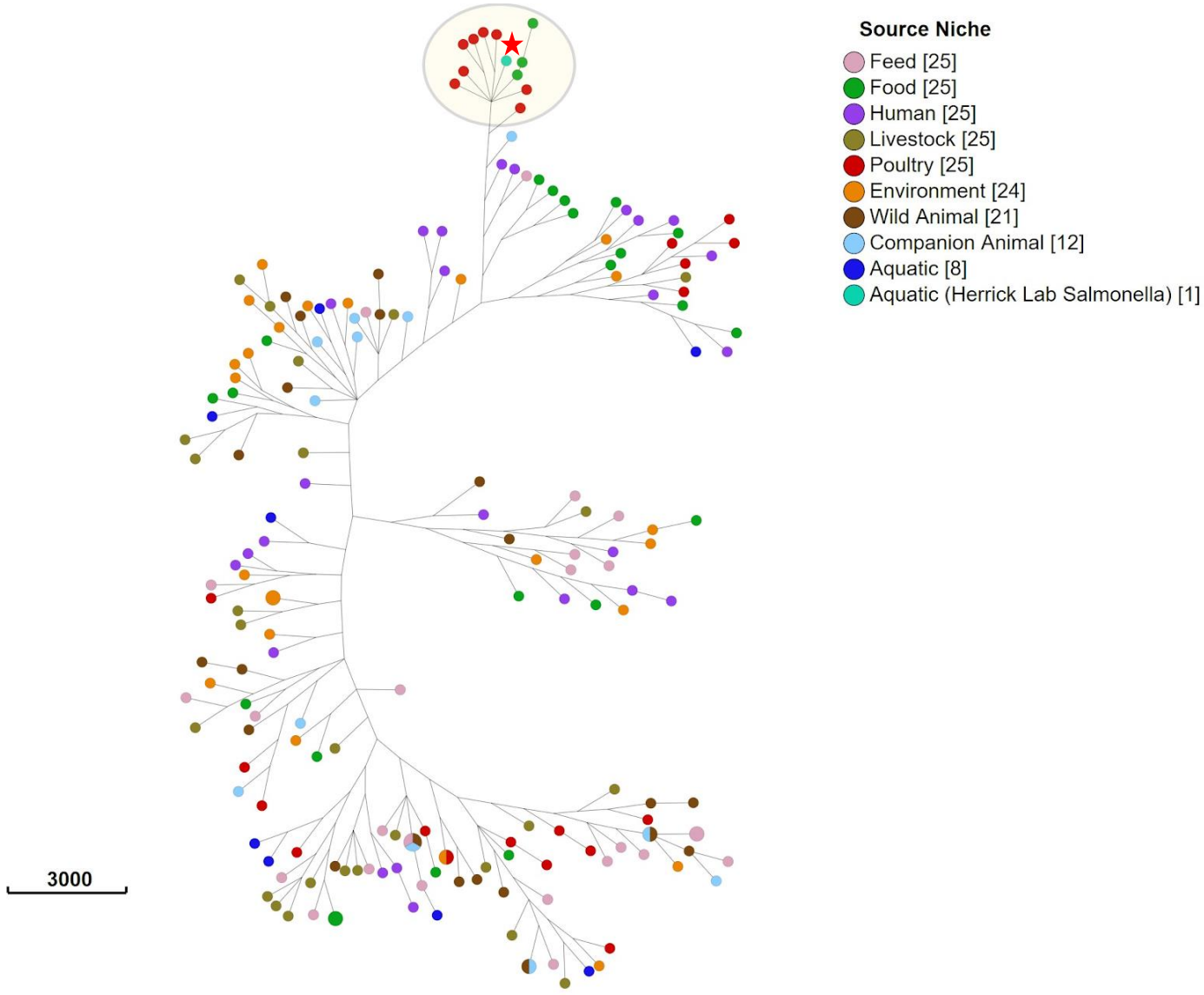


Figure 7: MSTree depicting randomly selected S. Infantis from 9 different source "niches" along with the single *S.* Infantis strain (FHS-02) isolated during this study. Larger circles represent multiple *S. enterica* with indistinguishable cgMLSTs. The scale represents the number of allelic differences between isolates. MSTree was logarithmically scaled to improve visual interpretation. Stars represent *S. enterica* isolated by the Herrick lab. The colored circle indicates *S. enterica* isolates from Enterobase that are genetically similar to our isolated *S.* Infantis FHS-02

Since these serotypes are associated with human infections more often than most other serotypes, *S.* Typhimurium and Infantis from the Herrick Lab were compared to other *Salmonella* of the same serotype isolated from human sources. This was done by

comparing our *S.* Typhimurium and Infantis to those that were isolated from humans, generating a GrapeTree using a neighbor joining algorithm, and identifying the branch that our isolates shared with other human-isolated *S.* Typhimurium or Infantis. GrapeTrees composed of only the closest relatives are shown in figures 8 and 9 respectively.

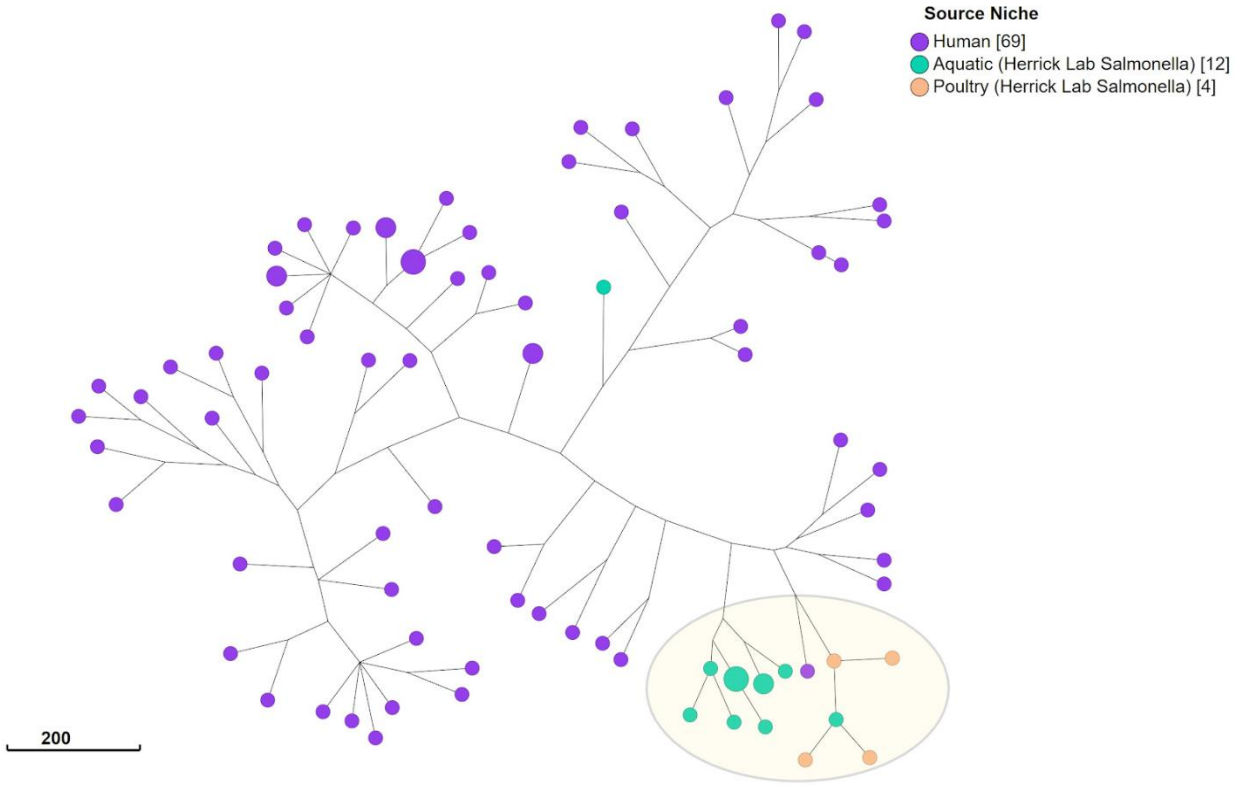


Figure 8: MSTree of *S.* Typhimurium isolated during this study with closest relatives isolated from humans. Our isolates were first compared to all Typhimurium on Enterobase that were isolated from a human. Human-isolated Typhimurium used in this tree were those that shared a branch with our isolated Typhimurium. Larger circles represent multiple *S. enterica* with indistinguishable cgMLSTs. The scale represents the number of allelic differences between isolates. MSTree was logarithmically scaled to improve visual interpretation. The colored circle indicates *S. enterica* isolates from Enterobase that are genetically similar to our isolated *S. enterica*

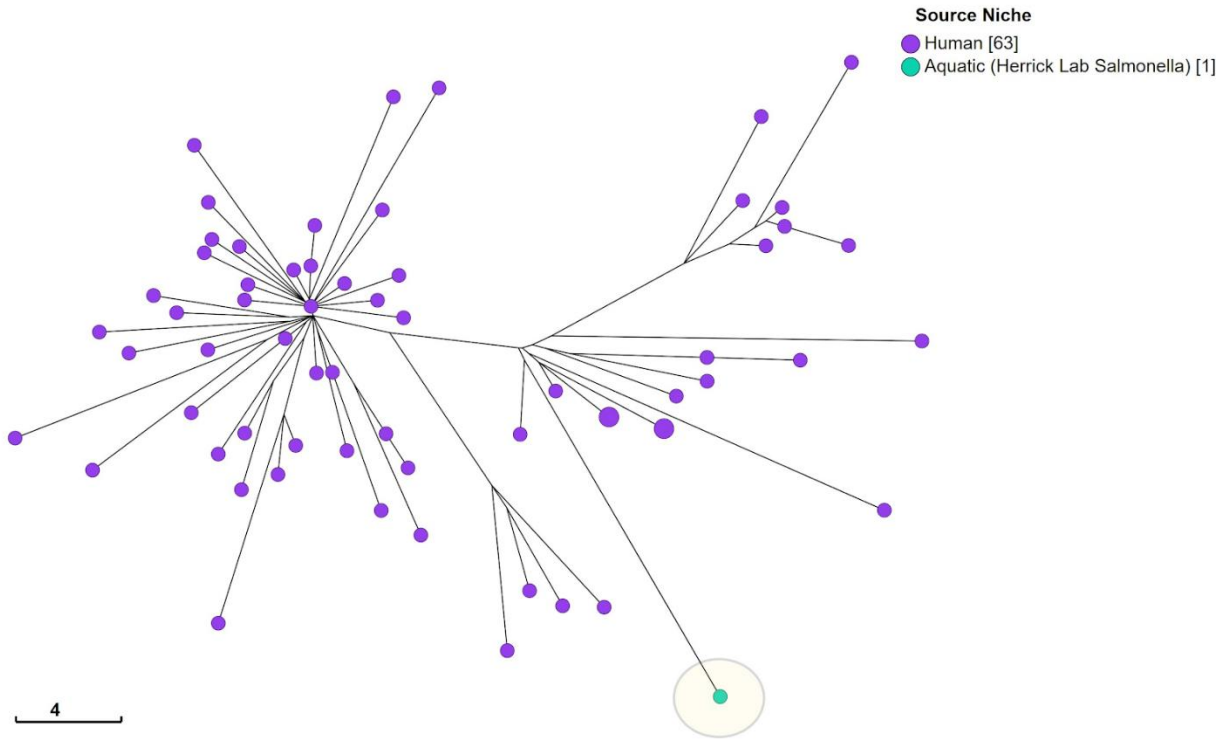


Figure 9: MSTree of *S.* Infantis FHS-02 isolated during this study with closest relatives isolated from humans according to Enterobase. Larger circles represent multiple *S. enterica* with indistinguishable cgMLSTs. The scale represents the number of allelic differences between isolates. MSTree was logarithmically scaled to improve visual interpretation. The colored circle indicates *S. enterica* isolates from Enterobase that are genetically similar to our isolated *S. enterica*

Identification of antibiotic resistance genes

The primary tool used to identify resistance genes within our isolates was ABRicate (Seemann, 2016). A percent identity of 80% and a percent coverage of 90% were chosen as thresholds to account for the presence of gene fragments. Of interest were those resistance genes that were present on plasmids as opposed to the chromosome. In nearly 84% of our isolates, resistance genes were located on one or more plasmids. Plasmid sequences identified by MOB_Suite were separated into individual FASTA files.

These plasmid FASTA sequences were individually run through ABRicate to check for ARGs. MOB_Suite also outputs a FASTA file identified as “chromosome” (anything that was not identified as a plasmid), and this was run through ABRicate to verify ARGs were only on a plasmid. All isolates also possessed at least one aminoglycoside resistance gene within the chromosome (*aac6-Iaa* or an *aac6-Iy*) (Figure 10). The ARGs of certain isolates such as those of HJ-07 were plasmid-borne, but HJ-08’s ARGs were not. Both HJ-07 and HJ-08 are the same serotype, share highly similar core genomes, and were isolated from the same source (Figure 5) at the same time (Table 1); however, HJ-07 possessed a plasmid and HJ-08 did not. The resistance genes *strA, strB, tetA, tetR,* and *sulII* were identified as plasmid-borne in HJ-07, whereas they were identified by ABRicate in the “chromosomal” called by MOB_Suite fraction in strain HJ-08. WMD-05 was also predicted by MOB_Suite to have a plasmid nearly identical to the plasmid in HJ-07. This plasmid contained the same predicted ARGs identified by ABRicate as well. Similar results were seen for strains DG-06 and DG-07. DG-07 possessed resistance genes *strA, strB, tetA,* and *tetR* on a plasmid, whereas DG-06 possessed the same resistance genes within its chromosome according to MOB_Suite.

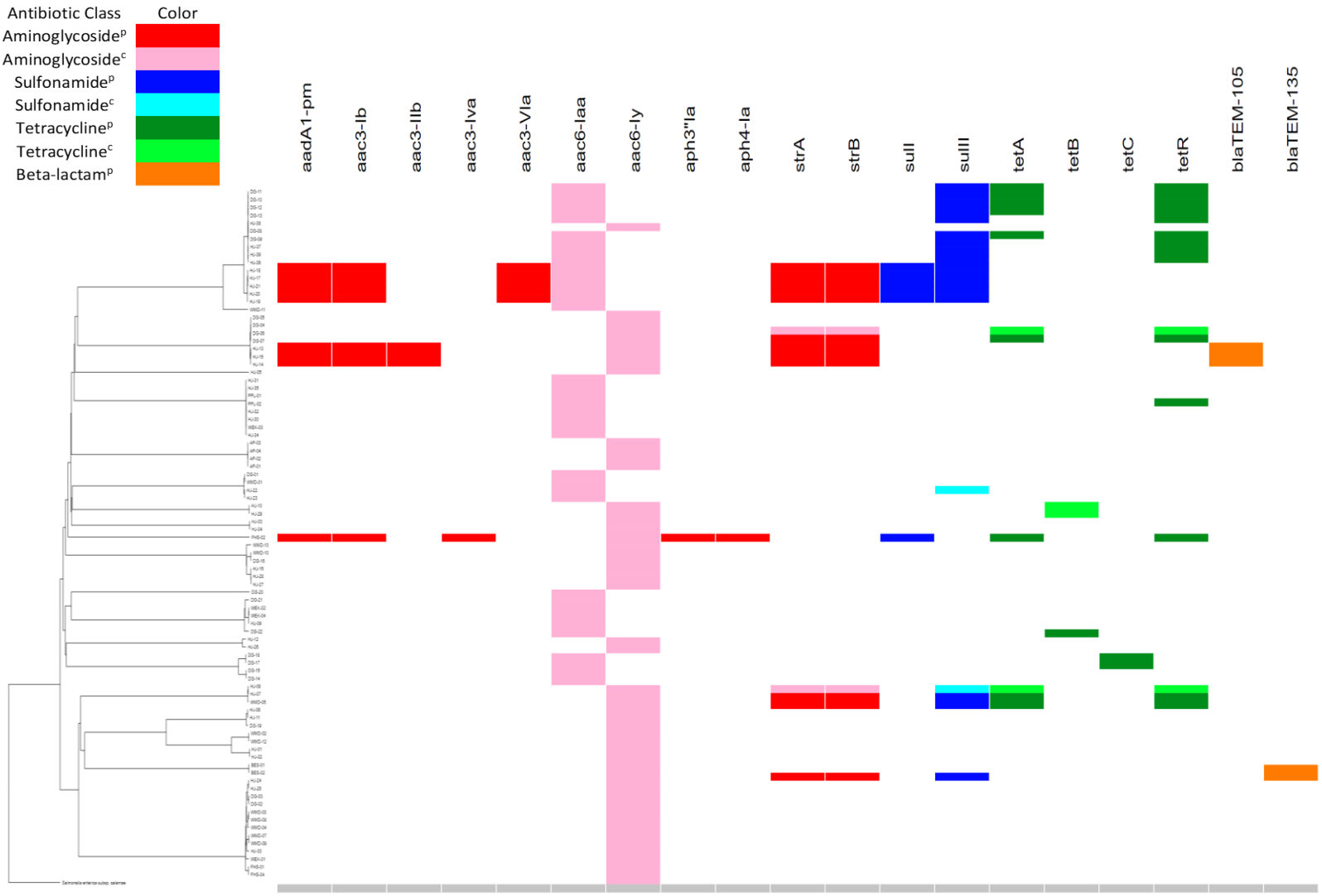


Figure 10: Antibiotic resistance genes identified in the 88 *S. enterica* collected. A neighbor-joining GrapeTree generated using cgMLST data was exported and uploaded to Phandango along with ARG metadata (Hadfield *et a.*, 2018). All gene predictions were performed with ABRicate using the ARG-ANNOT database. Red/pink denotes aminoglycoside resistance genes, blue/light blue denotes sulfonamide resistance genes, green/light green denotes tetracycline resistance genes, and orange denotes beta-lactam resistance genes. Dark colors (red, blue, green, and orange) represent plasmid-borne ARGs. Light colors (Pink, light blue, and light green) represent chromosome-encoded ARGs

[P]: Present on a plasmid

[C]: Present in the chromosome

Antibiotic minimum inhibitory concentrations (MICs) were determined for isolates HJ-01 – HJ-26 using Sensititre™ NARMS gram-negative plates. Most isolates tested that possess ARGs (Figure 10) were found to be resistant to the corresponding antibiotics (Table 2). An exception to this was HJ-03, which was resistant to streptomycin, sulfisoxazole, and tetracycline despite not having detected ARGs for any of these antibiotics. Another isolate, HJ-21, possessed resistance genes that could confer resistance to aminoglycosides, sulfonamides, and tetracycline, yet it displayed no phenotypic resistance when tested.

Table 2. Isolates that displayed phenotypic resistance to one or more antibiotics. HJ-15 has intermediate susceptibility to amoxicillin and clavulanic acid. HJ-20 has intermediate susceptibility to gentamycin. Numbers are only reported when MICs exceed the resistance breakpoint of that antibiotic for *Salmonella enterica*. Dashes indicate susceptibility to an antibiotic

		Antibiotic[a]					
Isolate	Phenotypic resistance	Str	Sul	Tet	Amp	Gen	Amo/Cla
HJ-03	Str, Sul, Tet	> 64	> 256	> 32	-	-	-
HJ-07	Str, Sul, Tet	> 64	> 256	> 32	-	-	-
HJ-08	Str, Sul, Tet	> 64	> 256	> 32	-	-	-
HJ-10	Tet	-	-	> 32	-	-	-
HJ-13	Amp, Gen, Str, Tet	> 64	-	> 32	> 32	> 16	-
HJ-14	Amp, Gen, Str, Tet	> 64	-	> 32	> 32	> 16	-
HJ-15	Amp, Gen, Str, Tet	> 64	-	> 32	> 32	> 16	= 16[b]
HJ-17	Gen, Str, Sul, Tet	> 64	> 256	> 32	-	> 16	-
HJ-18	Gen, Str, Sul, Tet	> 64	> 256	> 32	-	> 16	-
HJ-19	Gen, Str, Sul, Tet	> 64	> 256	> 32	-	> 16	-
HJ-20	Gen, Str, Sul, Tet	> 64	> 256	> 32	-	= 16[b]	-

[a]: Str- Streptomycin, Sul- Sulfisoxazole, Tet- Tetracycline, Amp- Ampicillin, Gen- Gentamycin, Amo/Cla- Amoxicillin and clavulanic acid

[b]: Isolates that display intermediate susceptibility to an antibiotic

As described previously, the program MOB_Suite was used to identify putative plasmids. When available, hybrid-assembly graphs were used to validate the predictions of MOB_Suite. ABRicate was then used to check predicted plasmids for ARGs if ARGs were identified within the whole genome of an isolate. In total, 26 plasmids were found to contain one or more ARGs of the 31 isolates with ARGs present in their genomes. Many of these plasmids are large and likely transmissible (Figures 11-14). Other plasmids identified were small, but still potentially mobilizable (Figures 15-16), whereas plasmids such as pHJ-07 lacked the machinery required to mobilize or conjugate (Figure 17).

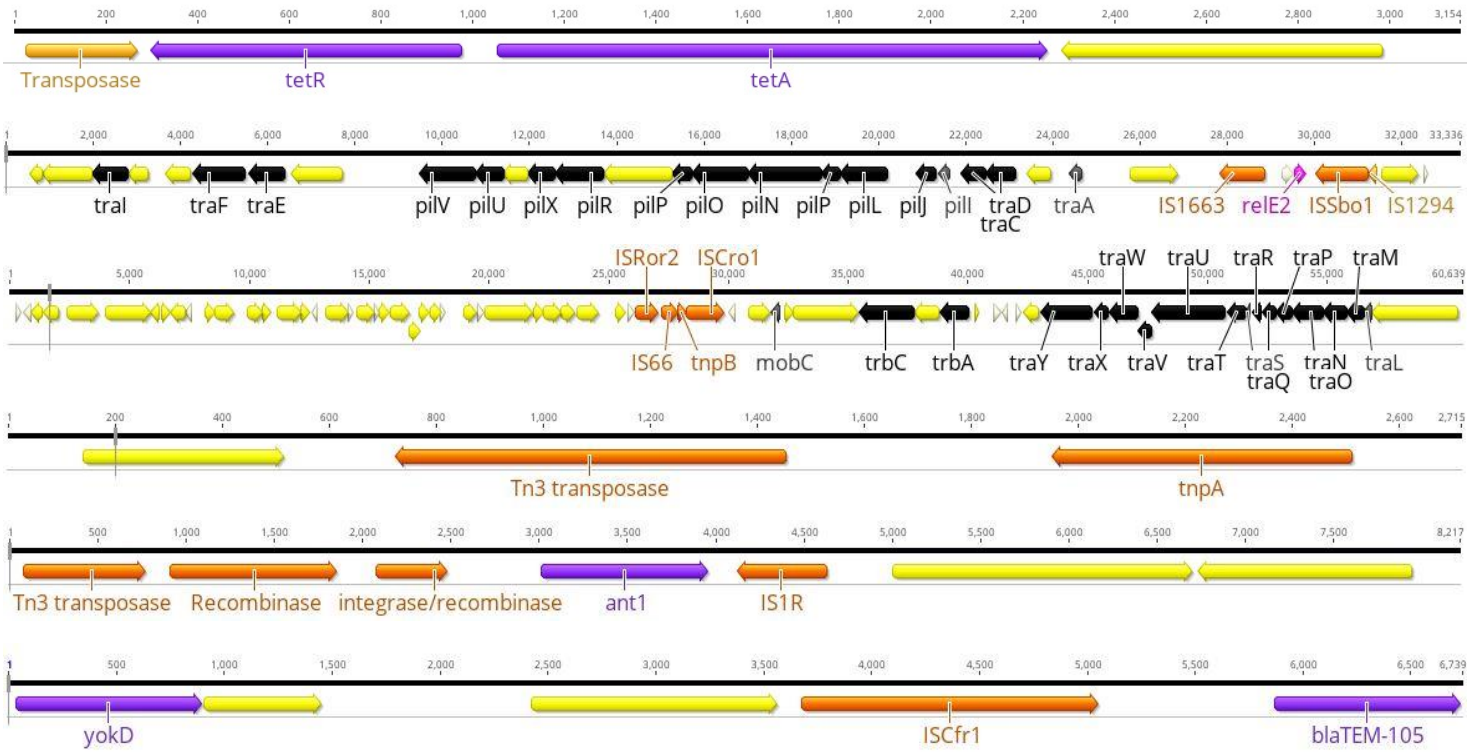


Figure 11: Plasmid map of pHJ-15, an IncI1 plasmid identified in HJ-15 by MOB_Suite. **Pink** arrows represent toxin/antitoxin genes. **Orange** arrows represent transposon/integron-associated genes. **Yellow** arrows represent genes hypothetical and miscellaneous genes. **Purple** arrows represent antibiotic resistance genes identified by both ABRicate and Prokka. **Black** arrows represent genes associated with mobility (including *mob*, *tra*, and *pil*, *trb*, *vir*, and *mbe* genes). Light colors of genes are those that had an alignment score below 200 using BLASTx

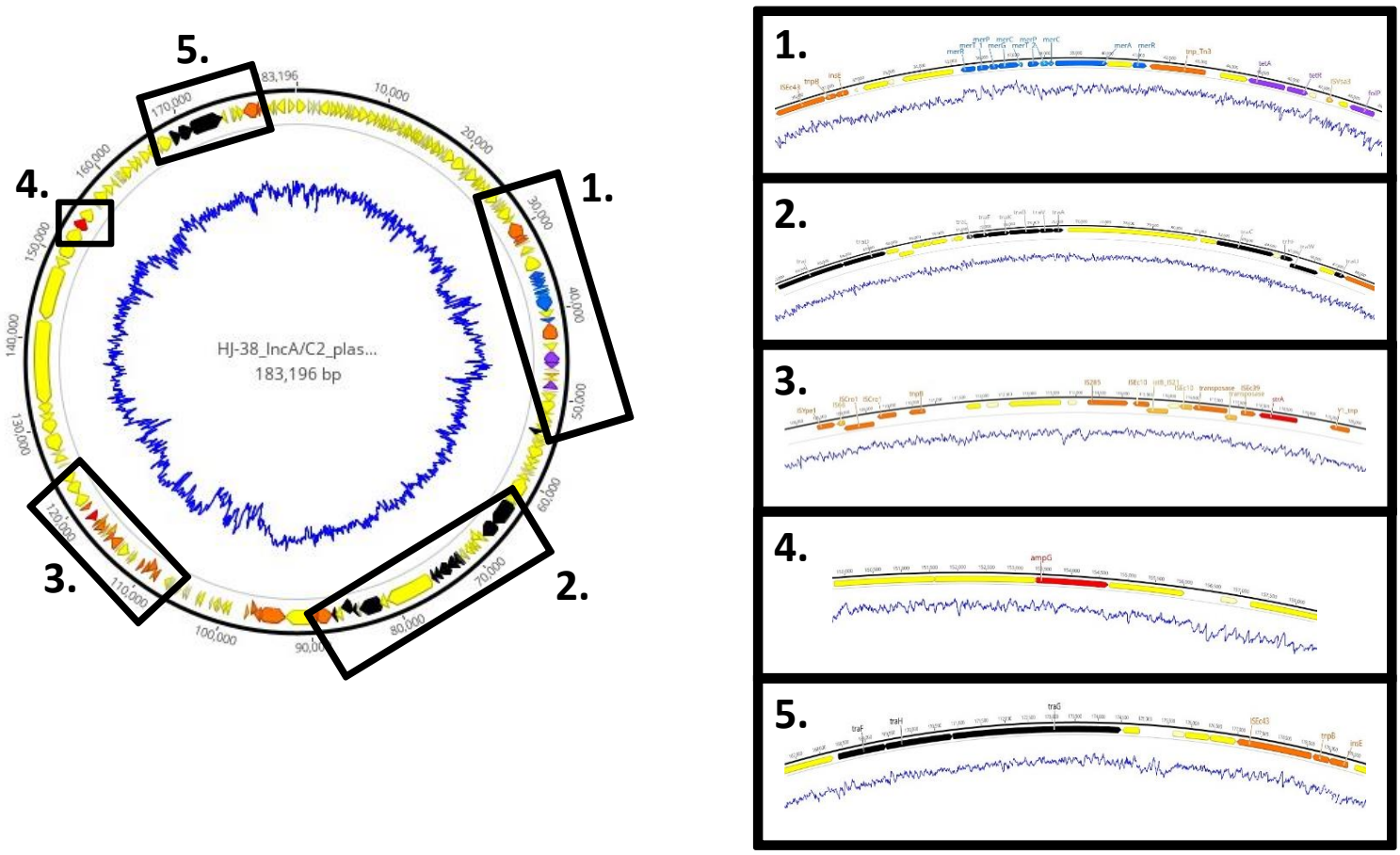


Figure 12: Plasmid map of pHJ-38.1, an IncA/C2 plasmid identified in HJ-38 by MOB_Suite. **Red** arrows represent antibiotic resistance genes only identified by Prokka. **Purple** arrows represent antibiotic resistance genes identified by both ABRicate and Prokka. **Pink** arrows represent toxin/antitoxin genes. **Orange** arrows represent transposon/integron-associated genes. **Yellow** arrows represent genes hypothetical and miscellaneous genes. **Blue** arrows represent heavy metal resistance genes. **Black** arrows represent genes associated with mobility (including *mob*, *tra*, and *pil*, *trb*, *vir*, and *mbe* genes). Light-colored genes are those that had an alignment score below 200 using BLASTx. Inner blue ring represents percent G+C

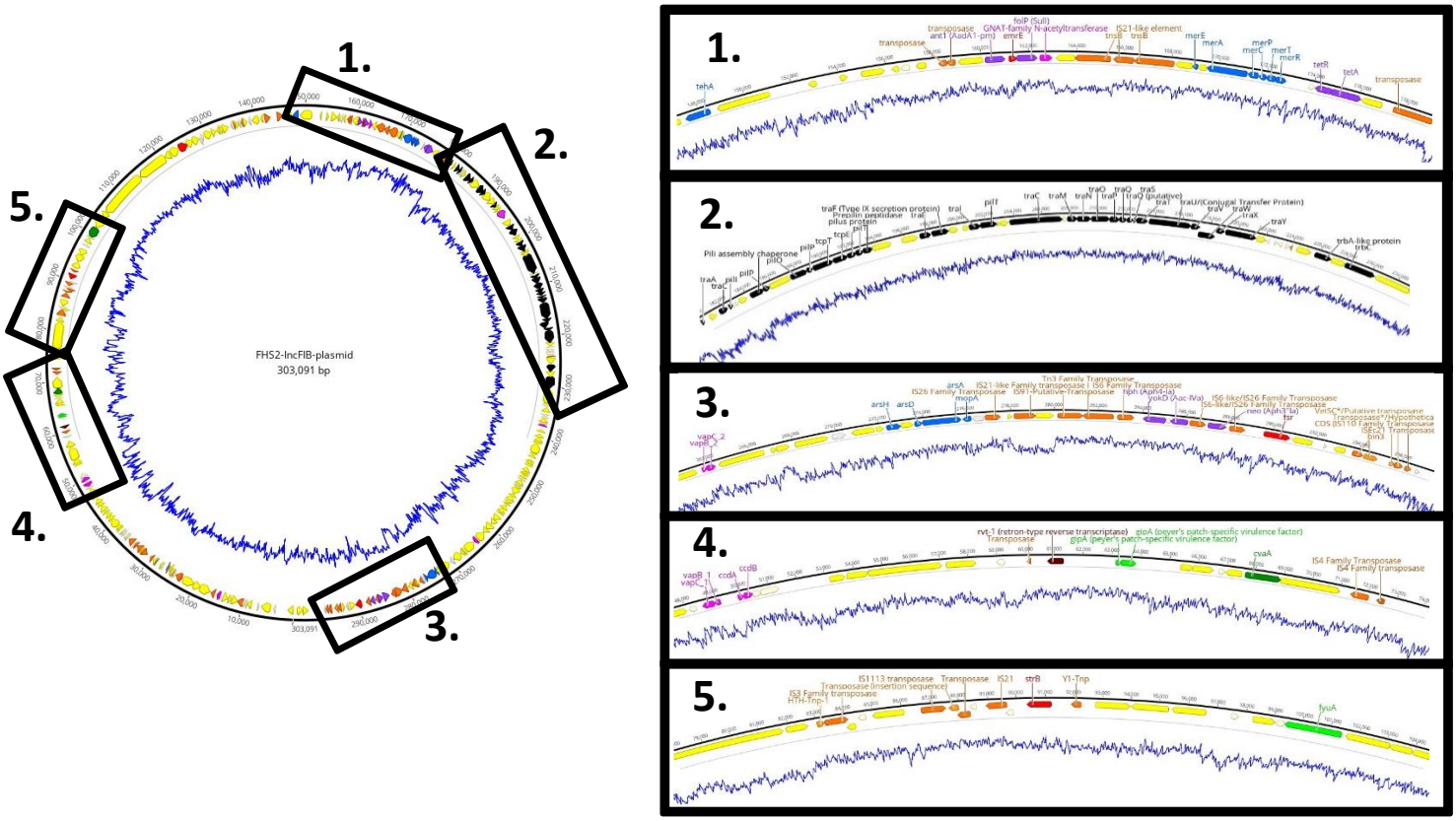


Figure 13: Plasmid map of pFHS-02, an IncFIB plasmid identified in FHS-02 by MOB_Suite. **Red** arrows represent antibiotic resistance genes only identified by Prokka. **Purple** arrows represent antibiotic resistance genes identified by both ABRicate and Prokka. **Pink** arrows represent toxin/antitoxin genes. **Orange** arrows represent transposon/integron-associated genes. **Yellow** arrows represent genes hypothetical and miscellaneous genes. **Green** arrows represent virulence genes. **Green** arrows represent bactericidal genes. **Blue** arrows represent heavy metal resistance genes. **Black** arrows represent genes associated with mobility (including *mob*, *tra*, and *pil*, *trb*, *vir*, and *mbe* genes). Light-colored genes are those that had an alignment score below 200 using BLASTx. Inner blue ring represents percent G+C

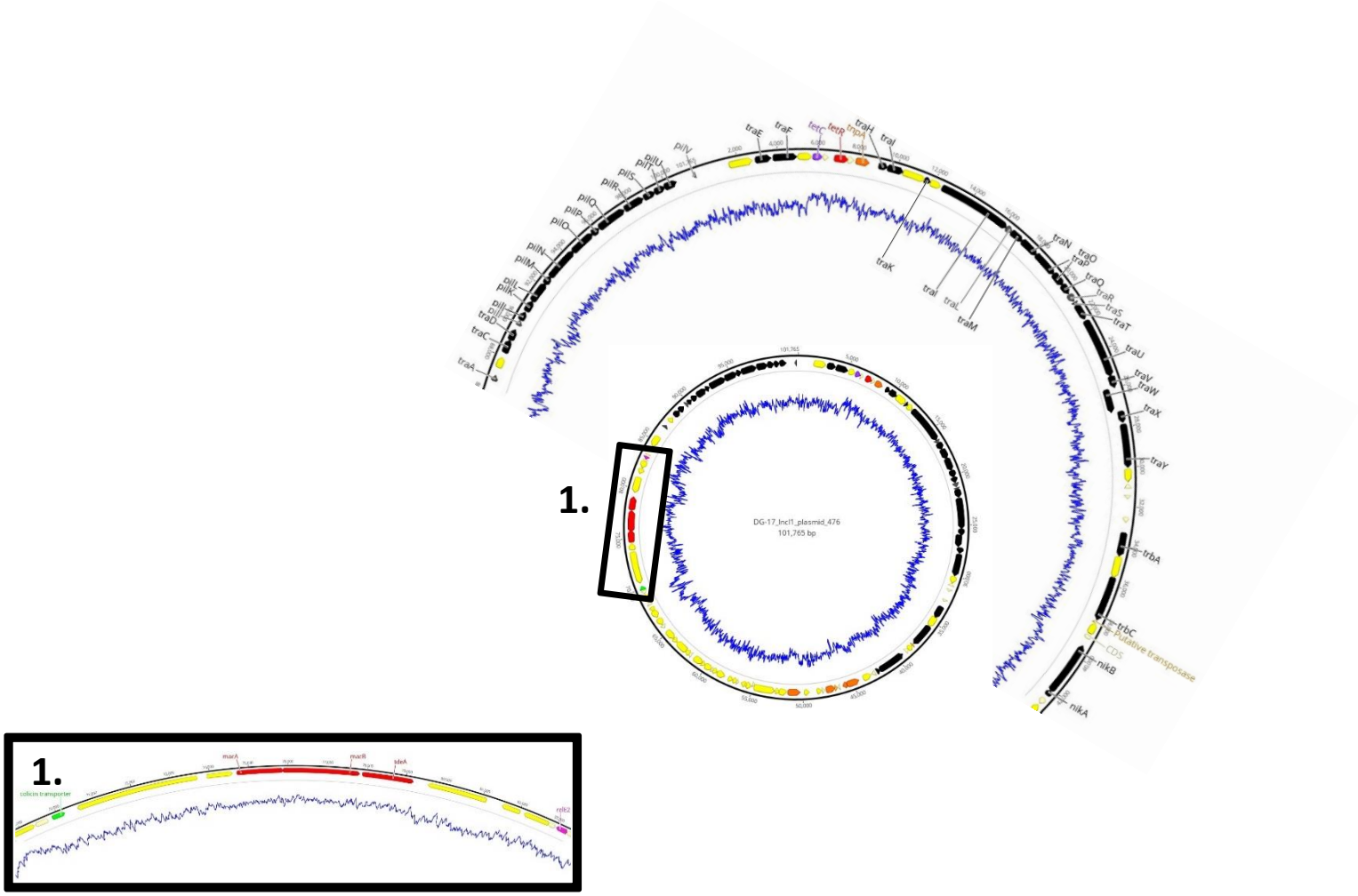


Figure 14: Plasmid map of pDG-17, an IncI1 plasmid identified in DG-17 by MOB_Suite. **Red** arrows represent antibiotic resistance genes only identified by Prokka. **Purple** arrows represent antibiotic resistance genes identified by both ABRicate and Prokka. **Pink** arrows represent toxin/antitoxin genes. **Orange** arrows represent transposon/integron-associated genes. **Yellow** arrows represent genes hypothetical and miscellaneous genes. **Green** arrows represent virulence genes. **Black** arrows represent genes associated with mobility (including *mob*, *tra*, and *pil*, *trb*, *vir*, and *mbe* genes). Light-colored genes are those that had an alignment score below 200 using BLASTx. Inner blue ring represents percent G+C

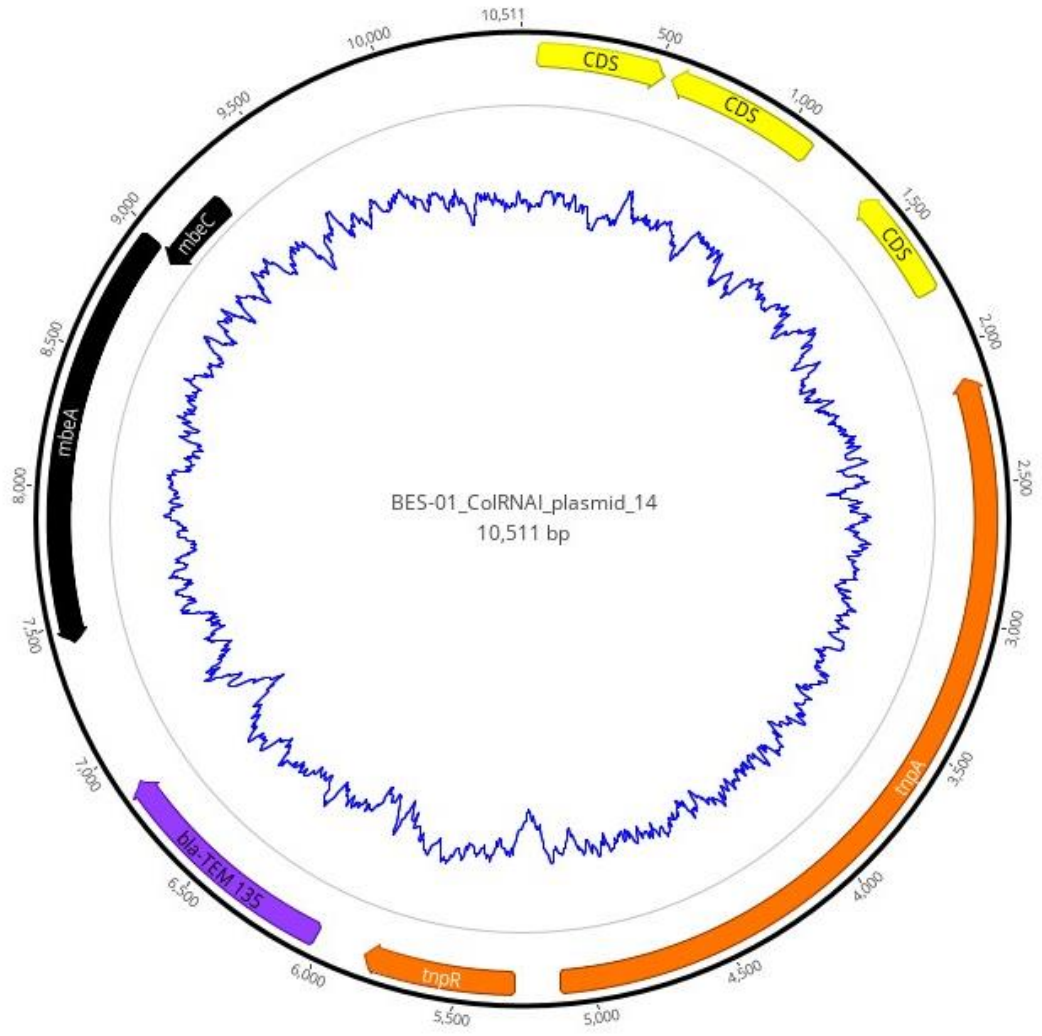


Figure 15: Plasmid map of pBES-01, a ColRNAI plasmid identified in BES-01 by MOB_Suite. **Orange** arrows represent transposon/integron-associated genes. **Yellow** arrows represent genes hypothetical and miscellaneous genes. **Purple** arrows represent antibiotic resistance genes identified by both ABRicate and Prokka. **Black** arrows represent genes associated with mobility (including *mob*, *tra*, and *pil*, *trb*, *vir*, and *mbe* genes). Inner blue ring represents percent G+C

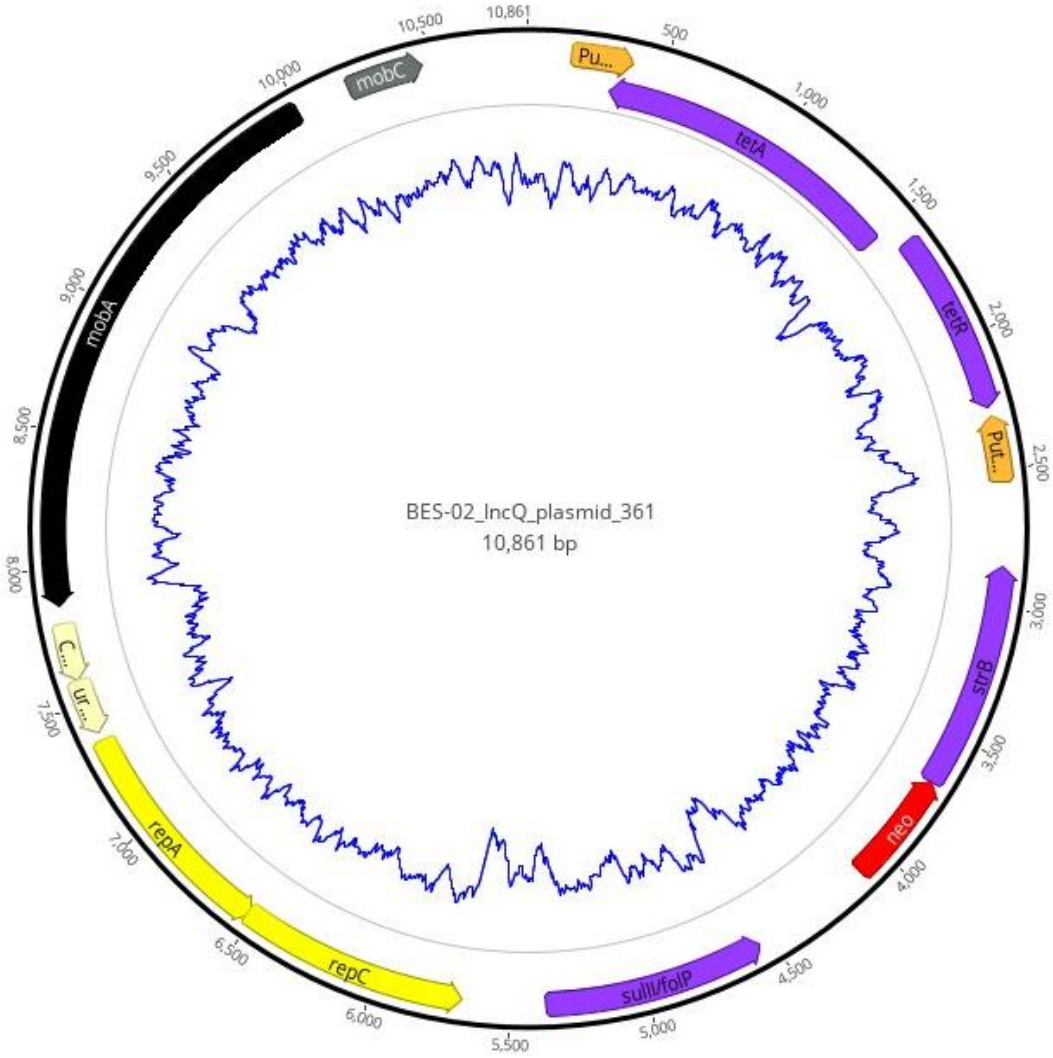


Figure 16: Plasmid map of pBES-02, a ColRNAI plasmid identified in BES-02 by MOB_Suite. **Red** arrows represent antibiotic resistance genes identified only by Prokka. **Purple** arrows represent antibiotic resistance genes identified by both ABRicate and Prokka. **Orange** arrows represent transposon/integron-associated genes. **Yellow** arrows represent genes hypothetical and miscellaneous genes. **Black** arrows represent genes associated with mobility *(including* mob, tra, *and* pil, trb, vir, *and* mbe *genes)*. Light-colored genes are those that had an alignment score below 200 using BLASTx. Inner blue ring represents percent G+C

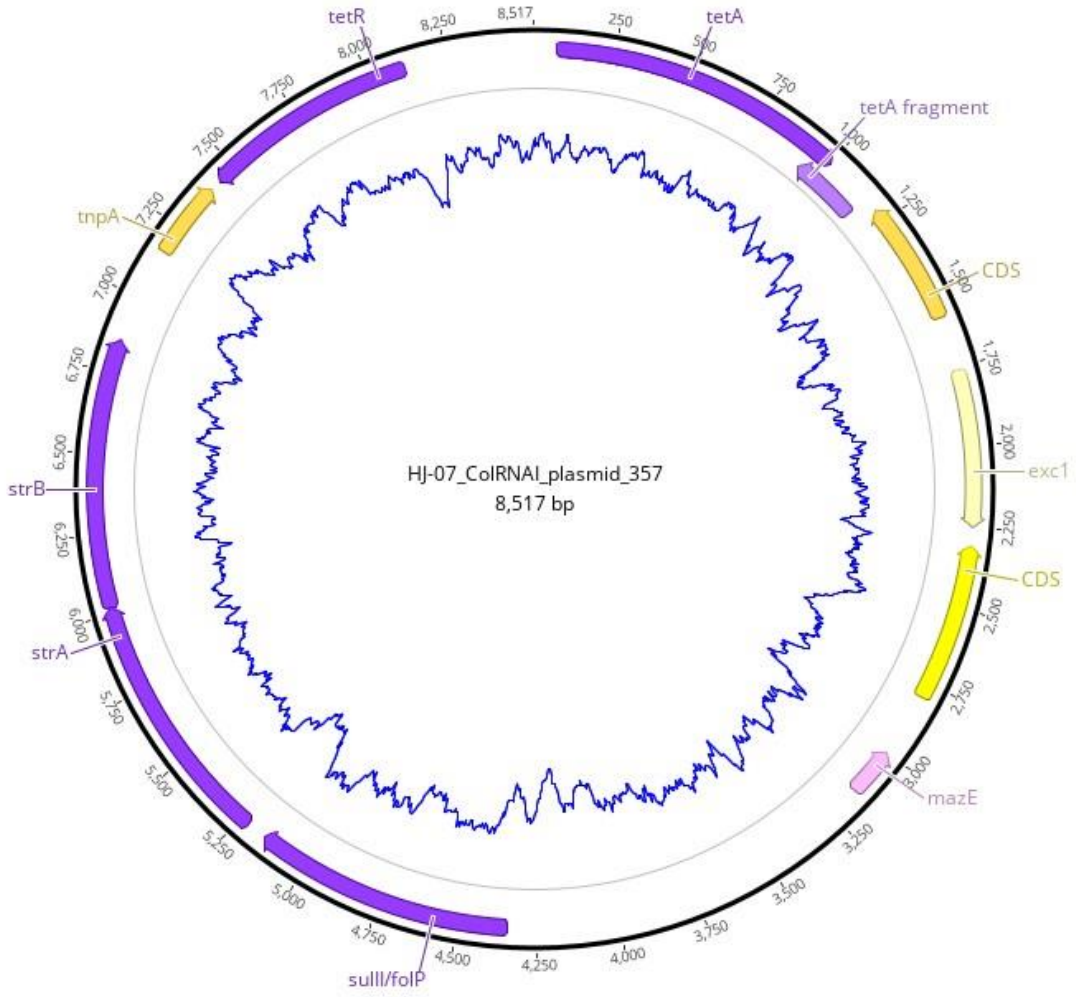


Figure 17: Plasmid map of pHJ-07, a ColRNAI plasmid identified in HJ-07 by MOB_Suite. **Purple** arrows represent antibiotic resistance genes identified by both ABRicate and Prokka. **Orange** arrows represent transposon/integron-associated genes. **Yellow** arrows represent genes hypothetical and miscellaneous genes. Light-colored genes are those that had an alignment score below 200 using BLASTx. Inner blue ring represents percent G+C

Discussion

Sequencing and typing of *Salmonella enterica*

In the age of big data, sequencing platforms continue to improve, as do bioinformatics tools designed to parse valuable information from them. The advancement of 3rd generation sequencing technologies like the MinION™ can aid in the production of high-quality whole genome assemblies by combining long-read sequence data with short-read sequence data (Wick *et a.*, 2017). In this study we showed that hybrid assembly strategies produce better assemblies than using purely short-read data (Figure 4). Assemblies that utilize both short and long-read sequence data produce assemblies with fewer, longer contigs, resulting in a more complete and more accurate representation of the organism's genome. This is especially valuable in the context of a major human pathogen such as *S. enterica*.

For years, *S. enterica* has mainly been characterized using the O and H antigens that determine an isolate's serotype. Recently, it has been proposed that using multi-locus sequence typing (MLST) instead of serotyping allows for more stringent characterization of *S. enterica* isolates (Achtman *et al.*, 2012). Using MLST schemes such as core genome MLST (cgMLST) would allow epidemiologists to better track and respond to outbreaks caused by specific strains of *S. enterica*. A great amount of diversity was found within the 88 isolates collected by our lab (Table 1). Some isolates such as HJ-30, 32, 35, PPL-01, 02, and WEK-03 were found to be identical at the core genome level. This is especially interesting because these isolates were collected at different times and from

different sources. While these isolates all originated from Cook's Creek, the distance between Cook's Creek Park, the source of strains HJ-30, 32, and 35, and Cook's Creek Rt. 11, the source of strains PPL-01, 02, and WEK-03, would be astronomical to a bacterium. One explanation is that there could be a common source for each isolate, presumably upstream of both locations. In any case, since these isolates are essentially identical strains, they likely have not been separated from each other for long. Another might be this strain has established itself within one or both locations, possibly seeding the downstream from the upstream location, for example. All isolates grouped together in clades (Figures 5, 6), likely indicating that each serovar may have a common ancestor.

Clinically-significant serotypes such as Typhimurium and Infantis were isolated from every source tested, except Black's Run which wasn't extensively sampled. Here we define "clinically-significant" as any serotype that was one of the top 20 to cause infections in the US in 2016 (CDC, 2016). Seven of the 19 serotypes our lab identified were members of the top 20 in 2016. The presence of multiple "clinically-significant" serotypes in stream sediment and poultry litter is a cause for greater concern.

Antibiotic resistance and plasmids

Thirty one of the 88 isolates identified contained one or more antibiotic resistance genes. This included all Typhimurium, as well as the one Infantis isolate collected. Of those 31 isolates, 26 of them contained resistance genes exclusively on plasmids (except for aminoglycoside resistance genes *aac6-Iaa* and *aac6-Iy*). Interestingly, some isolates

which are identical at the serotype and practically identical at the core genome levels differ by the presence or absence of a plasmid. This is the case with HJ-07 and H-08; According to MOB_Suite and ABRicate, HJ-07 possesses 4 resistance genes on a small plasmid (Figure 17), whereas HJ-08 possesses 4 resistance genes within its chromosome. Interestingly, there were no mobility genes such as *tra* or *mob* genes present on pHJ-07, suggesting that this plasmid is only vertically transferred. Strain HJ-08 could have lost the plasmid or HJ-07 could have acquired it via transformation from the environment. A highly similar plasmid was also identified by MOB_Suite in WMD-05. Interestingly, WMD-05 was isolated from Cook's Creek Arboretum on 9/30/2018, whereas HJ-07 and HJ-08 were isolated from Pleasant Run on 10/2/2016. This large temporal gap makes finding two isolates, WMD-05 and HJ-07, with nearly identical plasmids strange. Even stranger is that the sources of HJ-07/HJ-08 and WMD-05 are not connected. One explanation for these observations could be a common contamination source shared by both Pleasant Run and Cook's Creek Arboretum, resulting in the same serotype being at two different sites. One caveat to this is that MOB_Suite, like any bioinformatic tool, is not perfect. MOB_Suite is limited by the quality of the assembly which can affect MOB_Suite's ability to detect plasmid-associated genes such as *rep* genes, relaxase genes, and *oriT* sites. Another issue with MOB_Suite is that it can create what are referred to as mosaicisms (Robertson and Nash, 2018). This occurs when plasmid sequences are incorrectly partitioned, resulting in partial, separate plasmid sequences. On the other hand, hybrid assembly has been shown to resolve plasmids, even large ones, incredibly well (George *et al.*, 2017). An example of this can be seen in figure 4, where

in FHS2, a single, contiguous, circular contig in addition to the chromosome was resolved. This contig is pFHS-02 seen in figure 14.

The mere presence of antibiotic resistance genes did not necessarily equate to phenotypic resistance and vice versa. As can been seen in table 2, HJ-03 was resistant to the sulfonamide sulfisoxazole, tetracycline, and streptomycin, but contained no predicted antibiotic resistance genes at all. This could be explained by chromosomal point mutations altering the targets of these antibiotics enough to where the antibiotics have little-to-no effect. By contrast, HJ-21 possessed multiple resistance genes, but displayed no phenotypic resistance (data not shown). This could be due to a lack of expression of these genes, potentially because of a disruption in an upstream promoter.

Sixty-four of the 88 *S. enterica* strains isolated possessed one or more plasmids. These plasmids, some bearing antibiotic resistance genes, spanned a wide range of sizes. Some plasmids were tiny, containing just over 2000bp, and often did not encode antibiotic resistance genes, nor did they encode genes allowing mobilization. Of the plasmids with antibiotic resistance genes, the range of sizes was variable; the smallest was just over 8500bp (Figure 18), and the largest was over 303kb (Figure 14). While many plasmids can be mobilizable, requiring only an *oriT* and a relaxase at minimum, these cannot conjugate without the mate pair formation system of another, often larger plasmid (Smillie *et al.*, 2010). This means that plasmids such as those found in BES-01 and BES-02 which lack such genes would require the aid of a larger plasmid to mobilize to another cell. Plasmids like those in HJ-07 are non-mobilizable and non-transmissible, meaning they are limited to vertical gene transfer with their host unless acquired from the

environment directly via transformation or a conjugative plasmid were to enter the same bacterium.

In contrast, we have identified multiple, large plasmids that are likely to be conjugative. FHS-02's IncF1B, HJ-15's IncI1, and DG-17's IncI1 plasmids contain two different mate pair formation systems. The *tra* Type IV secretion system (T4SS) creates a thick, rigid pilus that is used for conjugation primarily in solid media, whereas *pil* genes encode a thin, flexible pilus responsible for conjugation exclusively in liquid systems (Yoshida *et al.,* 1999). The presence of both these systems suggests these may have a broad host range. For plasmids, the goal is replication. In addition to having multiple antibiotic resistance genes, FHS-02's IncF1B and HJ-38's Inc A/C2 plasmids also possess a suite of heavy metal resistance genes. Both plasmids have genes encoding resistance to mercury, and FHS-02's IncF1B plasmid encodes genes conferring resistance to molybdenum (*mopA*), arsenic (*ars* genes), and tellurite (*tehA*). Possession of these heavy metal resistance genes might help maintain plasmids without the pressure of antibiotics. To further aid in persistence, plasmids can also contain what are known as plasmid addiction systems. These systems are composed of a toxin and an antitoxin, killing the host unless the plasmid is properly replicated (Unterholzner *et al.,* 2013). These systems were found in several of our multidrug resistance plasmids (Figures 12, 14-15). The combination of these features provides many tools that could explain perpetuation and dissemination of these plasmids.

In this study, we attempted to compare endogenous plasmids present in *S. enterica* to exogenous plasmids isolated directly from environmental samples without cultivation. To do this, *S. enterica* isolates were checked for the absence of plasmids by

extracting plasmid DNA from each isolate and running the DNA on a gel after treating DNA with a restriction exonuclease to eliminate chromosomal DNA. The absence of a plasmid is crucial, as plasmids will exclude other plasmids of the same incompatibility group (Novick *et al.*, 1976). HJ-04, a *Salmonella* Uganda, possessed no plasmids and was chosen as a capture host strain for environmental plasmids. HJ-04 was first selected for resistance to rifampicin, an antibiotic whose resistance is rare in environmental bacteria, at a concentration of 500 μg/mL by gradient plating (Brown and Carlton, 1981). This new strain was named HJ4R. To capture plasmids, bacteria from stream sediment and poultry litter were conjugated with HJ4R on a filter and in liquid media. These two conjugation media were used due to the fact fact that some plasmids conjugate more readily in liquids, and others in solids. To select for transconjugants, filter-mated cells were transferred to plates containing both rifampicin and tetracycline (25 μg/mL). Tetracycline has been commonly used by farmers, so we used Tet^R as a selective agent. Three putative plasmids were captured into HJ4R, one from litter and two from stream sediment. The plasmid from litter came from a filter conjugation, and the two from sediment came from liquid conjugations. Due to the nature of the media from which they were isolated, it follows that the plasmids captured likely encode a suite of pilus genes designed for a solid or liquid environment. After plasmid capture, it was found that HJ4R no longer grew at the rate it did without plasmids. Where TSB with HJR4 became turbid after 16-24 hours, the presence of a Tet^R plasmid made HJ4R take up to 5 days for similar levels of growth. DNA extraction was hampered due to low amounts of growth, and plasmid mini-preps designed to isolate plasmid DNA resulted in a very low abundance of these plasmids (data not shown).

Conclusion

The prevalence of "clinically-significant" *Salmonella enterica* in streams and poultry litter reveals a real potential danger. While all *S. enterica* can infect humans, certain serotypes including Typhimurium and Infantis are more often implicated in human infections than others (CDC, 2016). The advent of whole-genome sequencing has allowed scientists to predict outbreaks of clinically-significant *S. enterica*, emphasizing the value of these data in an epidemiological capacity (Zhang *et al.*, 2019). The direct proximity of sampled sites to agricultural operations implicates these sites as reservoirs for the *S. enterica* that were isolated. This is supported by the evidence that identical serotypes and closely related core genome sequence types are often connected with agriculture, especially poultry (Figures 6 and 7). Poultry litter is often re-used after composting, being more cost-effective than purchasing new litter. This process, however, does not eliminate all *S. enterica* and may explain the persistence of *S. enterica* in poultry (Ahmed *et al.*, 2012). Agriculture and other environments have been found to be important reservoirs, not just for *S. enterica*, but also antibiotic resistance genes in general (Klemm *et al.*, 2018).

While *S. enterica* infections are typically self-limiting, septic infections require antibiotic intervention. We identified multiple plasmids conferring genotypic and sometimes phenotypic resistance on their host strains. Plasmids found in HJ-13, HJ-14, and HJ-15, BES-01, and BES-02 each possessed beta-lactamases which could confer resistance to multiple beta-lactams. Resistance to 3rd and 4th generation cephalosporins in *S. enterica* due to the presence beta-lactamase genes on plasmids has been shown

previously (McCollister *et al.*, 2016). These antibiotics are considered last resorts, and beta-lactamase genes on potentially transmissible plasmids is of great concern because other bacteria could become resistant to cephalosporins without direct exposure.

Future studies are required to better elucidate why *S. enterica* persists in these environments, particularly stream sediment which has more direct exposure to weather and other environmental influences that makes it a more dynamic environment than poultry litter. Comparison of plasmids found in *S. enterica* to possible plasmids that could be acquired from the environment (endogenous vs exogenous) would also help reveal the what ARGs are present on mobile elements that *S. enterica* could acquire from these sources. Furthermore, examining mobile elements such as transposons, integrons, and phages is essential to begin fully grasping the network that underlies genetic exchange and resistance gene acquisition.

References

1. Acheson D, Hohmann EL. 2001. Nontyphoidal salmonellosis. Clin Infect Dis. 32:263-269.
2. Achtman M, Wain J, Weill FX, Nair S, Zhou Z, Sangal V, Krauland MG, Hale JL, Harbottle H, Usebeck A, Dougan G, Harrison LH, Brisse S, the *S. enterica* MLST study group. 2012. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. PLOS Pathogen 8:1002776. https://doi.org/10.1371/journal.ppat.1002776.
3. Afgan E, Baker D, Batut B, Beek M, Bouvier D, Čech M, Chilton J, Celments D, Coraor N, Grüning B, Gueler A, Hillman-Jackson J, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible, and collaborative biomedical analyses: 2018 update. Nucl Acid Res. 46:537-544. https://doi.org/10.1093/nar/gky379.
4. Ahmed ZAM, Sedik ZM, Alharery MD, Khalaf MA, Nasr SA, Abdelrahman HA. 2012. Microbial ecology of composting dead poultry and their wastes. Global Vet. 9:683-690. https://doi.org/10.5829/idosi.gv.2012.9.6.66182.
5. Akiyama T, Presedo J, Khan AA. 2013. The *tetA* gene decreases tigecycline sensitivity of *Salmonella enterica* isolates. Int J Antimicrob Agents. 42:133-140.
6. Alikhan N-F, Zhou Z, Sergeant MJ, Achtman M (2018) A genomic overview of the population structure of *Salmonella*. PLoS Genet 14(4): e1007261. https://doi.org/10.1371/journal.pgen.1007261.
7. Alpuche-Aranda CM, RAcoosin EL, Miller SI. 1994. *Salmonella* stimulate macrophage micropinocytosis and persist within spacious phagosomes. J Experim Med. 179:601-608.
8. Alvarez J, Sota M, Vivanco AB, Perales I, Cisterna R, Rementeria A, Garaizar J. 2004. Development of a multiplex PCR technique for detection and epidemiological typing of *Salmonella* in human clinical samples. J Clin Microbiol. 42:1734-1738.
9. Amin AR, Attur MG, Thakker GD, Patel PD, Vyas PR, Patel RN, Patel IR, Abramson SB. 1996. A novel mechanism of action of tetracyclines: effects on nitric oxide synthases. PNAS. 93:14014-14019.
10. Amos GC, Zhang L, Hawkey PM, Gaze WH, Wellington EM. 2014. Functional metagenomic analysis reveals rivers are a reservoir for diverse antibiotic resistance genes. Vet Micrbiol. 171: 441-447.

11. Anderson ES, Ward LR, de Saxe MJ, de Sa JDH. 1977. Bacteriophage-typing designations of *Salmonella Salmonella typhimurium*. Epidemiol and Infect. 78:297-300.
12. Andino A, Hanning I. 2014. *Salmonella enterica*: survival, colonization, and virulence differences among serovars. The Scientific World Journal. 2015:1-16. http://dx.doi.org/10.1155/2015/520179.
13. Andrews WH, Wang H, Jacobson A, Hammack T. 2009. *Salmonella*, Bacteriological analytical manual, 8th ed, vol 1. United States Department of Agriculture, Washington, DC.
14. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
15. Apata DF. 2009. Antibiotic resistance in poultry. Int J Poultry Science. 8:404-408.
16. Ash RJ, Mauck B, Morgan M. 2002. Antibiotic resistance of gram-negative bacteria in rivers, United States. Emerg Infect Dis. 8:713-716.
17. Bailey JK, Pinyon JL, Anatham S, Hall RM. 2010. Commensal *Escherichia coli* of healthy humans: a reservoir for antibiotic-resistance determinants. J Med Microbiol. 59:1331-1339.
18. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19:455-477. https://doi.org/10.1089/cmb.2012.0021.
19. Baquero F, Martínez JL, Cantón R. 2008. Antibiotics and antibiotic resistance in water environments. Curr Opin Biotechnol. 19:260-265.
20. Baron, S. *In* Giannella RA (ed), Medical microbiology, 4th ed, vol. 1. 1996. The University of Texas Medical Branch at Galveston, Galveston, TX.
21. Baucheron S, IMberechts H, Chaslus-Dancia E, Cloeckaert A. 2004. The *AcrB* multidrug transporter plays a major role in high-level fluoroquinolone resistance in *Salmonella enterica* serovar Typhimurium phage type DT204. Microbial Drug Res. 8:281-289.
22. Bäumler AJ, Kusters JG, Stojiljkovic I, Heffron F. 1994. *Salmonella* Typhimurium loci involved in survival within macrophages. Infect and Immun. 62:1623-1630.
23. Blanc-Potard AB, Groisamn EA. 1997. The *Salmonella selC* locus contains a pathogenicity island mediating intramacrophage survival. J EMBO. 16:5376-5385.
24. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114-2120. https://doi.org/10.1093/bioinformatics/btu170.

25. Boyd D, Peters GA, Cloeckaert A, Boumedine KS, Chaslus-Dancla E, Imberechts H, Mulvey MR. 2001. Complete nucleotide sequence of a 43-kilobase genomic island associated with the multidrug resistance region of *Salmonella enterica* serovar Typhimurium DT104 and its identification in phage type DT120 and serovar Agona. J Bacteriol. 183:5725-5732.
26. Braden CR. 2006. *Salmonella enterica* serotype Enteritidis and eggs: a national epidemic in the United States. Clin Infect Dis. 43:512-517.
27. Briganti F, Pierattelli R, Scozzafava A, Supuran CT. 1996. Carbonic anhydrase inhibitors. Part 37. Novel classes of isozyme I and II inhibitors and their mechanism of action. Kinetic and spectroscopic investigations on native and cobalt-substituted enzymes. Eu J Med Chem. 31:1001-1010.
28. Briggs CE, Fratamico PM. 1999. Molecular characterization of an antibiotic resistance gene cluster of *Salmonella* Typhimurium DT104. Antimicrob Agents Chemother. 43:846-849.
29. Brown AC, Grass Je, Richardson LC, Nisler AL, Bicknese AS, Gould LH. 2017. Antimicrobial resistance in *Salmonella* that caused foodborne disease outbreaks: United States, 2003-2012. Epidemiol and Infect. 145:766-774.
30. Brown BJ, Carlton BC. 1981. Gene mutation, p 230, *In* Gerhardt P, Nester EW (ed), Manual of Methods for General Bacteriology. American Society of Microbiology, Washington, DC.
31. Burgess BA, Noyes NR, Bolte DS, Hyatt DR, van Metre DC, Morley PS. 2014. Rapid *Salmonella* detection in experimentally inoculated equine faecal and veterinary hospital environmental samples using commercially available lateral flow immunoassays. Equine Vet. 47:119-122.
32. Cabello FC. 2006. Heavy use of prophylactic antibiotics in aquaculture: a growing problem for human and animal health and for the environment. Environ Microbiol. 8:1137-1144.
33. Carattoli A. 2003. Plasmid-mediated antimicrobial resistance in *Salmonella enterica*. Curr Issues Mol Biol 5:113-122.
34. Carattoli A, Zankari E, García-Fernández A, Larsen MV, Lund O, Villa L, Aarestrup FM, Hasman H. 2014. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob. Agents Chemother. 58:3895-3903.
35. Calomiris JJ, Armstrong JL, Seidler RJ. 1984. Association of metal tolerance with multiple antibiotic resistance of bacteria isolated from drinking water. Appl Environ Microbiol. 47:1238-1242.
36. Center for Disease Control and Prevention. 2002. Outbreak of multidrug-resistant *Salmonella* Newport-United States, January-April 2002. J Amer Med Assoc. 51:545-548.

37. Cheung T, Chu Y, Chu M, Ma C, Yung R, Kam K. 2005. Plasmid-mediated resistance to ciprofloxacin and cefotaxime in clinical isolates of *Salmonella enterica* serotype Enteriditis in Hong Kong. J Antimicro Chemother. 56:586-589.
38. CDC. 2016. National Enteric Disease Surveillance: *Salmonella* Annual Report, 2016, p 5. https://www.cdc.gov/nationalsurveillance/pdfs/2016-Salmonella-report-508.pdf.
39. Cirulli ET, Goldstien DB. 2010. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nature Rev Genet. 11:415-425.
40. Clarke J, Wu HC, Jayasignhe L, Patel A, Reid S, Bayley H. 2009. Continuous base identification for single-molecule nanopore DNA sequencing. Nature Nanotech. 4:265-270.
41. Coburn B, Grassl GA, Finlay BB. 2006. *Salmonella*, the host and disease: a brief review. Immuno. And Cell Bio. 85:112-118.
42. Davis BD. 1987. Mechanisms of bactericidal action of aminoglycosides. Microbiol Rev. 51:341-350.
43. Dechet AM, Scallan E, Gensheimer K, Hoekstra R, Gunderman-King J, Lockett J, Wrigley D, Chege W, Sobel J, Multistate Working Group. 2006. Outbreak of multidrug-resistant *Salmonella enterica* serotype Typhimurium definitive type 104 infection linked to commercial ground beef, Northeastern United States, 2003-2004. Clin Infect Dis. 42:747-752.
44. DeFrancesco KA, Cobbold RN, Rice DH, Besser TE, Hancock DD. 2004. Antimicrobial resistance of commensal *Escherichia coli* from dairy cattle associated with recent multi-resistant salmonellosis outbreaks. Vet Microbiol. 98:55-61.
45. Deiwick J, Salcedo SP, Boucrot E, Gilliland SM, Henry T, Petermann N, Waterman SR, Gorvel JP, Holden DW, Méresse S. 2006. The translocated *Salmonella* effector proteins *SseF* and *SseG* interact and are required to establish an intracellular replication niche. Infect Immun. 74:6965-6972.
46. den Bakker HC, Moreno Switt AI, Govoni G, Cummings CA, RAnieri ML, Degoricja L, Hoelzer K, Rodriguez-Rivera LD, Brown S, Bolchacova E, Furatado MR, Wiedmann M. 2011. Genome sequencing reveals diversification of virulence factor content and possible host adaptation in distinct subpopulations of *Salmonella enterica*. BMC Genom. 12:425.
47. den Bakker HC, Allard MW, Bopp D, Brown EW, Fontana J, Iqbal Z, Kinney A, Limberger R, Musser KA, Shudt M, Strain E, Wiedmann M, Wolfgang WJ. 2014. Rapid whole-genome sequencing for surveillance of *Salmonella enterica* serovar Enteriditis. Emerg Infect Dis. 20:1306-1314.

48. Deng X, Ran L, Wu S, Ke B, He D, Yang X, Zhang Y, Ke C, Klena JD, Yan M, Feng Z, Kan B, Liu X, Mikoleit M, Varma JK. 2012. Laboratory-based surveillance of non-typhoidal *Salmonella* infections in Guangdong Province, China. Foodborne Pathog Dis. 9:305-312.
49. Diarra MS, Delaquis P, Rempel H, Bach S, Harlton C, Aslam M, Pritchard J, Topp E. 2014. Antibiotic resistance and diversity of *Salmonella enterica* serovars associated with broiler chickens. J Food Protec. 77:40-49.
50. Diarrassouba F, Diarra MS, Bach S, Delaquis P, Pritchard J, Topp E, Skura BJ. 2007. Antibiotic resistance and virulence genes in commensal *Escherichia coli* and *Salmonella* isolates from commercial broiler chicken farms. J Food Protec. 70:1316-1327.
51. Di Conza J, Ayala JA, Power P, Mollerach M, Gutkind G. 2002. Novel class 1 integron (Ins21) carrying $bla_{\text{CTX-M-2}}$ in *Salmonella enterica* serovar Infantis. Antimicrob. Agents Chemother. 46:2257-2261.
52. Doumith M, Godbole G, Ashton P, Larkin L, Dallman T, Day Ma, Day Mi, Muller-Pebody B, Ellington MJ, de Pinna E, Johnson AP, Hopkins KL, Woodford N. 2016. Detection of the plasmid-mediated *mcr-1* gene conferring colistin resistance in human and food isolates of *Salmonella enterica* and *Escherichia coli* in England and Wales. J Antimicrob Chemother. 71:2300-2305.
53. Dutil L, Irwin RJ, Finley R, Ng LK, Avery BP, Boerlin P, Bourgault AM, Cole L, Daignault D, Desruisseau A, Demczuk W, Hoang L, Horsman GB, Ismail J, Jamieson FB, Maki A, Pacagnella A, Pillai DR. 2010. Ceftiofur resistance in *Salmonella enterica* serovar Heidelberg from chicken meat and humans, Canada. Emerg Infect Dis. 16:48-54.
54. Edel W, Kampelmacher EH. 1969. *Salmonella* isolation in nine European laboratories using a standaradized technique. Bull World Health Organ. 41:297-306.
55. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjorson K, Chauduri B, Christians F, Ciecero R, Clark S, Dalal R, Dewinter A, Dixon J, Foguet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundguist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. 2009. Real-time DNA sequencing from single polymerase molecules. Science. 323:133-138.
56. Egorova S, Timinouni M, Demartin M, Granier SA, Whichard JM, Sangal V, Fabre L, Deluané A, Pardos M, Millemann Y, Espié E, Achtman M, Grimont PAD, Weill FX. 2008. Ceftriaxone-resistant *Salmonella enterica*

serotype Newport, France. Emerg Infect Dis. 14:954-957. type III effectors in *Salmonella enterica* subspecies 1 serovar Typhimurium. Infect Genet Evol. 5:1-9.

57. Ehrbar K, Hardt WD. 2010. Bacteriophage-encoded type III effectors in *Salmonella enterica* subspecies 1 serovar Typhimurium. Infect Genet Evol. 5:1-9
58. Esiobu N, Armenta L, Ike J. 2010. Antibiotic resistance in soil and water environments. Inter J Environ Health Res. 12:133-144.
59. Fábrega A and Vila J, 2013. *Salmonella enterica* serovar typhimurium skills to succeed in the host: virulence and regulation. Clin Microbiol Rev 26:308-341.
60. Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 32:2047-3048. https://doi.org/10.1093/bioinformatics/btw354.
61. Fierer J, Guiney DG. 2001. Diverse virulence traits underlying different clinical outcomes of *Salmonella* infection. J Clin Invest. 107:775-780.
62. Figueira R, Holden DW. 2012. Function of the *Salmonella* pathogenicity island 2 (SPI-2) type III secretion system effectors. Microbiol. 158:1147-1161.
63. Finley RL, Collignon P, Larsson DG, McEwen SA, Li XZ, Gaze WH, Reid-Smith R, Timinouni M, Graham DW, Topp E. 2013. The scourge of antibiotic resistance: the important role of the environment. Clin Infect Dis. 57:704-710.
64. Foster JW, Hall HK. 1991. Inducible pH homeostasis and the acid tolerance response of *Salmonella* Typhimurium. J Bacteriol. 173: 5129-5135.
65. Foster ED, Deardorff A. 2017. Open Science Framework (OSF). J Med Libr Assoc. 105:203-206. https://doi.org/10.5195/jmla.2017.88.
66. Francis CL, Ryan TA, Jones BD, Smith SJ, Falkow S. 1993. Ruffles induced by *Salmonella* and other stimuli direct micropinocytosis of bacteria. Nature. 364:639-642. (Letter.)
67. Frye JG, Jackson CR. 2013. Genetic mechanisms of antimicrobial resistance identified in *Salmonella enterica*, *Escherichia coli*, and *Enteroccocus* spp. Isolated from U.S. food animals. Front Microbiol. 4:135.
68. Galán JE. 1999. Interaction of *Salmonella* with host cells through centisome 63 type III secretion system. Curr Opin Microbiol. 2:46-50.
69. Gay K, Robicsek A, Strahilevitz J, Park CH, Jacoby G, Barrett TJ, Medalla F, Chiller TM, Hooper DC. 2006. Plasmid-mediated quinolone resistance in non-typhi serotypes of *Salmonella enterica*. Clin Infect Dis. 43:297-304.

70. George S, Pankhusrt L, Hubbard A, Votintseva A, Stoesser N, Sheppard AE, Mathers A, Norris R, Navickaite I, Eaton C, Iqbal Z, Crook DW, Phan HTT. 2017. Resolving plasmid structures in *Enterobacteriaceae* using the MinION nanpore sequencer: assessment of MinION and MinION/Illumina hybrid data assembly approaches. Microb Genom. 3:e000118. https://doi.org/10.1099/mgen.0.000118.
71. Ghosh S, LaPara TM. 2007. The effects of subtherapeutic antibiotic use in farm animals on the proliferation and persistence of antibiotic resistance among soil bacteria. Multidiscip J Microbial Ecol. 1:191-203.
72. Gilliver MA, Bennett M, Begon M, Hazel SM, Hart A. 1999. Enterobacteria: antibiotic resistance found in wild rodents. Nature. 401:233-234.
73. Ginocchio CC, Rahn K, Clarke RC, Galán JE. 1997. Naturally occurring deletions in the centisome 63 pathogenicity island of environmental isolates of *Salmonella* spp. Infect Immun. 65:1267-1272.
74. Glenn TC. 2011. Field guide to next-generation DNA sequencers. Mol Ecol Resour. 11:759-769.
75. Goñi-Urriza M, Capdepuy M, Arpin C, Raymond N, Caumette P, Quentin C. 2000. Impact of an urban effluent on antibiotic resistance of riverine *Enterobacteriaceae* and *Aeromonas* spp. Appl Environ Microbiol. 66:125-132.
76. Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz M, McCombie WR. 2015. Oxford Nanopore sequencing, hybrid error correction, and *de novo* assembly of a eukaryotic genome. Genome Res. 25:1750-1756.
77. Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 17:333-351.
78. Gordon MA, Graham SM, Walsh AL, Wilson L, Phiri A, Molyneux E, Ziljstra EE, Heyderman RS, Hart CA, Molyneux ME. 2008. Epidemics of invasive *Salmonella enterica* serovar Enteriditis and *S. enterica* serovar Typhimurium infection associated with multidrug resistance among adults and children in Malawi. Clin Infect Dis. 46:963-969.
79. Grüning B, Dale R, Sjödin A, Chapman BA, Rowe J, Tomkins-Tinch CH, Valieris R, Köster J, The Bioconda Team. 2018. Bioconda: sustainable and comprehensive software distribution for the life sciences. Nature Methods. 15:475-476.
80. Gulig PA, Danbara H, Guiney DG, Lax AJ, Norel F, Rhen M. 1993. Molecular analysis of *spv* virulence genes of the *Salmonella* virulence plasmids. Molec Microbiol. 7:825-830.

81. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf, M, Landraud L, Rolain JM. 2014. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother. 58:212-220. https://doi.org/10.1128/AAC.01310-13.
82. Gurevich A, Saveliev V, Vyanhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29:1072-1075. https://doi.org/10.1093/bioinformatics/btt086.
83. Haack SK, Duris JW, Kolpin DW, Fogarty LR, Johnson HE, Gibson KE, Focazio M, Schawb KJ, Hubbard LE, Foreman WT. 2015. Genes indicative of zoonotic and swine pathogens are persistent in stream water and sediment following a swine manure spill. Appl Environ Micrbiol. 81:3430-3441.
84. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aaenensen DM, Harris SR. 2018. Phandango: an interactive viewer for bacterial population genomics. Bioinformatics. 34:292-293. https://doi.org/10.1093/bioinformatics/btx610.
85. Haley BJ, Kim SW, Pettengill J, Luo Y, Karns JS, Van Kessel JAS. 2016. Genomic and evolutionary analysis of two *Salmonella enterica* serovar Kentucky sequence types isolated from bovine and poultry sources in North America. PLOS One. 11:e0161225. https://doi.org/10.1371/journal.pone.0161225.
86. Helinski DR. 2004. Introduction to plasmids: a selective view of their history, p 1-21, *In* Funnell BE, Phillips GJ (ed), Plasmid Biology. ASM Press, Washington DC.
87. Hendriksen RS, Vieira AR, Karlsmose S, Lo Wo Fong DM, Jensen AB, Wegener HG. 2011. Global monitoring of *Salmonella* serovar distribution from the world health organization global foodborne infections network country data bank: results of quality assured laboratories from 2001 to 2007. Foodborne Pathog Dis. 8:887-900.
88. Herrick JB, Haynes R, Heringa S, Brooks SM, Sobota LT. 2014. Coselection for resistance to multiple late-generation human therapeutic antibiotics encoded on tetracycline resistance plasmids captured from uncultivated stream and soil bacteria. J Appl Microbiol. 117:380-389.
89. Heuer H, Schmitt H, Smalla K. 2011. Antibiotic resistance gene spread due to manure application of agricultural fields. Curr Opin Microbiol. 14:236-243.
90. Hirsch R, Ternes T, Haberer K, Kratz KL. 1999. Occurrence of antibiotics in the aquatic environment. The Science Total Environ. 255:109-118.
91. Ho TD, Figueroa-Bossi N, Wang M, Uzzau S, Bossi L, Slauch JM. 2002. Identification of *GtgE*, a novel virulence factor encoded on the gifsy-2

bacteriophage of *Salmonella enterica* serovar Typhimurium. J Bacteriol. 184:5234-5239.

92. Hoffmann, S. 2017. Data from "USDA foodborne illness cost calculator." United States Department of Agriculture Economic Research Service. https://www.ers.usda.gov/webdocs/DataFiles/48464/Salmonella.xlsx?v=41919.
93. Hoffmann M, Zhao S, Pettengill J, Luo Y, Monday SR, Abbott J, Ayers SL, Cinar HN, Muruvanda T, Li C, Allard MW, Whichard J, Meng J, Brown EW, McDermott PF. 2014. Comparative genomic analysis and virulence differences in closely related *Salmonella enterica* serotype Heidelberg isolated from humans, retail meats, and animals. Gen Biol Evol. 6:1046-1068.
94. Holt PS, Geden CJ, Moore RW, Gast RK. 2007. Isolation of *Salmonella enterica* serovar Enteriditis from houseflies (*Musca domestica*) found in rooms containing *Salmonella* serovar Enteriditis-challenged hens. Appl Environ Microbiol. 73:6030-6035.
95. Horby PW, O'Brien SJ, Adak GK, Graham C, Hawker JI, Hunter P, Lane C, Lawson AJ, Mitchell RT, Reacher MH, Threlfall EJ, Ward LR. 2003. A national outbreak of multi-resistant *Salmonella enterica* serovar Typhimurium definitive phage type (DT) 104 associated with consumption of lettuce. Epidemiol and Infect. 130:169-178.
96. Jackson BR, Griffin PM, Cole D, Walsh KA, Chai SJ. 2013. Outbreak-associated *Salmonella enterica* serotypes and food commodities, United States, 1998-2008. Emerg Infect Dis. 19:1239-1244.
97. Jain R, Rivera MC, Lake JA. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. PNAS. 96:3801-3806.
98. Jones TF, Ingram LA, Cieslak PR, Vugia DJ, Tobin-D'Angelo M, Hurd S, Medus C, Cronquist A, Angulo FJ. 2008. Salmonellosis outcomes differ substantially by serotype. J Infect Dis. 198:109-114.
99. Jørgensen F, Bailey R, Williams S, Henderson P, Wareing DRA, Bolton FJ, Frost JA, Ward L, Humphrey TJ. 2002. Prevalence and numbers of *Salmonella* and *Campylobacter* spp. on raw, whole chickens in relation to sampling methods. Int J of Food Microbiol. 76:151-164.
100. Jorgensen JH, Turnidge JD. 2015. Susceptibility test methods: dilution and disk diffusion methods, p 1253-1273 *In* Jorgensen J, Pfaller M, Carroll K, Funke G, Landry M, Richter S, Warnock D (ed), *Manual of Clinical Microbiology, Eleventh Edition*. ASM Press, Washington, DC.
101. Kinde Hailu, Castellan DM, Kass PH, Ardans A, Cutler G, Breitmeyer RE, Bell DD, Ernst RA, Kerr DC, Herbert EL, Willoughby D, Riemann HP, Snowdon JA, Kuney DR. 2004. The occurrence and distribution of

Salmonella Enteriditis and other serovars on California egg layig premises: a comparison of two sampling methods and two culturing techniques. Avian Dis. 48:590-594.

102. Kingston DJ. 1981. A comparison of culturing drag swabs and litter for identification of infections with *Salmonella* spp. In commercial chicken flocks. Avian Dis. 25:513-516.
103. Klemm EJ, Wong VK, Dougan G. 2018. Emergence of dominant multidrug-resistant bacterial clades: lessons from history and whole-genome sequencing. PNAS. 115:12872-12877. https://www.pnas.org/cgi/doi/10.1073/pnas.1717162115.
104. Landegren U, Kaiser R, Sanders J, Hood L. 1988. A ligase-mediated gene detection technique. Science. 241:1077-1080.
105. Latasa C, Roux A, Toledo-Arana A, Ghigo JM, Gamazo C, Penadés JR, Lasa I. 2005. BapA, a large secreted protein required for biofilm formation and host colonization of *Salmonella enterica* serovar Enteriditis. Molec Microbiol. 58:1322-1339.
106. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ. 2015. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomolec Detec Quant.3:1-8
107. Leekitcharoenphon P, Hendriksen RS, Le Hello S, Weill FX, Baggesen DL, Jun SR, Ussery DW, Lund O, Crook DW, Wilson DJ, Aaestrup FM. 2016. Global genomic epidemiology of *Salmonella enterica* serovar Typhimurium DT104. Appl Environ Microbiol. 82:2516-2526.
108. Leekitcharoenphon P, Nielsen EM, Kaas RS, Lund O, Aarestrup FM. 2014. Evaluation of whole genome sequencing for outbreak detection of *Salmonella enterica*. PLOS One 9:e87991. https://doi.org/10.1371/journal.pone.0087991.
109. Lekunberri I, Balcázar JL, Borrego CM. 2017. Detection and quantification of the plasmid-mediated *mcr-1* gene conferring colistin resistance in wastewater. Int J Antimicrob Agents. 50:734-736.
110. Levings RS, Lightfoot D, Partridge SR, Hall RM, Djordjevic SP. 2005. The genomic island SGI1, containing the multiple antibiotic resistance region of *Salmonella enterica* serovar Typhimurium DT104 or variants of it, is widely distributed in other *S. enterica* serovars. 2005. J Bacteriol. 187: 4401-4409.
111. Levy SB. 2001. Antibiotic resistance: consequences of inaction. Clin Infect Dis. 33: S124-S129.
112. Levy SB. 2002. The antibiotic paradox: how the misuse of antibiotics destroys their curative powers, 2nd ed, vol 1. Perseus Publishing Services, Cambridge, MA.

113. Levy SB, Fitzgerald GB, Macone AB. 1976. Spread of antibiotic-resistant plasmids from chicken to chicken and from chicken to man. Nature. 260:40-42.
114. Levy SB, Marshall B. 2004. Antibacterial resistance worldwide: causes, challenges, and responses. Nature Med. 10:122-129.
115. Libby SJ, Lesnick M, Hasegawa P, Weidenhammer E, Guiney DG. 2000. The *Salmonella* virulence plasmid *spv* genes are required for cytopathology in human monocyte-derived macrophages. Cell Microbiol. 2:49-58.
116. Maka Ł, Popowska M. 2016. Antimicrobial resistance of *Salmonella* spp. isolated from food. Rocz Panstw Zakl Hig. 67:343-358.
117. Maloy S, Edwards R. 1999. *Salmonella* information. www.salmonella.org.
118. Malorny B, Hoorfar J, Bunge C, Helmuth R. 2003. Multicenter validation of the analytical accuracy of *Salmonella* PCR: towards an international standard. Applied and Env Microbiol. 69:290-296. https://doi.org/10.1128/AEM.69.1.290-296.2003.
119. Marcus SL, Brumell JH, Pfeifer CG, Finlay BB. 2000. *Salmonella* pathogenicity islands: big virulence in small packages. Microbes and Infect. 2:145-156.
120. Martínez JL. 2008. Antibiotics and antibiotic resistance genes in natural environments. Science. 321:365-367.
121. Mather AE, Reid SW, Maskell DJ, Parkhill J, Fookes MC, Harris SR, Brown DJ, Coia JE, Mulvey MR, Gilmour MW, Petrovska L, de Pinna E, Kuroda M, Akiba M, Izumiya H, Connor TR, Suchard MA, Lemey P, Mellor DJ, Haydon DT, Thomson NR. 2013. Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. Science 341:1514-1517.
122. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik J, Ali J, Dante M, Du F, Hou S, Layman D, Leonard S, Nguyen C, Scott K, Holmes A, Grewal N, Mulvaney E, Ryan E, Sun H, Florea L, Miller W, Stoneking T, Nhan M, Waterson R, Wilson RK. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature. 413:852-856.
123. McCollister B, Kotter CV, Frank DN, Taylor W, Jobling MG. 2016. Whole-genome sequencing identifies *in vivo* acquisition of a *blaCTX*$_{\text{-M-27}}$-carrying IncFII transmissible plasmid as the cause of ceftriaxone treatment failure for an invasive *Salmonella enterica* serovar Typhimurium infection. Antimicrob Agents Chemother. 60:7224-7235.
124. McCoy RC, Taylor RW, Blauwkamp TA, Kelley JL, Kertesz M, Pushkarev D, Petrov DA, Fiston-Lavier AS. 2014. Illumina TruSeq synthetic long-reads empower *de novo* assembly and resolve complex,

highly-repetitive transposable elements. PLOS One. 9:e106689. https://doi.org/10.1371/journal.pone.0106689.

125. Mellon M, Benbrook C, Benbrook KL. 2001. Estimates of antimicrobial abuse in livestock. UCS Publications, Cambridge, MA.
126. Metzker ML. 2010. Sequencing technologies- the next generation. Nat Rev Genet. 11:31-46.
127. Mohammed M, Le Hello S, Leekitcharoenphon P, Hendriksen R. 2017. The invasome of *Salmonella* Dublin as revealed by whole genome sequencing. BMC Infect Dis. 17:544. https://doi.org/10.1186/s12879-017-2628-x.
128. Mølbak K, Baggesen DL, Araestrup FM, Ebbesen JM, Engberg J, Frydendahl K, Gerner-Smidt P, Ptersen AM, Wegener HC. 1999. An outbreak of multidrug-resistant, quinolone-resistant *Salmonella enterica* serotype Typhimurium DT104. N Engl J Med. 341:1420-1425.
129. Nakamura Y, Itoh T, Matsuda H, Gojobori T. 2004. Biased biological functions of horizontally transferred genes in prokaryotic genomes. Nature Genet. 36:760-766. (Letter.)
130. Nesme J, Cécillon S, Delmont TO, Monier JM, Voggel TM, Simonet P. 2014. Large-scale metagenomic-based study of antibiotic resistance in the environment. Curr Biol. 24:1096-1100.
131. Noda T, Murakami K, Etoh Y, Okamoto F, Yatsuyanagi J, Sera N, Furuta M, Onozuka D, Oda T, Asai T, Fujimoto S. 2015. Increase in resistance to extended-spectrum cephalosporins in *Salmonella* isolated from retail chicken products in Japan. PLOS One. 10:e0121469 https://doi.org/10.1371/journal.pone.0121469
132. Novick RP, Clowes RC, Cohen SN, Curtiss R, Datta N, Falkow S. 1976. Uniform nomenclature for bacterial plasmids: a proposal. Bacteriol Rev. 40:168-189.
133. Ohl ME, Miller SI. 2001. *Salmonella*: A model for bacterial pathogenesis. Annu Rev of Med. 52:259-274.
134. Oxford Nanopore Technologies™. 2017. Albacore v. 2.0.1. https://nanoporetech.com/.
135. Paesold G, Guiney DG, Eckmann L, Kagnoff MF. 2002. Genes in the *Salmonella* pathogenicity island 2 and the *Salmonella* virulence plasmid are essential for *Salmonella*-induced apoptosis in intestinal epithelial cells. Cell Microbiol. 4:771-781.
136. Pei R, Kim SC, Carlson KH, Pruden A. 2006. Effect of river landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes (ARG). Water Res. 40:2427-2435.

137. Perez JM, Cavalli P, Roure C, Renac R, Gille Y, Freydiere AM. 2003. Comparison of four chromogenic media and hektoen agar for detection and presumptive identification of *Salmonella* strains in human stools. J Clin Microbiol. 41:1130-1134. https://doi.org/10.1128/JCM.41.3.1130-1134.2003.
138. Rabsch W, Andrews HL, Kingsley RA, Prager R, Tschäpe H, Adams LG, Bäumler AJ. 2002. *Salmonella enterica* serotype Typhimurium and its host-adapted variants. Infect and Immun. 70:2249-2255.
139. Raffatellu M, George MD, Akiyama Y, Hornsby MJ, Nuccio SP, Paixao TA, Butler BP, Chu H, Santos RL, Berger T, Mak TW, Tsolis RM, Bevins CL, Solnick JV, Dandekar S, Bäumler AJ. 2009. Lipocalin-2 resistance confers an advantage to *Salmonella enterica* serotype Typhimurium for growth and survival in the inflamed intestine. Cell Host Microbe. 5:476-486.
140. Rajbanshi A. 2008. Study on heavy metal resistant bacteria in Guheswori sewage treatment plant. Our Nature. 6:52-57. https://doi.org/10.3126/on.v6i1.1655
141. Rathman M, Barker LP, Falkow S. 1997. The unique trafficking pattern of *Salmonella* Typhimurium-containing phagosomes in murine macrophages is independent of the mechanism of bacterial entry. Infect Immun. 65:1475-1485.
142. Rhen M, Maskell D, Mastroeni P, Threlfall J (ed). 2007. *Salmonella* Molecular Biology and Pathogenesis. British Library Cataloguing-in-Publication Data, Norfolk, UK.
143. Robertson J, Nash JHE. 2018. MOB-suite: software tools for clustering, reconstructing, and typing of plasmids from draft assemblies. Microb Genom. 4. https://doi.org/10.1099/mgen.0.000206.
144. Rychlik I, Gregorova D, Hradecka H. 2005. Distribution and function of plasmids in *Salmonella enterica*. Vet Microbiol. 112:1-10.
145. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molec Bio and Evol. 4:406-425. https://doi.org/10.1093/oxfordjournals.molbev.a040454.
146. Schmieger H, Schicklmaier P. 1999. Transduction of multiple drug resistance of *Salmonella enterica* serovar Typhimurium DT104. Microbiol Letters. 170:251-256. (Letter.)
147. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 30:2068-2069. https://doi.org/10.1093/bioinformatics/btu153.
148. Seemann T. 2016. ABRicate: mass screening of contigs for antibiotic resistance genes. https://github.com/tseemann/abricate.

149. Shipp CR, Rowe B. 1980. A mechanized microtechnique for *Salmonella* serotyping. J Clin Path. 33:595-597.
150. Sivapalasingam S, Barrett E, Kimura A, Duyne SV, De Witt W, Ying M, Frisch A, Phan Q, Gould E, Shillam P, Reddy V, Cooper T, Hoekstra M, Higgins C, Sanders JP, Tauxe RV, Slutsker L. 2003. A multistate outbreak of *Salmonella enterica* serotype Newport infection linked to mango consumption: impact of water-dip disinfestation technology. Clin Infect Dis. 37:1585-1590.
151. Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. 2010. Mobility of plasmids. Microbio and Molec Biol Rev. 74:434-452.
152. Stone GG, Oberst RD, Hays MP, McVey S, Chengappa MM. 1994. Detection of *Salmonella* serovars from clinical samples by enrichment broth cultivation-PCR procedure. J Clin Microbiol. 32:1742-1749.
153. Su LH, Wu TL, Chiu CH. 2012. Development of carbapenem resistance during therapy for non-typhoidal *Salmonella* infection. Clin Microbiol Infect. 18:E91-94.
154. Takeuchi A. 1967. Electron microscope studies of experimental *Salmonella* infection. I. Penetration into the intestinal epithelium by *Salmonella* Typhimurium. Am J Pathol. 50:109-136.
155. Thomas CM, Nielsen KM. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nature Rev Microbiol. 3:711-721.
156. Tolen TN, Xie Y, Hernandez AC, Gabriel F, Everett K. 2015. Complete genome sequence of *Salmonella enterica* serovar Typhimurium myophage mushroom. Genome Announc. 3:2e00154-15.
157. Toro M, Retamal P, Ayers S, Barreto M, Allard M, Brown EW, Gonzalez-Escalona N. 2016. Whole-genome sequencing analysis of *Salmonella enterica* serovar Enteriditis isolates in Chile provides insights into possible transmission between gulls, poultry, and humans. Appl Environ Microbiol. 82:6223-6232.
158. Tosini F, Visca P, Luzzi I, Dionisi AM, Pezzella C, Petrucca A, Carattoli A. 1998. Class 1 integron-borne multiple-antibiotic resistance carried by IncFl and IncL/M plasmids in *Salmonella enterica* serotype Typhimurium. Antimicrob Agents Chemother. 42:3053-3058
159. University of Kentucky College of Agriculture. Updated November 2017. Poultry production manual. University of Kentucky, Lexington, KY.
160. Unterholzner SJ, Poppenberger B, Rozhon W. 2013. Toxin-antitoxin systems: biology, identification, and application. Mob Genet Elements. 3:e26219. https://doi.org/10.4161/mge.26219.
161. Van TT, Nguyen HN, Smooker PM, Coloe PJ. 2012. The antibiotic resistance characteristics of non-typhoidal *Salmonella enterica* isolated from

food-producing animals, retail meat, and humans in South East Asia. Int J Food Microbiol. 154:98-106.

162. Van Benden CA, Keene WE, Strang RA, Werker DH, King AS, Mahon B, Hedberg K, Bell A, Kelly MT, Balan VK, Mac Kenzie WR, Fleming D. 1999. Multinational outbreak of *Salmonella enterica* serotype Newport infections due to contaminated alfalfa sprouts. J Amer Med Assoc. 281:158-162.
163. Vasquez-Torres A, Xu Y, Jones-Carson J, Holden DW, Lucia SM, Dinauer MC, Mastroeni P, Fang FC. 2000. *Salmonella* pathogenicity island 2-dependent evasion of the phagocyte NADPH oxidase. Science. 287:1655-1658.
164. Wain J, Keddy KH, Hendriksen RS, Rubino S. 2013. Using next generation sequencing to tackle non-typhoidal *Salmonella* infections. J Infect Dev Ctries. 7:1-5.
165. Wehrli W. 1983. Rifampin: mechanisms of action and resistance. Rev of Infect Dis. 5:S407-S411.
166. Wick RR. 2017. Porechop. https://github.com/rrwick/Porechop
167. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom. 3:e000132. https://doi.org/10.1099/mgen.0.000132.
168. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genomes assemblies from short and long sequencing reads. PLOS Comput Biol. 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
169. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of *de novo* genome assemblies. Bioinfor. 31:3350-3352.
170. Winokur PL, Vonstein DL, Hoffamn LJ, Uhlenhopp EK, Doern GV. 2001. Evidence for transfer of CMY-2 *AmpC* β-lactamase plasmids between *Escherichia coli* and *Salmonella* isolates from food and animals and humans. Antimicrob Agents Chemother. 45:2716-2722.
171. Wolfson JS, Hooper DC. 1985. The fluoroquinolones: structures, mechanisms of action, and resistance and spectra activity *in vitro*. Antimicrob Agents Chemother. 28:581-586.
172. Wong KK, McClelland M, Stillwell LC, Sisk EC, Thurston SJ, Saffer JD. 1998. Identification and sequence analysis of a 27-kilobase chromosomal fragment containing a *Salmonella* pathogenicity island located at 92 minutes on the chromosome map of *Salmonella enterica* serovar Typhimurium LT2. Infect and Immun. 66:3365-3371.
173. Wood MW, Jones MA, Watson PR, Siber AM, McCormick BA, Hedges S, Rosqvist R, Wallis TS, Galyov EE. 2000. The secreted effector protein of

Salmonella Dublin, SopA, is translocated into eukaryotic cells and influences the induction of enteritis. Cell Microbiol. 2:293-303.

174. Wuyts V, Denayer S, Roosens NHC, Mattheus W, Bertrand S, Marchal K, Dierick K, De Keersmaecker SCJ. 2015. Whole genome sequence analysis of *Salmonella* Enteriditis PT4 outbreaks from a national reference laboratory's viewpoint. PLOS Curr Outbreaks. 1. https://doi.org/10.1371/currents.outbreaks.aa5372d90826e6cb0136ff66bb7a62fc.
175. Xi C, Zhang Y, Marrs CF, Ye W, Simon C, Foxman B, Nriagu J. 2009. Prevalence of antibiotic resistance in drinking water treatment and distribution systems. Appl Environ Microbiol. 75: 5714-5718.
176. Xu J, Xu Y, Wang H, Guo C, Qiu H, He Y, Zhang Y, Li X, Meng W. 2015. Occurrence of antibiotics and antibiotic resistance genes in sewage treatment plant and its effluent-receiving river. Chemosphere. 119:1379-1385.
177. Yachison CA, Yoshida C, Robertson J, Nash JHE, Kruczkiewicz P, Taboada EN, Walker M, Reimer A, Christianson S, Nichani A, PulseNet Canada Steering Committee, Nadon C. 2017. The validation and implications of using whole genome sequencing as a replacement for traditional serotyping for a national *Salmonella* reference laboratory. Front Microbiol. 8:1044.
178. Yang Y, Li B, Zou S, Fang HH, Zhang T. Fate of antibiotic resistance genes in sewage treatment plant revealed by metagenomic approach. Water Res. 62:97-106. Antimicrob Chemother. 67:2640-2644.
179. Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VPJ, Nash JHE, Taboada EN. 2016. The *Salmonella in silico* typing resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft *Salmonella* genome assemblies. PLOS One 11:e0147101. https://doi.org/10.1371/journal.pone.0147101.
180. Yoshida T, Kim SR, Komano T. 1999. Twelve *pil* genes are required for biogenesis of the R64 this pilus. J Bacteriol. 181:2038-2043.
181. Zhang S, Yin Y, Jones MB, Zhang Z, Kaiser BLD, Dinsmore BA, Fitzgerald C, Fields PI, Deng X. 2015. *Salmonella* serotype determination utilizing high-throughput genome sequence data. J Clin Microbiol. 53:1685-1692.
182. Zhang S, Li S, Gu W, Bakker HD, Boxrud D, Taylor A, Roe C, Driebe E, Engelthaler DM, Allard M, Brown E, McDermott P, Zhao S, Bruce BB, Trees E, Fields PI, Deng X. 2019. Zoonotic source attribution of *Salmonella enterica* serotype Typhimurium using genomic surveillance data, United States. Emerg Infect Dis. 25:82-91. https://doi.org/10.3201/eid2501.180835.

183. Zhao S, Qaiyumi S, Friedman S, Singh R, Foley SL, White DG, McDermott PF, Donkar T, Bolin C, Munro S, Baron EJ, Walker RD. 2003. Characterization of *Salmonella enterica* serotype Newport isolated from humans and food animals. J Clin Micrbiol. 41:5366-5371.

184. Zhao S, White DG, Friedman SL, Glenn A, Blickenstaff K, Ayers SL, Abbott JW, Hall-Robinson E, McDermott PF. 2008. Antimicrobial resistance in *Salmonella enterica* serovar Heidelberg isolates from retail meats, including poultry, from 2002 to 2006. Appl Environ Microbiol. 74:6656-6662.

185. Zhou Z, Alikhan NF, Sergeant MJ, Luhmann N, Vaz C, Francisco AP, Carriço JA, Achtman M. 2018. GrapeTree: visualizations of core genomic relationships among 100,000 bacterial pathogens. Genome Res. 28: 1395-1404. https://doi:10.1101/gr.232397.117.

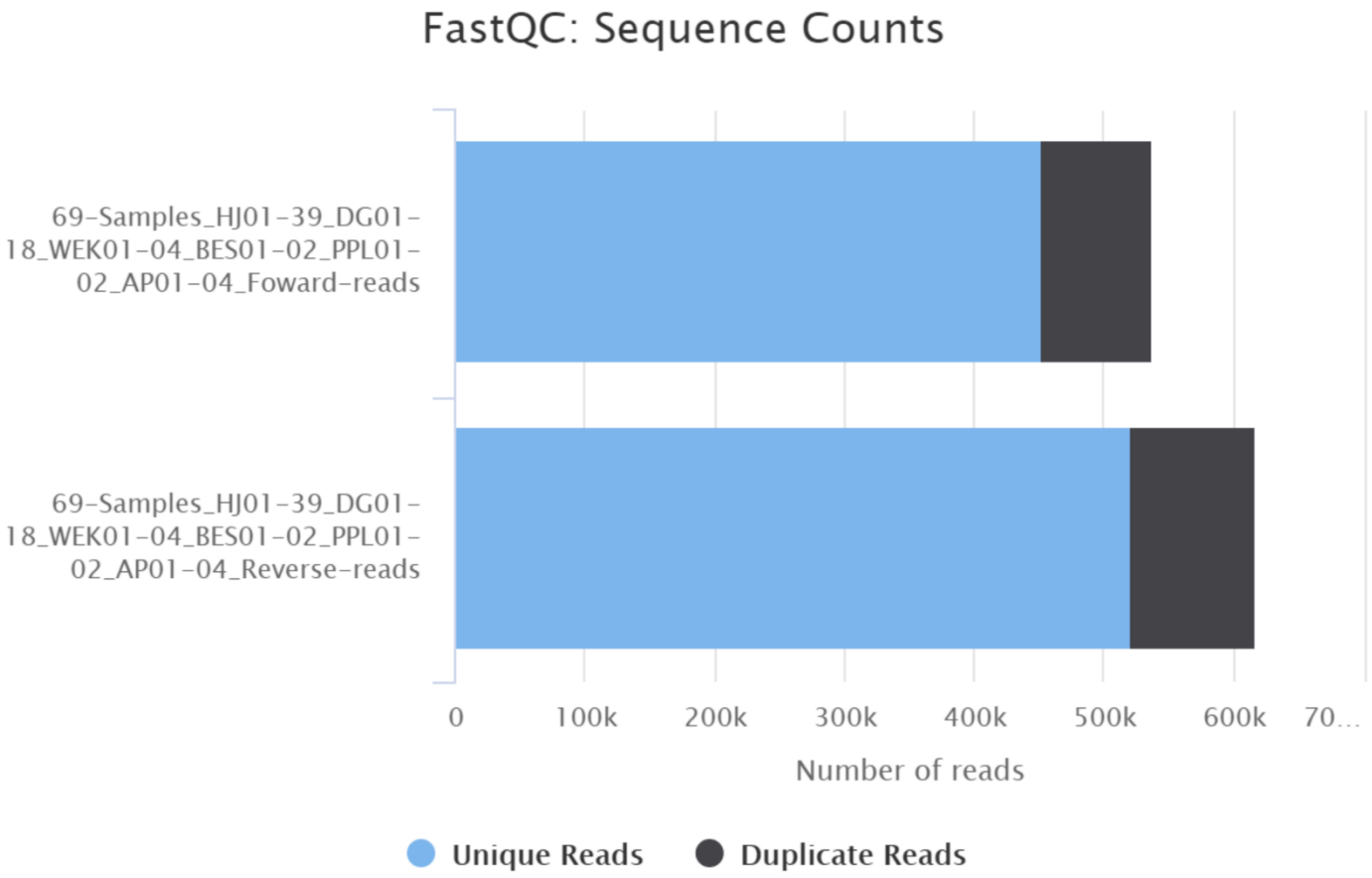
186. Zhu YG, Johnson TA, Su JQ, Qiao M, Guo GX, Stedtfeld RD, Hashham SA, Tiedje JM. 2012. Diverse and abundant antibiotic resistance genes in Chinese swine farms. PNAS. 110:3435-3440.

Appendix

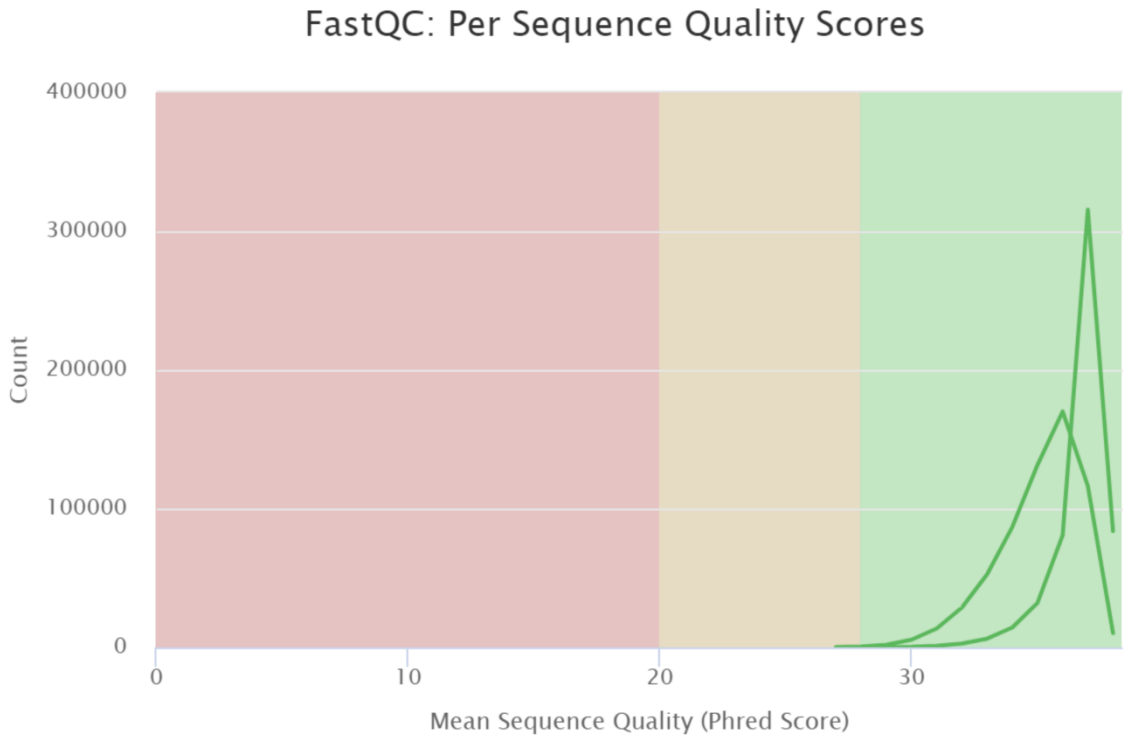
A

Sample Name	% Dups	% GC	Length	M Seqs
69-Samples_HJ01-39_DG01-18_WEK01-04_BES01-02_PPL01-02_AP01-04_Foward-reads	15.8%	52%	236 bp	0.5
69-Samples_HJ01-39_DG01-18_WEK01-04_BES01-02_PPL01-02_AP01-04_Reverse-reads	15.6%	52%	201 bp	0.6

B



C



D

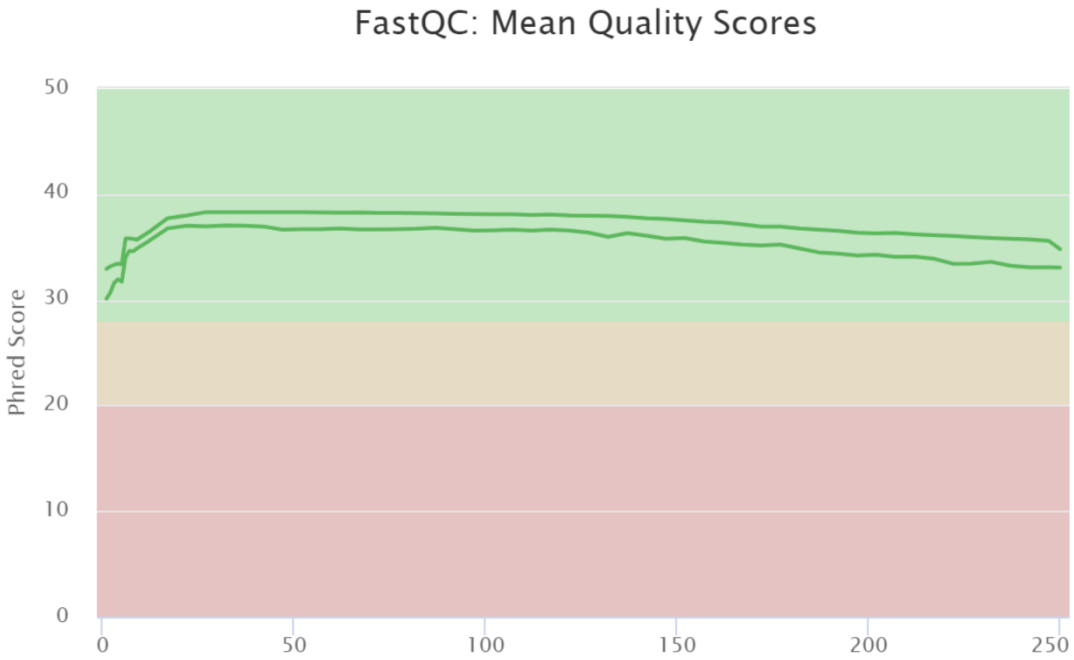
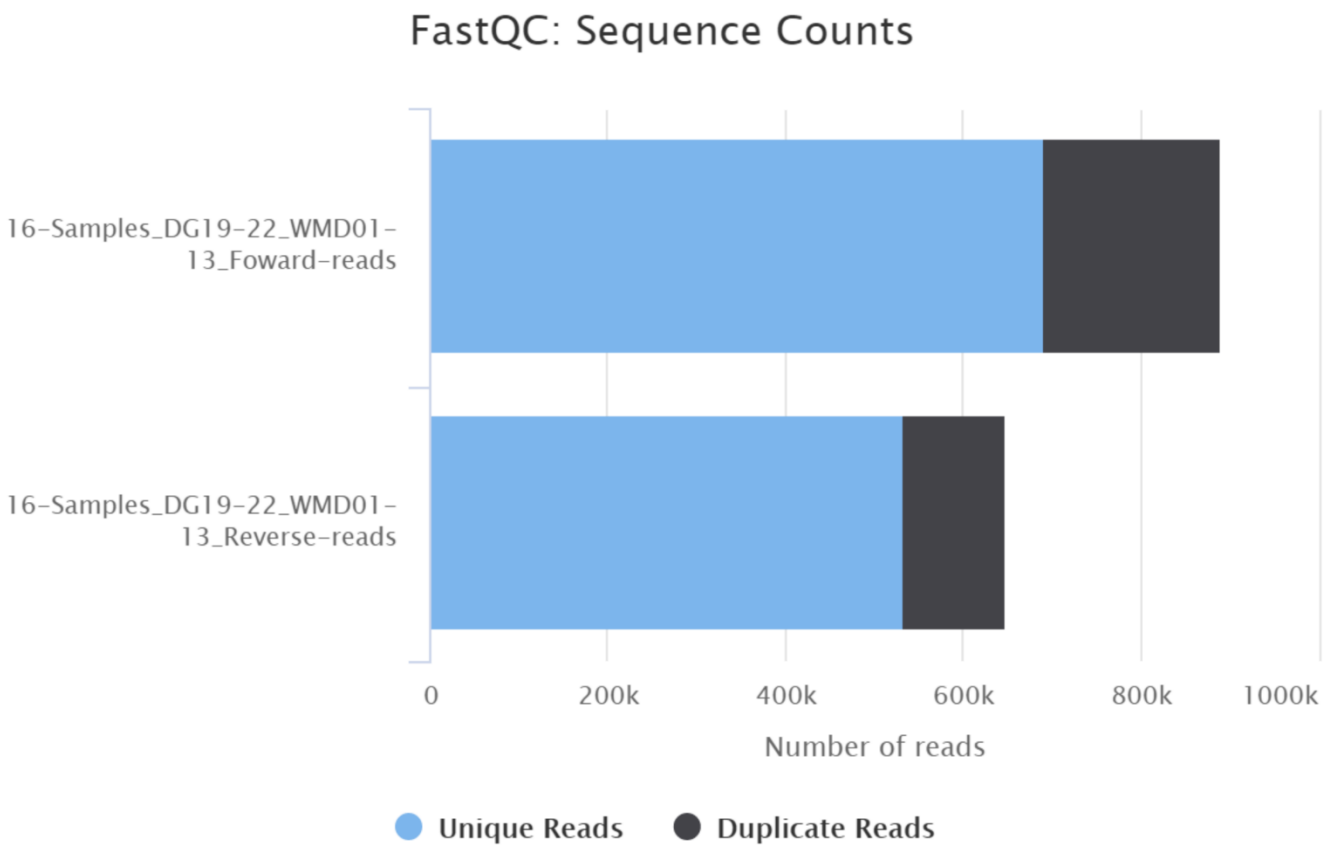


Figure S1: MultiQC outputs depicting overall base and sequence quality for short-read sequences from isolates HJ01-39, DG01-18, FHS01-04, WEK01-04, BES01-02, and AP01-04. (A) Depiction of overall sequence metrics for FastQC reports. % Dups- Percentage of reads that were duplications, %GC- average percent GC for sequences, Length- Average length of sequences, M Seqs- number of sequences in millions. (B) Plot showing the total number of sequences. (C) FastQC plot showing average sequence quality. (D) FastQC plot showing average per base quality score

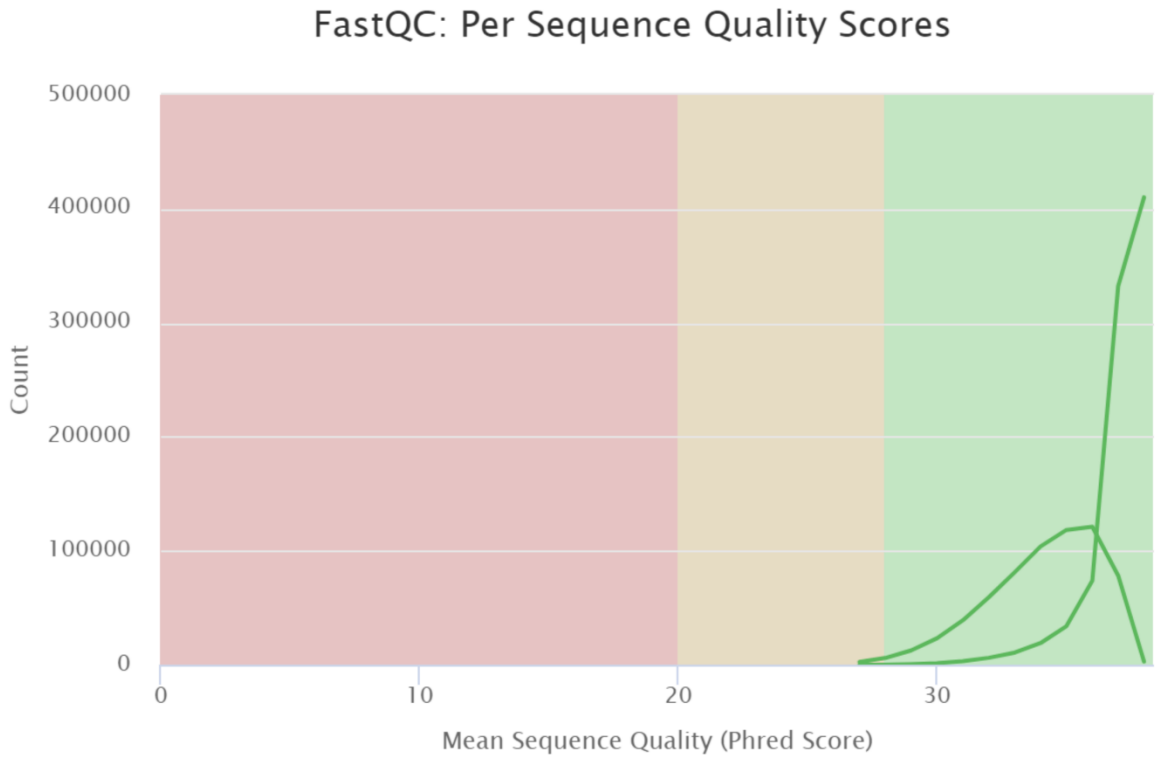
A

Sample Name	% Dups	% GC	Length	M Seqs
16-Samples_DG19-22_WMD01-13_Foward-reads	22.4%	52%	147 bp	0.9
16-Samples_DG19-22_WMD01-13_Reverse-reads	17.7%	52%	128 bp	0.6

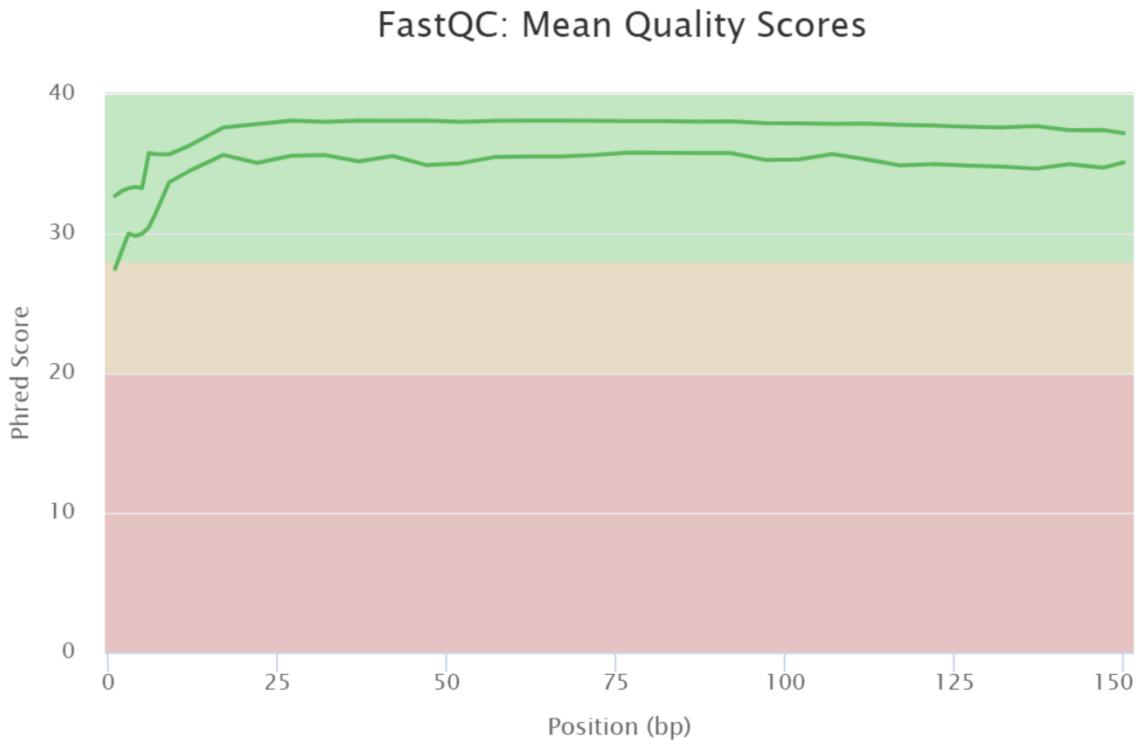
B



C



D

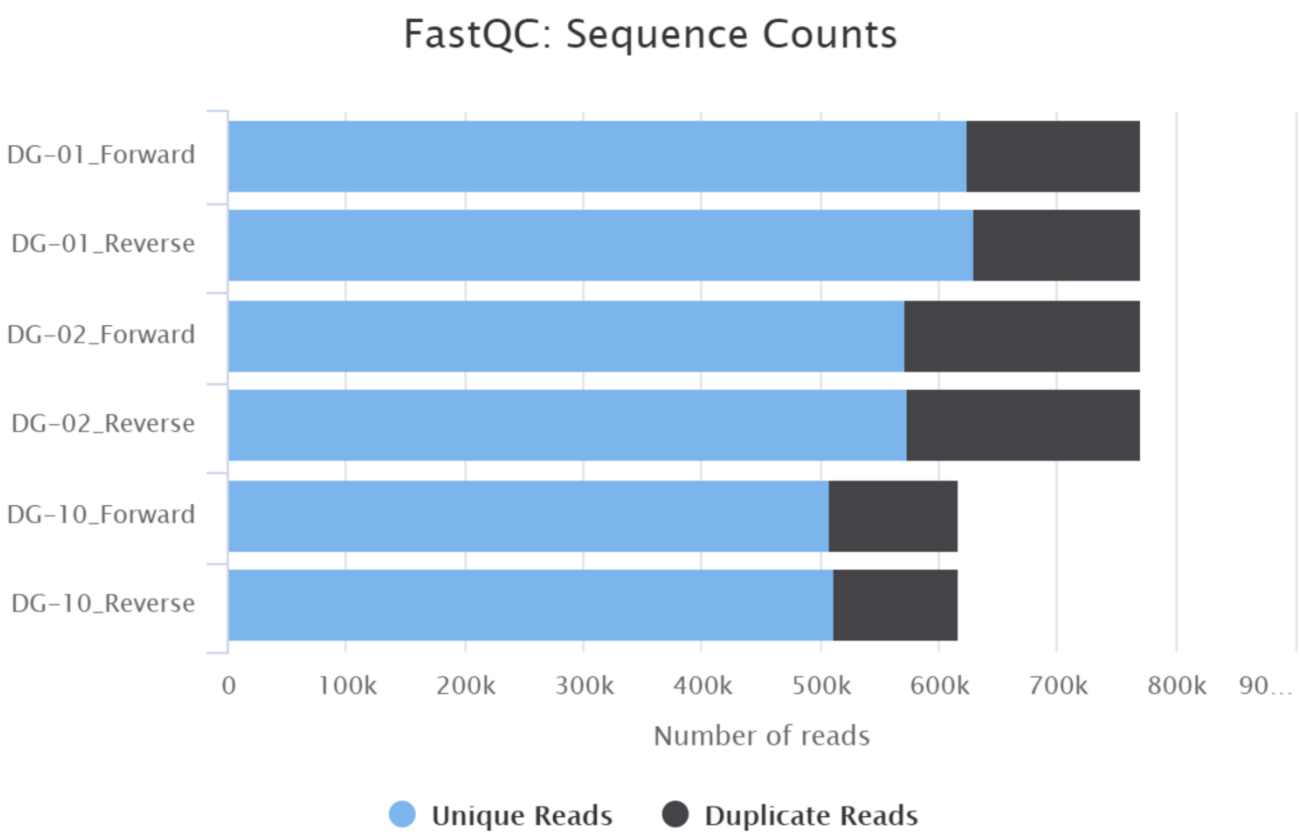


S2: MultiQC outputs depicting overall base and sequence quality for short-read sequences from isolates DG19-22 and WMD01-13. (A) Depiction of overall sequence metrics for FastQC reports. % Dups- Percentage of reads that were duplications, %GC- average percent GC for sequences, Length- Average length of sequences, M Seqs- number of sequences in millions. (B) Plot showing the total number of sequences. (C) FastQC plot showing average sequence quality. (D) FastQC plot showing average per base quality score

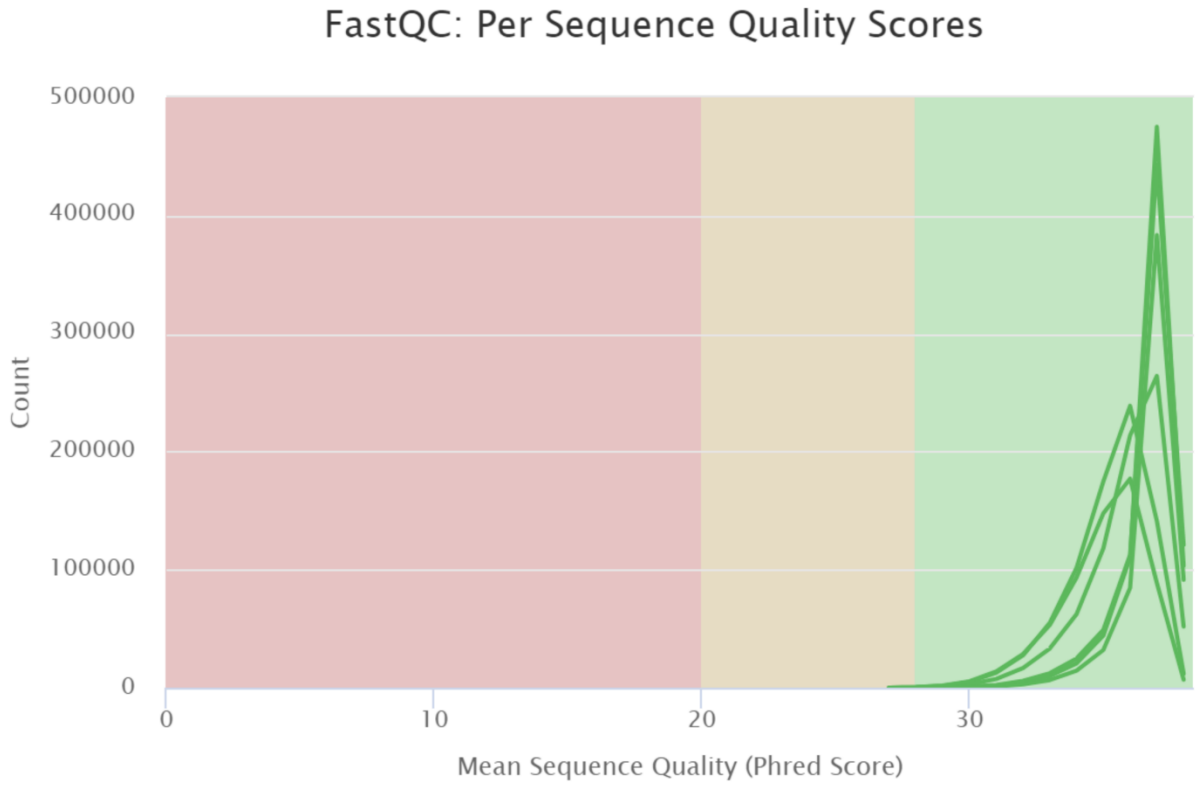
A

Sample Name	% Dups	% GC	Length	M Seqs
DG-01_Forward	18.8%	52%	234 bp	0.8
DG-01_Reverse	18.2%	52%	201 bp	0.8
DG-02_Forward	25.8%	52%	229 bp	0.8
DG-02_Reverse	25.6%	52%	210 bp	0.8
DG-10_Forward	17.6%	52%	237 bp	0.6
DG-10_Reverse	17.0%	52%	196 bp	0.6

B



C



D

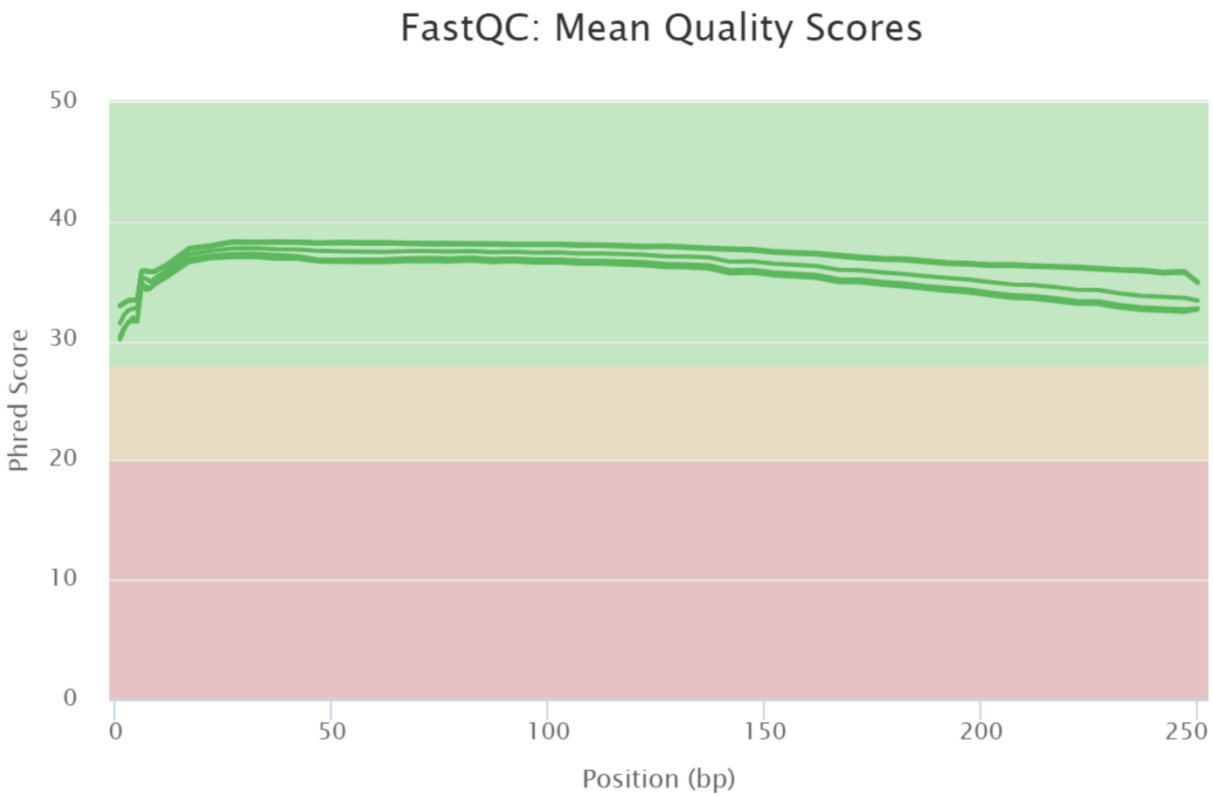


Figure S3: MultiQC outputs depicting overall base and sequence quality for short-read sequences from isolates DG-01, DG-02, and DG-10. (A) Depiction of overall sequence metrics for FastQC reports. % Dups- Percentage of reads that were duplications, %GC- average percent GC for sequences, Length- Average length of sequences, M Seqs- number of sequences in millions. (B) Plot showing the total number of sequences. (C) FastQC plot showing average sequence quality. (D) FastQC plot showing average per base quality score