


Spring 2016

# Next-generation sequencing of a multi-drug resistance plasmid captured from stream sediment

Kevin G. Libuit  
*James Madison University*

Follow this and additional works at: <https://commons.lib.jmu.edu/master201019>

 Part of the [Bioinformatics Commons](#), [Ecology and Evolutionary Biology Commons](#), [Environmental Microbiology and Microbial Ecology Commons](#), [Other Genetics and Genomics Commons](#), and the [Other Immunology and Infectious Disease Commons](#)

---

## Recommended Citation

Libuit, Kevin G., "Next-generation sequencing of a multi-drug resistance plasmid captured from stream sediment" (2016). *Masters Theses*. 114.  
<https://commons.lib.jmu.edu/master201019/114>

This Thesis is brought to you for free and open access by the The Graduate School at JMU Scholarly Commons. It has been accepted for inclusion in Masters Theses by an authorized administrator of JMU Scholarly Commons. For more information, please contact [dc\\_admin@jmu.edu](mailto:dc_admin@jmu.edu).

Next-Generation Sequencing of a Multi-Drug Resistance Plasmid Captured from Stream  
Sediment

Kevin G. Libuit

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 2016

---

FACULTY COMMITTEE:

Committee Chair : Dr. James B. Herrick

Committee Members/ Readers:

Dr. Steve Cresawn

Dr. Kyle Seifert

## **Acknowledgements**

Dr. James Herrick has been my academic advisor and mentor for the past four years. Working under his guidance has been the most rewarding experience of my life. It has been a privilege to learn from such an incredible researcher and educator. I owe the completion of this thesis to his insight, unending verve for discovery and trust in me.

I also thank my committee members, Dr. Kyle Seifert and Dr. Steve Cresawn, who have offered invaluable advice throughout the completion of this project and my collaborators, Dr. Stephen Turner and Adrian Pelin, who have made major contributions to the genome assembly and annotation pipeline. I am grateful to Erika Gehr for her work in the isolation and characterization of pEG1-1 as well as her mentorship in the lab, and to my lab partner, Curtis Kapsak, for his friendship and help with MinION sequencing.

## Table of Contents

|   |     |
|---|-----|
| Acknowledgements.....   | ii  |
| Table of Contents.....  | iii |
| List of Tables.....   | iv  |
| List of Figures.....  | v   |
| Abstract.....   | vi  |
| Introduction.....   | 1   |
| Antibiotics and the Threat of Resistance.....   | 1   |
| Horizontal Gene Transfer and Antibiotic Resistance.....                                       | 4   |
| Next Generation DNA Sequencing.....   | 7   |
| DNA Sequencing of Multi-Drug Resistance Plasmids.....   | 9   |
| Methods.....  | 11  |
| Plasmid pEG1-1.....   | 11  |
| Plasmid isolation and purification.....   | 11  |
| Electroporation.....  | 13  |
| Ion Torrent Personal Genome Machine Sequencing.....   | 13  |
| Oxford Nanopore Technologies' MinION Sequencing.....  | 14  |
| Genome Assembly and Annotation.....   | 15  |
| Results.....  | 17  |
| Developing Methods to Prepare Plasmid DNA<br>for Sequencing on Next-Generation Platforms..... | 17  |
| Sequencing on the Ion Torrent PGM.....  | 20  |
| Sequencing on Oxford Nanopore Technologies' MinION.....                                       | 23  |
| Hybrid Assembly.....  | 25  |
| Annotation.....   | 28  |
| pEG1-1 Resistance Phenotype.....  | 31  |
| Functional Modules of pEG1-1.....   | 32  |
| Discussion.....   | 36  |
| Sequencing, Assembly and Annotation.....  | 36  |
| pEG1-1.....   | 38  |
| Conclusion.....   | 41  |
| Literature Cited.....   | 43  |

## List of Tables

|  |    |
|--|----|
| 1. Ion Torrent PGM run analysis options..... | 14 |
| 2. Complete list of pEG1-1 genes.....        | 29 |

## List of Figures

|  |    |
|--|----|
| 1. Original and scaled-up plasmid preparation protocols. ....  | 18 |
| 2. Plasmid DNA preparation for MinION sequencing.....  | 20 |
| 3. Ion Torrent PGM sequence diagnostics.....   | 22 |
| 4. Oxford Nanopore Technologies' MinION sequence diagnostics. ....   | 24 |
| 5. Sequencing, assembly and annotation pipeline for plasmid pEG1-1.....  | 26 |
| 6. Global alignment of the pEG1-1 assembly and pB8.....  | 28 |
| 7. The seven functional modules mapped along the multi-drug resistance IncP-1 $\beta$<br>plasmid pEG1-1 genome. .... | 33 |
| 8. Genetic map of the Gene Load 1 & 2 modules. ....  | 34 |

## Abstract

Plasmids in agriculturally-impacted bodies of water may play a significant role in the dissemination of antibiotic resistance. Previously, Erika Gehr, as part of her M.S. thesis work in our laboratory, captured environmental plasmids without cultivation of host bacteria from stream sediment into *Escherichia coli*. Individual plasmids were capable of conferring resistance to a surprising array of antibiotics including aminoglycosides and extended-spectrum  $\beta$ -lactams. In this study, we developed a method to sequence multi-drug resistance plasmids using both Oxford Nanopore MinION and Ion Torrent Personal Genome Machine sequencers. Plasmid pEG1-1 was sequenced on both platforms and a hybrid assembly utilizing data from both sequencing platforms generated a single 73,320 bp contig that was annotated using automated and manual techniques. Analysis of the genome revealed pEG1-1 to be an IncP-1 $\beta$  plasmid with two mobile genetic elements – a *tn21*-related transposon and an *in104* complex integron – both of which carry multiple antibiotic resistance genes. These findings suggest that plasmids in stream sediment are prone to the incorporation of mobile genetic elements that introduce a broad range of antibiotic resistance genes into their genome. This could cause serious risk to human health since IncP-1 $\beta$  plasmids are capable of transferring into nearly all Gram-negative bacteria, including fecal pathogens that get introduced to stream sediment.

## Introduction

**Antibiotics and the Threat of Resistance.** Antibiotics have saved millions of lives around the world and have been recognized as one of the most successful forms of chemotherapy in the history of medicine (Aminov, 2010). Antibiotics have cured and prevented otherwise fatal diseases, enabled major advances in surgery, and even helped to extend the life expectancy of humans (Gould & Bal, 2013; Ventola, 2015). Morbidity and mortality caused by infectious diseases have drastically decreased in developing countries due to antibiotic use (World Health Organization (WHO), 2014). In the 1960's, the use of antibiotics had such a drastic effect on public health that the Surgeon General of the United States of America, William Stewart, proclaimed "The time has come to close the book on infectious diseases. We have basically wiped out infection in the United States" (Upshur, 2008). Unfortunately, nearly fifty years later, the book on infectious diseases is wide open and a new chapter is unfolding.

Antibiotic resistance (AR) has been deemed one of our most serious health threats (Centers for Disease Control [CDC], 2013). AR infections have been reported around the globe and the number of instances continues to rise (Ventola, 2015). In the United States alone, serious infections attributed to a pathogen resistant to at least one antibiotic have reached rates of more than 2 million patients a year, with at least 23,000 dying as a result (CDC, 2013). For example, methicillin-resistant *Staphylococcus aureus* is now one of America's most lethal diseases, surpassing fatality rates of HIV/AIDS, Parkinson's disease, emphysema, and homicide combined (Gross, 2013). Other organisms of particular concern due to their recently acquired antibiotic resistance or increasing pathogenicity include *Mycobacterium tuberculosis*; *Streptococcus pneumoniae*;



*Acinetobacter* spp., *Stenotrophomonas maltophilia*, and *Pseudeomonas aeruginosa* (all intrinsically resistant to many antibiotics); *Salmonella enterica* and pathogenic *E. coli*; and Extended-Spectrum Beta-Lactamase- (ESBL) and Carbapenemase-Resistant Enterobacteriaceae (CRE).

Efforts to combat the growing problem of antibiotic resistance are being carried out by national and international bodies around the world (WHO, 2014). In March of 2015, the White House released the “National Action Plan for Combating Antibiotic-Resistant Bacteria” (The White House Office of the Press Secretary, 2015). In this document, the U.S. government declared that research and surveillance of antibiotic resistant organisms, including CRE and MRSA. They assert that, with a better understanding of how and why resistance spreads, policies can be made to curb its acceleration (The White House Office of the Press Secretary, 2015).

The predominant factor contributing to the increasing prevalence of antibiotic resistance is thought to be the overuse of antibiotics, especially in agriculture (Zur Wiesch *et al.*, 2011). In 2013, the Food and Drug Administration reported that 74% of the medically important antimicrobials sold were administered in animal feed—the majority of which is used for non-therapeutic growth promotion (United States Food and Drug Administration, 2012). The use of antibiotics as an additive to food has been shown to increase feeding efficiency, thereby promoting growth rate by up to 5% in dairy cattle, sheep and goats, relative to animals not treated with antibiotics (Khachtourians, 1998). However, these gains come with profound consequences. Agricultural use of antibiotics has been strongly linked to resistance (Smith *et al.*, 2002). Antibiotic-resistant bacteria

have been isolated on animals treated with antibiotics (Aarestrup, 1999), food products (Chadwick *et al.*, 1996), and even environments treated with animal waste (Chee-Sanford *et al.*, 2001; Brooks, 2005).

Work from our lab showed that the use of prophylactic antibiotics in turkeys increased the presence of antibiotic-resistant bacteria in litter (Brooks, 2005). Brooks quantified culturable bacteria resistant to tetracycline from the litter of tetracycline-treated turkeys and turkeys not treated with tetracycline. His results showed that tetracycline-resistant bacteria were present in significantly greater numbers in litter from turkeys that received tetracycline treatment (Brooks, 2005).

These data are especially concerning since contamination of agricultural soils can reach water systems through runoff. In Virginia, 68% of the state's rivers that are designated as Impaired by the Virginia Department of Environmental Quality (DEQ) are degraded by fecal bacteria. In the Potomac-Shenandoah River basin, that number jumps to 80% (DEQ, 2014).

*Escherichia coli* is used as an indicator of fecal contamination because it is not considered to persist in freshwater for more than a few days (Gordon *et al.*, 2002). While this holds true for *E. coli* in the water column, this may not be the case in stream sediment. A 1979 analysis of *E. coli* populations in sediment compared to the water column in streams in southwestern Idaho indicated that the *E. coli* concentration was 2 to 760 times greater in the sediment when compared to that in the water (Stephenson & Rychert, 1982). More recently, a mesocosm experiment was set up by inoculating water and sediment with fecal contaminants. By enumerating *E. coli* in the sediment and

freshwater over time, the authors showed that the decay rate of *E. coli* in sediment is about 70% lower than *E. coli* in water (Anderson *et al.*, 2005). Work in our lab compared the *E. coli* colony forming units obtained from stream water versus stream sediment (Gehr, 2013). Over a three month span, Gehr demonstrated that stream sediment had a consistently higher number of *E. coli* colony forming units than stream water and that specific strains appeared to persist over time in the sediment only (Gehr, 2013).

Persistence in stream sediment may allow fecal bacteria such as *E. coli* to interact with the large and diverse native microbial population (Nealson, 1997). Through horizontal gene transfer these fecal pathogens could introduce and/or gain access to the pool of genes present in bacterial populations native to stream sediment.

**Horizontal Gene Transfer and Antibiotic Resistance.** Horizontal gene transfer is the exchange of DNA amongst mature bacteria through viral infection (transduction), direct uptake of free DNA (transformation) or the transfer of plasmids through direct cell-to-cell contact (transconjugation). Whole-genome investigations of bacterial species have shown that significant portions of nearly all bacterial genomes contain foreign DNA that has been introduced via horizontal gene transfer (Ochman *et al.*, 2000; Pallen & Wren, 2007). For example, the sequencing of multiple, diverse *E. coli* genomes has revealed a conserved, “core genome” interspersed with multiple regions of high variability. These variable regions are strain-specific and often include mobile elements that code for increased virulence and pathogenicity (Dobrindt *et al.*, 2010). In *Salmonella*, for example, *ssrI* and *sodCI* genes have been found on prophages—genomes of bacteriophages that have been inserted into bacterial genomes (Figueroa-Bossi, 2001).

The horizontal exchange of genetic material offers a mechanism for some bacteria to rapidly adapt to changing environments (Claverys, 2000). The mass distribution of antibiotics in both medicine and agriculture has added a unique selective pressure to the environment. In response, horizontal gene transfer has aided in the emergence of resistance (Davies & Davies, 2010).

The preeminent means of antibiotic resistance transmission is through plasmid conjugation (Davies & Davies, 2010). Plasmids are mobile molecules of circularized DNA that can carry genes involved in virulence, pathogenicity and antibiotic resistance (Kelly *et al.*, 2009, Revilla *et al.*, 2008). Plasmid-mediated resistance genes have been associated with most clinical antibiotics, including those of “last resort”, drugs reserved for infections exhibiting resistance to all other antibiotics. (Bennet, 2008).

A major contributing factor to the dissemination of plasmid-mediated resistance is the capability of certain plasmids to conjugate into a broad-range of bacterial hosts. For example, plasmids from the incompatibility groups IncP can transfer and maintain themselves in nearly all Gram-negative bacteria (Dröge *et al.*, 2000). Incompatibility (Inc) groups are based on replication and partitioning systems; plasmids belonging to the same Inc group cannot coexist in a single cell because of competing replication and partitioning genes (Shintani *et al.*, 2010a). These genes are also contributing factors to a plasmid’s capacity to successfully conjugate into certain hosts.

The replication and partitioning system shared by all IncP-1 plasmids allow for successful conjugation in a broad range of hosts (Shintani *et al.*, 2010b). IncP-1 plasmids have been found in virtually all Gram-negative bacteria, including *E. coli* and

*Pseudomonas* spp., and in pathogens and commensals from a wide variety of sources, such as hospital and other clinical and veterinary settings, agriculturally-impacted soils and wastewater treatment plants (Bahl, 2009; Norberg *et al.*, 2011). Along with their broad host range, IncP-1 plasmids can carry genes encoding antibiotic resistance (Popowska & Krańczyk-Balska, 2013). These two factors make IncP-1 plasmids of particular interest to molecular and environmental biologists working to understand the underlying mechanisms driving the evolution and spread of antibiotic resistance.

IncP-1 $\beta$  plasmids, a subgroup within the incompatibility group IncP-1, have been found to confer multi-drug resistance in a wide range of bacterial hosts (Lee *et al.*, 2003). The Tra1 and Tra2 regions of the IncP-1 $\beta$  backbone—gene sequences that mediate conjugative transfer—are separated by regions of clustered restriction sites, hotspots for the integration of mobile genetic elements (Popowska & Krańczyk-Balska, 2013). Transposons carrying antibiotic resistance genes, such as the aminoglycoside resistance Tn5393c transposon found on plasmid pB4 (Tauch A. *et al.*, 2003), are commonly found between the Tra1 and Tra 2 regions of IncP-1 $\beta$  plasmids.

Because it is typical to find these plasmids in agricultural sites, clinical settings and wastewater treatment plants, IncP-1 plasmids have an opportunity to transmit antibiotic resistance genes to human pathogens through HGT. An *in vitro* conjugation assay of pB10, an IncP-1 $\beta$  plasmid carrying two multi-drug resistant transposons, demonstrated that the plasmid could be transferred to the food-borne pathogens *E. coli* and *Salmonella* spp. (Van Meervenne *et al.*, 2012). These data demonstrate the urgency to understand more about IncP-1 plasmids and their role in antibiotic resistance. The analysis of IncP-1 plasmid genomes has helped to uncover the importance of mobile

genetic elements, but the database is limited (Popowska & Kraqczyk-Balska, 2013). Currently, only five complete sequences of IncP-1 $\beta$  plasmids have been published (Popowska & Kraqczyk-Balska, 2013).

**Next Generation DNA Sequencing.** Modern DNA sequencing began in 1975 when the first complete sequence, that of bacteriophage phi X174, was published (Sanger & Coulson, 1975). Sanger & Coulson sequenced the phage DNA using their ‘plus and minus’ method of DNA sequencing. Two years later, Sanger et al. refined the ‘plus and minus’ method, creating a more efficient means of sequencing, termed the chain termination method (Sanger *et al.*, 1977). Sanger sequencing has since been utilized to sequence a broad range of genomes, but the chemistry of Sanger sequencing made increasing throughput (i.e. the amount of DNA being sequenced per run) expensive. In 2004, these constraints were lifted with the introduction of massively parallel sequencing technology, or next-generation sequencing (NGS) (Morey *et al.*, 2013).

What distinguishes NGS from Sanger sequencing is the capability of NGS platforms to produce read information on multiple samples of DNA in a single reaction (Morey *et al.*, 2013). This enabled massive amounts of data to be produced in a single sequencing run, bringing costs to unprecedented lows. In 2004, it cost nearly \$1,000 to sequence 10<sup>6</sup> bp. In 2015, thanks to improving technologies of NGS platforms, sequencing 10<sup>6</sup> bp cost less than a dime (National Institutes of Health, 2016).

The massive data output of these second-generation sequencers is due to chemistries that produce sequence data through cyclic parallel readings of clonally amplified, spatially separated amplicons (Mardis, 2008). While the details differ between

platforms, second-generation sequencers operate under the same, basic workflow principles (Morey et al., 2013). First, sample DNA is clonally amplified and denatured into single-stranded fragments. Examples of clonal amplification techniques are emulsion PCR and bridged amplification. Clonally amplified products are then spatially separated, usually across microchips or flowcells. Next, polymerization of the amplified product's complementary strand is initiated, releasing specific byproducts, e.g. hydrogen ion or pyrophosphate, upon the incorporation of a complimentary nucleotide. Since the sample DNA has been clonally amplified, polymerization byproducts are released in sufficient quantities to be detected. On NGS platforms, multiple samples of DNA can be sequenced in a single reaction since each reaction is occurring in a distinct spatial location, (Morey et al., 2013). With this innovative sequence chemistry, data could be produced more quickly and at a significantly lower cost than Sanger sequencing (NIH, 2016).

However, this technology does have its limitations. Clonal amplification of sample DNA can introduce amplification biases that can drastically decrease the quality of sequence data (Acinas et al., 2005). Also, the chemistries of NGS limit the read lengths produced to <500 bp (Morey et al., 2013). Short read lengths can make assemblies of large genomes difficult and sometimes impossible (Whiteford *et al.*, 2005). New sequencing technology has been introduced that no longer relies on clonal amplification and can produce reads upwards of tens of thousands of base pairs in length. This new wave of technology has established a third-generation of NGS platforms (Morey et al., 2013).

The two third-generation platforms commercially available are the Pacific Biosciences' Single-Molecule Real Time (SMRT) Sequencer and Oxford Nanopore

Technologies' MinION. While these platforms employ drastically different chemistries, both sequence DNA through single-molecule processing (Heather & Chain, 2016). Rather than clonally amplifying sample DNA, third-generation platforms generate sequencing reads from the original, single molecule of DNA. Third-generation sequencers also apply partial separation of samples which, like second-generation platforms, allows for massively parallel reactions to occur. Unlike second-generation platforms, the problem of amplification bias is not an issue for third generation sequencers since there is typically no amplification step. Also common amongst third-generation sequencers is the production of long sequencing reads. Long reads are particularly useful in assembling regions of a genome that contain large repeats (Heather & Chain, 2016).

The major drawback of third-generation platforms is their accuracy. Second-generation machines have average error rates lower than 1%, but an evaluation of the SMRT Sequencer system showed error rates as high as 15% (Carnerio *et al.*, 2013) and an assessment of the MinION reported an error rate of around 8% (Jain *et al.*, 2015). Rather than rely on the low accuracy data generated on third-generation platforms, or the short-read data generated on second-generation platforms, some researchers perform hybrid assemblies that utilize data obtained from both second- and third-generation sequencers (Koren *et al.*, 2012).

**DNA Sequencing of Multi-Drug Resistance Plasmids.** Previous work in our lab revealed the presence of multi-drug resistance plasmids in sediment of agriculturally-impacted streams (Gehr, 2013; Herrick *et al.*, 2014). Gehr used an exogenous capture method to capture (or conjugate) plasmids from samples of stream sediment of Shull



Run, a tributary of Mountain Run and Smith Creek in Rockingham County, Virginia (Gehr, 2013) directly into *E. coli* cells without culturing the plasmid donor cells. She then used a modified Stokes susceptibility assay to determine the antibiotic resistance profile of the captured tet<sup>R</sup> plasmids (Gehr, 2013). Gehr demonstrated that some of these tet<sup>R</sup> plasmids conferred decreased susceptibility to a surprising array of clinical antibiotics, including cefepime, a fourth generation cephalosporin, suggesting that a significant reservoir of antibiotic resistance genes may be present in stream sediment impacted by agricultural runoff (Gehr, 2013).

In this study, we sequenced a tet<sup>R</sup> plasmid captured by Erika Gehr, pEG1-1, and demonstrated how genome analysis of multi-drug resistance plasmids allows a better understanding of the reservoir of ARG present in stream sediment. We have developed methods to isolate plasmid DNA, sequence it on two NGS platforms—the Ion Torrent PGM and Oxford Nanopore Technologies' MinION— combine the data for hybrid assembly, and annotate the genome. Using these methods, we were able to classify pEG1-1 as an IncP-1 $\beta$  plasmid and uncovered the presence of two complex mobile genetic elements: a unique tn21-related transposon and an in104 complex integron, each of which carries multiple antibiotic resistance genes.

## Methods

**Plasmid pEG1-1.** Plasmid pEG1-1 is a stream sediment plasmid captured and characterized by Erika Gehr as part of her M.S. thesis research (originally designated “p1-1” in her study) (Gehr, 2013). It conferred decreased susceptibility to tetracycline, tobramycin, kanamycin, ticarcilin, piperacillin, piperacillin-tazobactam, and cefipime (Gehr, 2013). It was exogenously captured from a stream sediment sample from Shull Run, a tributary of Mountain Run and Smith Creek in Rockingham County, Virginia. It was kept frozen at at -20°C in 100 µl sterile ddH<sub>2</sub>O. The antibiotic resistance profile of pEG1-1 was determined by Gehr (and verified in this study) using a modified Stokes disc diffusion antibiotic susceptibility assay (Gehr, 2013; Herrick, *et al.*, 2014). Minimum inhibitory concentrations (MICs) of pEG1-1 were determined using Sensititre Gram Negative Xtra Plate Format (Trek Diagnostics Inc., Cleveland OH) according to the manufacturer's instructions and NCCLS standards.

**Plasmid isolation and purification.** A plasmid preparation procedure previously developed in our laboratory for the isolation of large, native, single-copy plasmids (Gehr, 2013) was scaled up in order to isolate and purify large quantities of pEG1-1. Transconjugant or transformant cells were grown in 30 ml of trypticase soy broth with shaking at 37° C overnight. Cells were harvested by centrifugation at 10,000 g for 5 min and resuspended in 2 ml of resuspension buffer (10 mM EDTA; 50 mM dextrose; 10 mM Tris-Cl, pH 8.0). Four milliliters of 0.2 M NaOH/1% SDS were added and the mixture kept at room temperature for 5 min to lyse the cells. Three milliliters of ammonium acetate and 3 ml of chloroform were added and the lysate immediately centrifuged at 16,000 x g for 10 min. The supernatant containing plasmid DNA was added to 4 ml of

30% polyethylene glycol 8000/1.5 M NaCl and chilled on ice for 15 min. Plasmid DNA was pelleted by centrifugation at 16,000 g and resuspended in 1 ml of sterile ddH<sub>2</sub>O. This solution was kept at 4° C for at least 24 hours to allow plasmid DNA to fully resuspend and then the DNA was stored at -20° C in sterile ddH<sub>2</sub>O.

The presence of plasmid DNA was confirmed by agarose gel electrophoresis. A 0.7% agarose gel was cast by boiling 0.35 g of agarose in 50 mL of bionic buffer (Sigma-Aldrich, St. Louis, MO) and adding 2.5 µL of 10,000X GelRed Nucleic Acid Gel Stain (Biotium, Hayward, CA). Ten microliters of plasmid DNA mixed with 2 µl of 6X loading dye was run on the gel at 3.5V/cm for 90 min. Gel photographs were taken with a Kodak DC 290 digital camera (Kodak, New Haven, CT) and analyzed using Kodak 1D Scientific Imaging System v.3.5.4.

To remove chromosomal DNA, plasmid DNA was treated with Plasmid-Safe™ ATP-Dependent DNase (Epicentre Technologies, Madison, WI). Eight-hundred and forty microliters of the plasmid DNA were added to an Eppendorf® RNA/DNA LoBind microcentrifuge tube (Sigma-Aldrich, St. Louis, MO) and mixed with 3 µl of Plasmid-Safe DNase, 100 µl of Plasmid-Safe 10X Reaction Buffer, 40 µl of 25 mM ATP solution and 6.25 µl of RNase A (Qiagen, Venlo, Netherlands). The mixture was incubated at 37° C for 30 min and the enzymes were inactivated by incubation at 70° C for 30 min. Plasmid DNA was purified in 200 µL aliquots using a 1X concentration of Ampure XP Beads (Beckman Coulter, Brea, CA) according to manufacturer's protocol and eluted using 30 µl of sterile ddH<sub>2</sub>O.

**Electroporation.** Plasmid DNA was precipitated with ethanol to remove residual salt and electroporated into the electrocompetent tetracycline susceptible ( $tet^S$ ) *E.coli* strain EC100 (Epicenter Technologies, Madison, WI using 1mm glass cuvettes) using a Bio-Rad GenePulser Xcell electroporator according to the manufacturer's instructions (Bio-Rad, Hercules, CA). EC100 was maintained and prepared for electroporation according to protocol 6.1.1 of the Bio-Rad® Gene pulser instruction manual.

Transformed cells were plated on tetracycline- ( $25 \mu\text{g ml}^{-1}$ ) amended TSA plates and incubated for 48 hours at  $37^\circ\text{C}$ . Colony presence was considered indicative of successful transformation. Electroporated transformants were used for antibiotic resistance profiling. Plasmids from the  $tet^R$  transformant isolates were isolated using the plasmid preparation protocol outlined above.

**Ion Torrent Personal Genome Machine Sequencing.** A 400 bp PGM library was constructed for plasmid pEG1-1 using the Ion Xpress™ Plus Fragment Library Kit (Thermo Fisher, Waltham, MA). DNA was fragmented using the Ion Plus Fragment Library Kit & Ion Shear™ Plus Reagents Kit. Ligation of PGM adapters, nick-repair, size selection and purification of the library was performed according to the manufacturer's instructions. Final library concentration was determined using the Qubit® dsDNA HS Assay Kit (Thermo Fisher, Waltham, MA). The prepared library was diluted to 100 pM with TE (10 mM Tris pH 8.0, 0.1 mM EDTA) and used to create a PGM template using the PGM™ Hi-Q™ OT2 Kit on the Ion OneTouch machine, according to manufacturer's instructions.

The template was prepared for sequencing using the Ion PGM™ Hi-Q™ Sequencing Kit (Thermo Fisher, Waltham, MA) and loaded onto an Ion 318™ Chip. Using the Torrent Suite™ Software, a sequencing run was executed using the “Generic Sequencing” template and the parameters listed in Table 1.

**Table 1.** Ion Torrent PGM run-analysis options.

| PGM Analysis Summary |                                  |
|----------------------|----------------------------------|
| Run Flows:           | 850                              |
| Flow Order:          | TACGTACGTCTGAGCATCGATCGATGTACAGC |
| Library Key:         | TCAG                             |
| 3' Adapter           | Ion P1B                          |
| Chip Type:           | 318C                             |
| Chip Data:           | Single                           |
| Barcode Set:         | IonXpress                        |
| Bead Loading Quality |                                  |
| Threshold:           | 30%                              |
| Plugins              | FileExporter <sup>a</sup>        |

<sup>a</sup>Converts sequence data to FastQ file format.

**Oxford Nanopore Technologies’ MinION Sequencing.** Plasmid DNA was fragmented using a partial restriction digest. Five-hundred-thirty seven microliters of purified plasmid DNA were mixed with 60 µL of NEBuffer 1.1 and 0.6 µL of 5,000 U/mL Sau3A1 enzyme (New England BioLabs, Ipswich, MA) and incubated at room temperature for one minute. The enzyme was immediately inactivated by incubating the solution at 70° C for 30 min.

The fragmented DNA was purified using a 1X concentration of Ampure XP Beads (Beckman Coulter, Brea, CA) according to the manufacturer’s protocol and eluted into 30 µl of sterile ddH<sub>2</sub>O. Successful fragmentation of plasmid DNA was confirmed by agarose gel electrophoresis (Figure 2) as described above. Fragmented DNA was quantified using the Qubit® dsDNA HS Assay Kit (Thermo Fisher, Waltham, MA). One

microgram of fragmented DNA was used to prepare a MinION sequencing library according to the manufacturer's protocol (Sequencing kit SQK-005, Oxford Nanopore, Oxford UK).

A sequencing run was created on MinKnow using the built-in python script "48Hr\_Sequencing\_Run\_SQK\_MAP005.py". MinION sequencing solution was prepared by mixing 75  $\mu$ L of 2x Running Buffer, 65  $\mu$ L ddH<sub>2</sub>O, 4  $\mu$ L Fuel Mix and 6  $\mu$ L of the prepared MinION sequencing library, and the solution was loaded into a R7.3 flow cell (Oxford Nanopore, Oxford UK). The preparation and loading of sequencing solution was repeated four times every six hours allowing for a total 24 hour sequencing run. MinION raw sequence data was basecalled using the Oxford Nanopore Metrichor software (2D Workflow, revision 1.9.1).

**Genome Assembly and Annotation.** PGM data were normalized to 100X coverage using the BBNorm normalization tool of the BBTool package (Bushnell, 2014) and MinION reads were error-corrected using Nanocorrect (Loman et al., 2015). Normalized PGM reads and Nanocorrected MinION reads were input into the SPAdes assembler V3.7 (Bankevich *et al.*, 2012) for a hybrid assembly using k-mer lengths 43, 53, 63, and 73 as well as the pipeline option "--careful". The reads were assembled into a single 74,302 kb contig. Because the plasmid genome was circular, a repeated sequence on both ends of the assembly was anticipated. To identify this region, the contig was split in half and the two halves were aligned to one another using Mauve (Darling *et al.*, 2010). This revealed a 982 bp repeated sequence on both ends of the assembly that was removed from one end. The final 73,320kb contig was first annotated using the automated annotation software Prokka (Seemann, 2014) and then manually annotated.

Manual annotation consisted of first assessing each annotation made by Prokka on the basis of % identity to the matched reference (plasmid pB8; <90% was removed), % coverage of the matched reference (<90% was removed), and e-value ( $>10^{-6}$  was removed). Prokka annotates genes as “hypothetical proteins” if the predicted gene cannot be matched to a reference within the large protein databases UniProt or Pfam (Seemann, 2014). These hypothetical proteins were manually extracted and locally aligned (BLASTp) to smaller, manually curated databases. The smaller databases used were CARD (McArthur *et al.*, 2013) and INTEGRALL (Soares, M. *et al.*, 2009). Areas of the genome >100 bp that did not receive a Prokka annotation were extracted and also aligned to the smaller databases described above. Prior to adding a manual annotation, the presence of a predicted protein-coding gene was verified using GLIMMER 2.1 (Delcher *et al.* 1999). Global alignments of homologous regions between pEG1-1 and the reference plasmid pB8 were also utilized to confirm annotations.

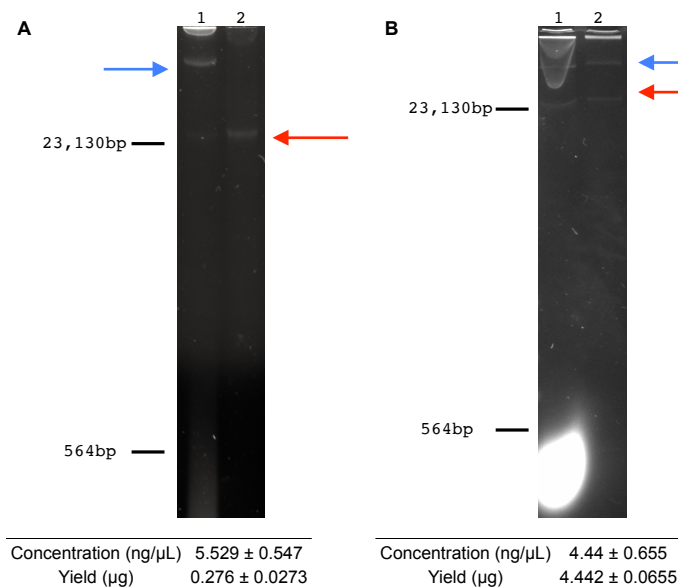
## Results

**Methods Development to Prepare Plasmid DNA for Sequencing on Next-Generation Platforms.** The removal of RNA and chromosomal DNA contaminants was verified through agarose gel electrophoresis (Figure 1). Lane 1 of both gels contains the raw product of a plasmid prep protocol. In Figure 1b, RNA was observed as fragments <500bp and chromosomal DNA was observed as smears near the top of Lane 1. Lane 2 of both gels contains plasmid DNA after Plasmid-Safe and RNase A treatment. In both gels, no RNA or chromosomal contaminants were observed in Lane 2.

Large quantities of DNA are necessary to sequence a plasmid on multiple NGS platforms. The MinION sequencing protocol requires 1 $\mu$ g of sample DNA and the PGM protocol requires at least 0.10 $\mu$ g. The average yield from the plasmid miniprep developed in our lab was only around 0.3  $\mu$ g (Figure 1). Running multiple plasmid preps one after another is impractical and time consuming, and running multiple preps in parallel is prone to error since many of the reactions within the protocol are time-sensitive. Instead, we elected to scale-up the original plasmid prep protocol by a factor of 20.

The efficacy of the scaled-up protocol was verified through agarose gel electrophoresis (Figure 1b). Lane 1 contains the raw product of the scaled-up plasmid prep. Plasmid DNA in both a supercoiled state (band at ~20kb) and relaxed conformation (band at >20kb) was observed. Greater quantities of RNA and chromosomal DNA contaminants compared to the original protocol (Figure 1a) were also present, but both contaminants were fully removed after Plasmid-Safe and RNase A treatment (Figure 1b). Concentration and yield of the plasmid DNA were obtained by first fragmenting the





**Figure 1.** Original and scaled-up plasmid preparation protocols. (A) Isolated pEG1-1 using the original protocol. (B) Isolated pEG1-1 using the scaled-up protocol. Samples were run with the following lane assignments - Lane 1: raw product from the plasmid prep, Lane 2: plasmid after Plasmidsafe and RNase treatment. The blue and red arrows indicate plasmid DNA in either a supercoiled or relaxed conformation, respectively. The molecular weights indicated to the left of each gel image are based on a Lambda DNA/HindIII marker run in each gel. The concentration of plasmid DNA was determined using the Qubit BR assay kit. The concentration was multiplied by the final volume of the sample to determine yield. The data shown is the average concentration and yield of 7 samples with a standard deviation expressed as a  $\pm$  value.

plasmid DNA with Sau3AI and then quantifying each sample with the Qubit dsDNA BR Assay Kit. The average yield of the scaled-up protocol was 4.44  $\mu$ g, a nearly 20-fold increase compared to the average yield obtained from the original plasmid prep.

All NGS platforms require sample DNA to be linear in order to ligate the appropriate oligonucleotide sequences and/or specialized adapter proteins required for each platform's sequencing chemistry (Morey et al., 2013). In addition, the PGM requires shearing to an average size of 400 bp. The PGM sequencing protocol employs an

enzymatic fragmentation method that has been optimized and commercially available for over five years (Ion Torrent, 2014). Plasmid DNA that has been treated with Plasmid-Safe and RNase can be input directly into the PGM protocol for successful sequencing. At the time of this study, however, the MinION sequencing protocol was still in beta testing through the MinION Access Programme and not yet optimized nor commercially available.

The suggested method for fragmenting the DNA library to be sequenced by the MinION was to use the Covaris G-tube. The G-tube shears DNA by using centrifugal force to push the sample through a precisely manufactured pore. The size of the fragments produced is dependent on the amount of centrifugal force applied. Unfortunately, our attempts to fragment plasmid DNA with a G-tube were unsuccessful: sequence runs utilizing the G-tube resulted in almost zero reads, suggesting that the plasmid was not being sufficiently linearized for proper ligation of adapter proteins during library preparation. It was hypothesized that the supercoiled state of plasmid DNA impeded successful shearing through the G-tube. Instead, fragmentation was accomplished through the use of partial restriction digestion with the four-base restriction enzyme Sau3AI.

Successful fragmentation was verified via agarose gel electrophoresis (Figure 2). Lanes 1 and 2 show the raw product of a scaled-up plasmid prep and circularized plasmid DNA after Plasmid-Safe and RNase A treatment, respectively. Lane 3 contains fragmented plasmid DNA after the partial restriction digest. A smear, indicating multiple fragments, from ~20kb to ~300kb was observed.



**Figure 2.** Plasmid DNA preparation for MinION sequencing. Lane 1: raw product from the scaled-up plasmid prep. Lane 2: purified plasmid product after Plasmidsafe and RNase treatment. Lane 3: fragmented plasmid DNA after partial digestion using the four-base restriction enzyme sau3AI. Molecular weights are estimates based on a Lambda DNA/HindIII marker run in the same gel.

**Sequencing on the Ion Torrent PGM.** For PGM sequencing, two<sup>1</sup> 400 bp libraries were prepared using plasmid pEG1-1 DNA that had been isolated and purified of RNA and chromosomal DNA contaminants using the methods described in the previous chapter. The libraries were pooled together and clonally amplified through emulsion PCR on Ion Sphere Particles<sup>TM</sup> (ISP), which were then loaded onto an Ion 318 Chip for sequencing on the Ion Torrent PGM. The sequence run diagnostics are presented in Figure 3.

The ISP Load and color map are measures of Ion Sphere Particles (ISP) loaded into the microwells of a sequencing chip (Figure 3a). The ISP Load is the percentage of microwells loaded with an ISP and is a metric of proper library preparation and chip-

---

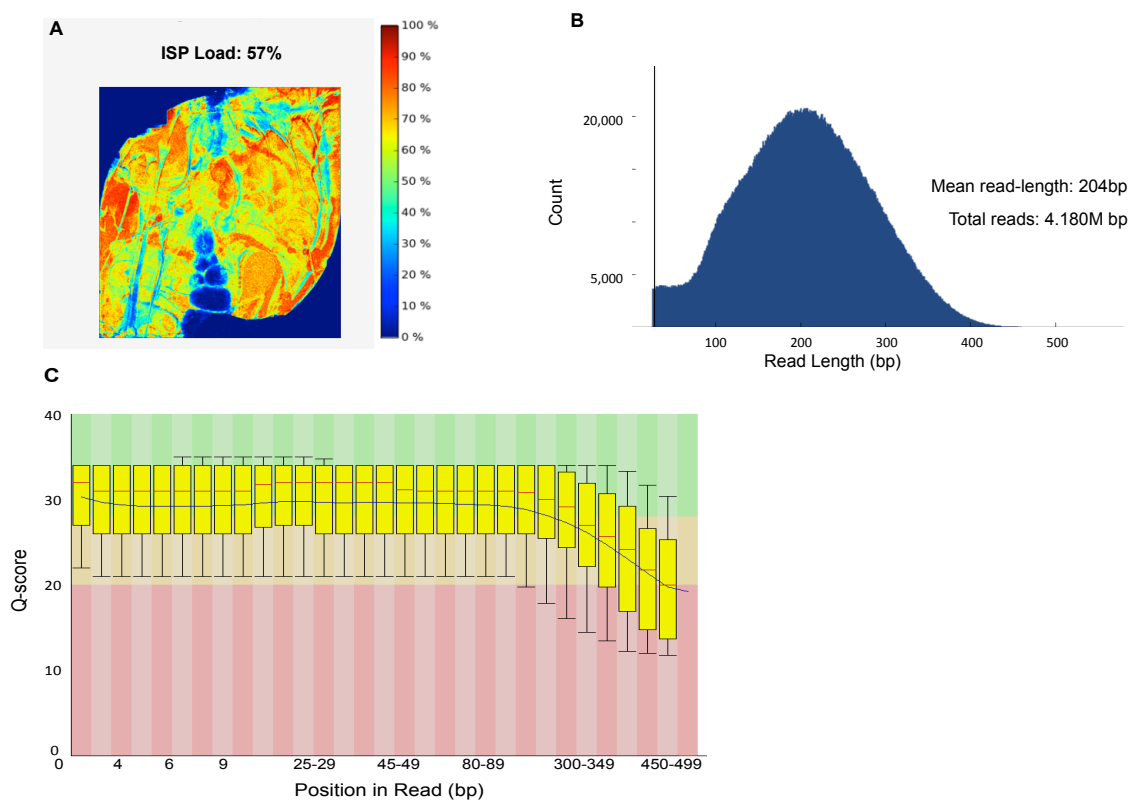
<sup>1</sup>Two separate PGM libraries were prepared with the assumption that each library was generated from two different plasmids. However, after performing a Modified Stokes disk-diffusion assay (Gehr, 2013), it was revealed that the two libraries were both generated from plasmid pEG1-1. Therefore, the PGM data from both libraries were combined for assembly.

loading (Ion Torrent, 2015). An ISP Load of 30% is a quality threshold set by the Torrent Suite™ software (Ion Torrent, 2015). An ISP Load below 30% is indicative of a failed run. The ISP Load generated during the sequence run of plasmid pEG1-1 was 57%, exceeding the quality threshold.

The color map is a representation of ISP loading distribution and is a second proxy for proper chip loading. The color blue indicates 0% ISP loading and red indicates 100% ISP loading in the represented area. The map is represented as a square, but the chip has rounded edges. For this reason, the top left and bottom right corners are completely blue. According to the manufacturer's protocol, an even distribution of ISP is ideal (Ion Torrent, 2015). The color map generated during the sequence run of plasmid pEG1-1 was a marbled mix of yellow and red, but contained a large region of blue on the bottom of the chip (Fig 3a). The chip-loading procedure was complex and dependent on a number of steps (Ion Torrent, 2014). The researcher conducting the chip-loading procedure had little experience at the time with such a protocol and it was assumed that the large region of low ISP density was a product of poor chip-loading technique.

The PGM produced 4.180 million reads with an average read-length of 204 bp (Figure 3b). Two 400 bp libraries were prepared; thus an average read-length of 400 bp was expected. However, similar to the chip-loading protocol, the Ion Xpress™ Plus Fragment Library protocol consists of many procedures new to the researcher conducting the library preparation. Operator error, in addition to the uneven ISP density, may have led to improper size selection of DNA fragments.

Q-score, or Phred quality score, is the standard metric of read quality and has an inverse, logarithmic relation to the probability of error of a base call (Ewing, 1998). For example,  $Q = 10$  indicates a 1 in 10 probability of error,  $Q = 20$  indicates a 1 in 100 probability of error,  $Q = 30$  indicates a 1 in 1,000 probability of error and so on. A plot of the range of Q-scores across all bases at each position was generated on FastQC (Andrews, 2010) (Figure 3c). The average Q-score of base calls in positions <300bp bp



**Figure 3.** Ion Torrent PGM sequence. (A) Map of Ion Sphere Particles (ISP) loaded onto the Ion 318™ Chip for sequencing of plasmid pEG1-1. Color indicates the percentage of wells occupied by an ISP in a given area. The ISP Load is the total percentage of wells loaded with an ISP. (B) Read-length histogram of sequence data produced. (C) FastQC-generated plot (Andrews, 2010) of the range of quality scores across all bases at each position. For each position a BoxWhisker plot is drawn. The central red line is the median value, the yellow box represents the inter-quartile range (25-75%), the upper and lower whiskers represent the 10% and 90% points, and the blue line represents the mean quality

was greater than  $Q = 20$  (i.e. >99% base call accuracy). Thus, despite an average read-length shorter than expected and regions of low ISP density, high quality reads were obtained from the PGM sequence run.

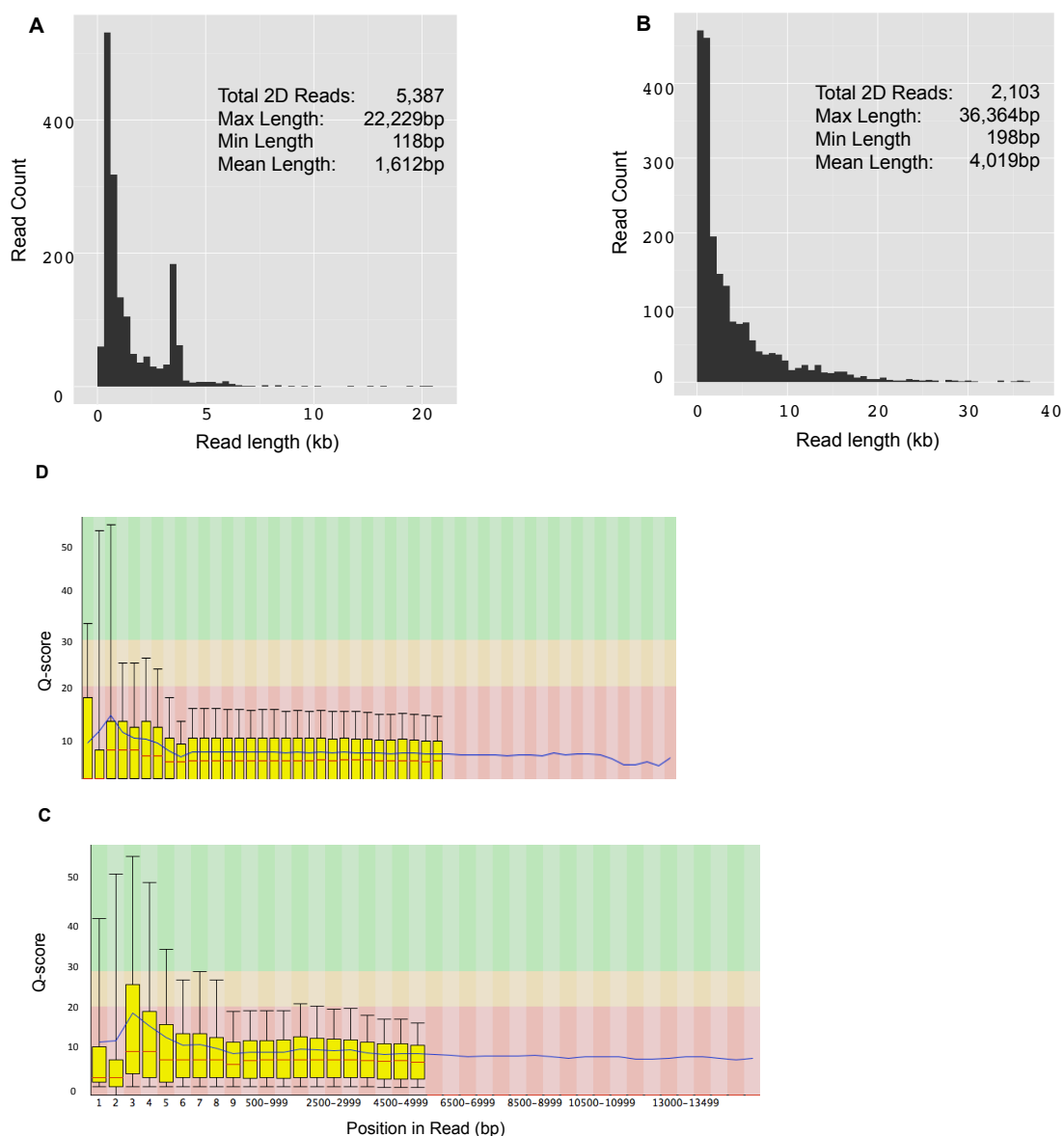
**Sequencing on Oxford Nanopore Technologies' MinION.** Plasmid pEG1-1 was isolated, purified of RNA and chromosomal DNA contaminants and fragmented via partial digest using the methods described above. Fragmented pEG1-1 DNA was used to generate two<sup>2</sup> MinION sequence libraries. As part of the library preparation protocol, a hairpin-adaptor was ligated to the end of DNA fragments (Sequencing kit SQK-005, Oxford Nanopore, Oxford UK). The hairpin-adaptor allows for both the template and complement strand of DNA to traverse a nanopore during a sequencing run. Reads produced by combining data from both strands are known as 2D reads and have a higher accuracy compared to reads produced from just the template strand, termed 1D reads (Loman et al., 2015). For this study, only the higher quality, 2D reads.

Two sequence runs (Run 1 and 2) were conducted for each library separately. Both runs were performed using R7.3 flow cells (Oxford Nanopore, Oxford UK). After sequencing, the raw data (i.e. current disruptions over time) of both runs were base called – i.e. converted to sequence reads – through the Oxford Nanopore program *Metrichor* (2D Workflow, revision 1.9.1), and separated into 1D and 2D reads. Sequence Run 1 and 2 diagnostics are based on the data of 2D reads (Figure 4).

The previous section showed that plasmid DNA can be fragmented to 20 kb using the methods we have developed. However, at the time of both sequence runs, the use of

---

<sup>2</sup>Two sequence runs on plasmid pEG1-1 were conducted with the same assumptions as noted previously with the PGM sequencing (i.e. it was falsely assumed that the two libraries were generated from two different plasmids). As with the PGM data, the data generated from both MinION runs were combined for the assembly.



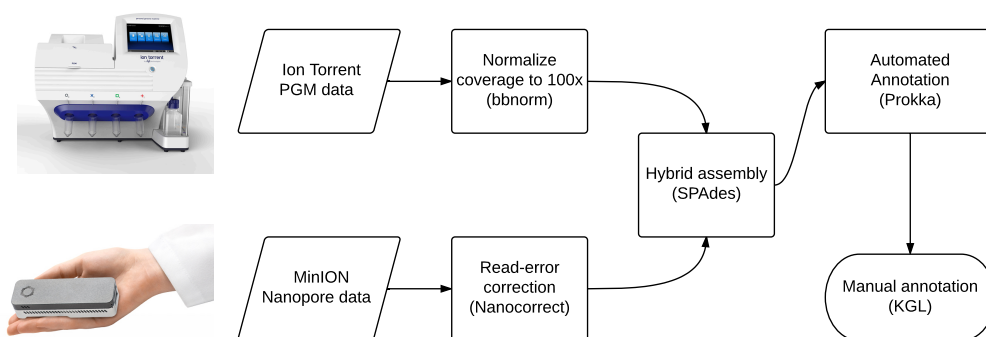
**Figure 4.** Oxford Nanopore Technologies' MinION sequence diagnostics. (A, B) Read-length histogram of sequence data produced in Run 1 and Run 2, respectively. (C, D) FastQC-generated plot (Andrews, 2010) of the range of quality scores across all bases at each position in Run 1 and Run 2, respectively. For each position a Box Whisker plot is shown. The central red line is the median value, the blue line represents the mean, the yellow box represents the inter-quartile range (25-75%), and the upper and lower whiskers represent the 10% and 90% points.

restriction digestion had not been optimized to generate 20 kb fragments. Fragmentation of plasmid DNA used in Run 1 and Run 2 were visualized on an agarose gel prior to preparing the sequence libraries. The gels revealed fragmented plasmid DNA ranging from 7 kb to <500 bp for the DNA used in Run 1 and from 10 kb to <500 bp for the DNA used in run 2. This was consistent with the distribution of read-lengths produced during each run (Fig 4a,b). Run 1 generated 5,387 2D reads with an average read-length of 1,612 bp and the largest read produced in Run 1 was 22,229 bp. Run 2 generated 2,103 2D reads with an average read-length of 4,019 bp and the largest read produced in Run 2 was 36,364 bp. The average read-lengths from Run 1 and 2 exceed the average PGM read-length by a factor of 10 and 20, respectively.

As with the PGM data, a range of Q-scores across all bases at each position was generated on FastQC (Andrews, 2010) for both Run 1 and 2 (Fig 4c, d). Both plots show that the Q-score of all base calls averages are ~10, (i.e. ~90% base-call accuracy).

**Hybrid Assembly.** Before combining PGM and MinION data for hybrid assembly, both datasets were adjusted to ensure optimal input-data quality (Figure 5) The PGM reads were normalized to 100x coverage using the program *bbnorm* (Bushnell, 2014). This was done to adjust for the uneven coverage bias associated with the platform (Quail *et al.* 2012). An assembly without normalization was attempted, but PGM depth-of-coverage, or the number of times a specific base was sequenced, ranged from >2,000X to 50X. An uneven depth-of-converge as drastic as this hinders the assembly process and can introduce error into constructed contigs (Chen *et al.*, 2013). MinION reads were





**Figure 5.** Sequencing, assembly and annotation pipeline for plasmid pEG1-1.

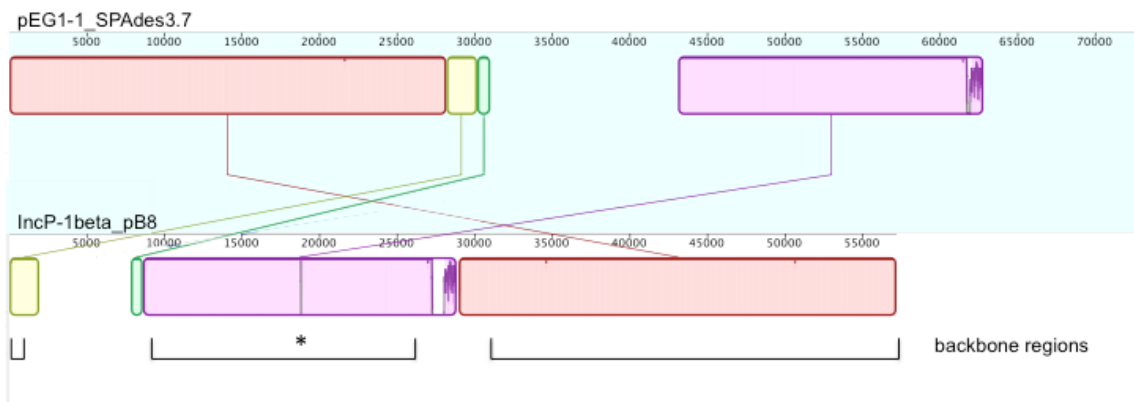
processed using Nanocorrect (Loman *et al.*, 2015) to increase the quality of the low-quality reads produced in the MinION sequence runs (Figure 4). Nanocorrect was designed to increase the quality of MinION reads through the use of a multiple alignment process and the raw signal traces produced during the sequence run (Loman *et al.*, 2015). Loman *et al.* showed that Nanocorrected MinION reads increased the accuracy of a *de novo* assembly from 80.5% to 95.9% (Loman *et al.*, 2015)

SPAdes, a *de Bruijn*-graph based assembler that constructs contigs (a continuous sequence constructed by overlapping reads) by merging assemblies based on multiple k-mer (substring of length  $k$ ) sizes (Bankevich *et al.*, 2012), was used to assemble the normalized PGM reads and Nanocorrected MinION reads into a single 74,212 bp contig (Figure 5). A 982 bp sequence was repeated on both ends of the contig, suggesting that a circular genome could be assembled. This region was removed from one end of the assembly, resulting in a final 73,230 bp contig.

Accurate *de novo* assemblies of bacterial genomes can be generated with a depth-of-coverage as low as 50X (Desai *et al.*, 2013), and the quality of SPAdes assemblies has been shown to drop with a depth-of-coverage over 700X (Lonardi *et al.*, 2015). For these reasons, an acceptable depth-of-coverage range between 50X and 700X was established. Depth-of-coverage of the assembly was determined by mapping all input data to the generated contig. The average depth-of-coverage was found to be 287X. Two regions (549 bp and 33 bp in length) had a depth-of-coverage <50X and no area exceeded 700X.

A BLASTn search (Altschul *et al.*, 1997) of the entire pEG1-1 contig against the NCBI nucleotide collection database (accessible at: <http://www.ncbi.nlm.nih.gov/nucleotide>) showed that our assembly matched most closely to plasmid pB8, a multi-drug resistance plasmid within the IncP-1 $\beta$  incompatibility group (Schlüter *et al.*, 2005). IncP-1 $\beta$  plasmids have genomes that consist of a conserved backbone region, interspersed with variable regions – collectively known as the “accessory region” of the plasmid – that often contain genetic elements such as transposons and integrons (Popowska & Krawczyk-Balska, 2013). Mauve (Darling *et al.*, 2010) was used to perform a global alignment between pB8 and the pEG1-1 assembly. The alignment revealed four homologous regions that completely encompass the conserved, backbone region of pB8 (Figure 6.)

When considering only the backbone region, pB8 and pEG1-1 align with >99.99% identity; only one 24 bp region of the pB8 backbone was not identical to the pEG1-1 assembly. The high degree of similarity between the conserved backbone of pB8 and the pEG1-1 assembly support the validity of the assembly. The accessory regions of pB8 have some similarity with the pEG1-1 assembly, but most was dissimilar. This was



**Figure 6.** Global alignment of the pEG1-1 assembly and pB8. The pairs of colored boxes indicate homologous regions between the pEG1-1 assembly (top) and pB8 reference genome (bottom). A similarity profile that is inversely proportional to the average alignment entropy over a region of the alignment is drawn within each box. The backbone regions of pB8 have been marked at the bottom of the figure. An asterisk has been placed at the single dissimilar region of 24 bp between the pB8 backbone and the pEG1-1 assembly.

not believed to be an indication of error, rather that pEG1-1 carries an accessory load that differs from pB8.

**Annotation.** The pEG1-1 assembly was annotated using Prokka, an automated annotation software that generates gene annotations rapidly by predicting coding regions, translating those regions into an amino acid sequence, and comparing the amino acid sequence to databases of known proteins using BLASTp (Seemann, 2014). Predicted coding regions that match to a known protein with an e-value  $<10^{-6}$  are annotated as the matched reference. If no match is found to a predicted coding region, it is annotated as a “hypothetical protein”. Each annotation generated by Prokka was manually inspected to confirm its validity based on %ID to the matched reference, e-value, and presence of ribosomal binding sites. The homologous regions of pB8 were also a large factor in the manual annotation process of the plasmid pEG1-1 backbone. In total, 78 gene annotations

were made on the pEG1-1 assembly, of which 6 remained “hypothetical proteins” after both automated and manual annotation (Table 2).

**Table 2.** Complete list of pEG1-1 genes.

| <b>Gene<sup>a</sup></b>   | <b>Predicted Function</b>      | <b>Start<sup>b</sup></b> | <b>Stop<sup>b</sup></b> | <b>Strand</b> |
|---------------------------|--------------------------------|--------------------------|-------------------------|---------------|
| <u>Accessory Module 1</u> |                                |                          |                         |               |
| <i>merR</i>               | involved in mercury resistance | 363                      | 512                     | R             |
| hypothetical protein      | unkown                         | 627                      | 1049                    | R             |
| hypothetical protein      | unkown                         | 1152                     | 2114                    | F             |
| <i>strB</i>               | aminoglycoside resistance      | 2320                     | 2949                    | R             |
| <i>repA</i>               | IncU plasmid replicase         | 3005                     | 3499                    | R             |
| <i>mazF</i>               | toxin/antitoxin system         | 3851                     | 4174                    | R             |
| <i>mazE</i>               | toxin/antitoxin system         | 4174                     | 4398                    | R             |
| hypothetical protein      | Uunknown                       | 5182                     | 5934                    | F             |
| <i>neo</i>                | aminoglycoside resistance      | 7199                     | 8014                    | F             |
| hypothetical protein      | unknown                        | 8698                     | 8889                    | R             |
| hypothetical protein      | unknown                        | 8952                     | 9080                    | F             |
| <i>tetC</i>               | tetracycline resistance        | 9136                     | 10326                   | R             |
| <i>tetA</i>               | tetracycline resistance        | 10419                    | 11054                   | F             |
| <i>Tnp IS200</i>          | transposase                    | 11077                    | 11412                   | R             |
| <i>tnpA</i>               | transposase                    | 13091                    | 14311                   | F             |
| <u>Initiaiton Module</u>  |                                |                          |                         |               |
| <i>trfA</i>               | IncP-1 $\beta$ replication     | 14934                    | 16154                   | R             |
| <i>ssb</i>                | IncP-1_ replication            | 16201                    | 16542                   | R             |
| <u>Tra1 Module</u>        |                                |                          |                         |               |
| <i>trbA</i>               | IncP-1 $\beta$ transfer        | 16656                    | 17018                   | F             |
| <i>trbB</i>               | IncP-1 $\beta$ transfer        | 17328                    | 18290                   | F             |
| <i>trbC</i>               | IncP-1 $\beta$ transfer        | 18307                    | 18771                   | F             |
| <i>trbD</i>               | IncP-1 $\beta$ transfer        | 18775                    | 19086                   | F             |
| <i>trbE</i>               | IncP-1 $\beta$ transfer        | 19083                    | 21641                   | F             |
| <i>trbF</i>               | IncP-1 $\beta$ transfer        | 21638                    | 22420                   | F             |
| <i>trbG</i>               | IncP-1 $\beta$ transfer        | 22417                    | 23337                   | F             |

Table 2. Continued.

| Gene <sup>a</sup>         | Predicted Function             | Start <sup>b</sup> | Stop <sup>b</sup> | Strand |
|---------------------------|--------------------------------|--------------------|-------------------|--------|
| <i>trbH</i>               | IncP-1 $\beta$ transfer        | 23340              | 23828             | F      |
| <i>virB10</i>             | IncP-1 $\beta$ transfer        | 23833              | 25233             | F      |
| <i>trbJ</i>               | IncP-1 $\beta$ transfer        | 25254              | 26018             | F      |
| <i>trbK</i>               | IncP-1 $\beta$ transfer        | 26028              | 26255             | F      |
| <i>trbL</i>               | IncP-1 $\beta$ transfer        | 26266              | 27984             | F      |
| <i>trbM</i>               | IncP-1 $\beta$ transfer        | 28002              | 28589             | F      |
| <i>trbN</i>               | IncP-1 $\beta$ transfer        | 28603              | 29238             | F      |
| <i>trbO</i>               | IncP-1 $\beta$ transfer        | 29267              | 29533             | F      |
| <i>trbP</i>               | IncP-1 $\beta$ transfer        | 29533              | 30231             | F      |
| <i>upf30.5</i>            | IncP-1 $\beta$ transfer        | 30247              | 30678             | F      |
| <i>upf31.0</i>            | IncP-1 $\beta$ transfer        | 30833              | 31507             | F      |
| <u>Accessory Module 2</u> |                                |                    |                   |        |
| <i>intl1</i>              | Integrase                      | 31754              | 32767             | R      |
| <i>aadA</i>               | Aminoglycoside resistance      | 33225              | 34064             | F      |
| <i>qacEdelta1</i>         | Integron/antiseptic resistance | 34228              | 34575             | F      |
| <i>sul1</i>               | Sulfonamide resistance         | 34569              | 35540             | F      |
| <i>floR</i>               | Florfenicol resistance         | 35757              | 36971             | F      |
| <i>tetR</i>               | Tetracycline resistance        | 37178              | 37804             | R      |
| <i>tetA</i>               | Tetracycline resistance        | 37908              | 39083             | F      |
| <i>gltC</i>               | Transposase                    | 39173              | 39895             | F      |
| <i>Tnp</i>                | Putative transposase           | 39987              | 41519             | R      |
| <i>groEL/inl1</i>         | Integrase fusion protein       | 41746              | 42399             | R      |
| <i>pse4</i>               | Betalactamase                  | 42557              | 43471             | F      |
| <i>aadA</i>               | Aminoglycoside resistance      | 43571              | 44380             | F      |
| <i>qacEdelta1</i>         | Integron/antiseptic resistance | 44544              | 44891             | F      |
| <i>sul1</i>               | Sulfonamide resistance         | 44885              | 45724             | F      |
| hypothetical protein      | Unknown                        | 45852              | 46352             | F      |
| <i>tniB</i>               | Transposase                    | 46321              | 47313             | R      |
| <i>tniA</i>               | Transposase                    | 47316              | 48995             | R      |
| <u>Tra2 Module</u>        |                                |                    |                   |        |
| <i>traC</i>               | Type IV secretion system       | 49546              | 53891             | R      |
| <i>traD</i>               | Type IV secretion system       | 53895              | 54284             | R      |
| <i>traE</i>               | Type IV secretion system       | 54306              | 56369             | R      |
| <i>traF</i>               | Type IV secretion system       | 56381              | 56917             | R      |
| <i>traG</i>               | Type IV secretion system       | 56914              | 58827             | R      |
| <i>traI</i>               | Type IV secretion system       | 58824              | 61064             | R      |
| <i>traJ</i>               | Type IV secretion system       | 61099              | 61473             | R      |
| <i>traK</i>               | Type IV secretion system       | 61847              | 62245             | F      |

**Table 2. Continued.**

| <b>Gene<sup>a</sup></b>            | <b>Predicted Function</b>        | <b>Start<sup>b</sup></b> | <b>Stop<sup>b</sup></b> | <b>Strand</b> |
|------------------------------------|----------------------------------|--------------------------|-------------------------|---------------|
| <i>traL</i>                        | IncP-1_ type IV secretion system | 62245                    | 62970                   | F             |
| <i>traM</i>                        | IncP-1_ type IV secretion system | 62970                    | 63410                   | F             |
| <u>Regulation/Stability Module</u> |                                  |                          |                         |               |
| <i>kfrC</i>                        | IncP-1_ regulation & stability   | 63613                    | 64266                   | R             |
| <i>kfrB</i>                        | IncP-1_ regulation & stability   | 64295                    | 64642                   | R             |
| <i>kfrA</i>                        | IncP-1_ regulation & stability   | 64813                    | 65844                   | R             |
| <i>korB</i>                        | IncP-1_ regulation & stability   | 66024                    | 67073                   | R             |
| <i>incC2</i>                       | IncP-1_ regulation & stability   | 67070                    | 67834                   | R             |
| <i>korA</i>                        | IncP-1_ regulation & stability   | 67831                    | 68133                   | R             |
| <i>kleF</i>                        | IncP-1_ regulation & stability   | 68247                    | 68777                   | R             |
| <i>kleE</i>                        | IncP-1_ regulation & stability   | 68779                    | 69108                   | R             |
| <i>kleB</i>                        | IncP-1_ regulation & stability   | 69253                    | 69468                   | R             |
| <i>kleA</i>                        | IncP-1_ regulation & stability   | 69527                    | 69763                   | R             |
| <i>korC</i>                        | IncP-1_ regulation & stability   | 69923                    | 70180                   | R             |
| <i>klcB</i>                        | IncP-1_ regulation & stability   | 70197                    | 71402                   | R             |
| <i>klcA</i>                        | IncP-1_ regulation & stability   | 71454                    | 71882                   | R             |
| <i>kluA</i>                        | IncP-1_ regulation & stability   | 72051                    | 72320                   | F             |
| <i>kluB</i>                        | IncP-1_ regulation & stability   | 72699                    | 72990                   | F             |

<sup>a</sup>Listed according to distance from *oriV* on the forward strand.

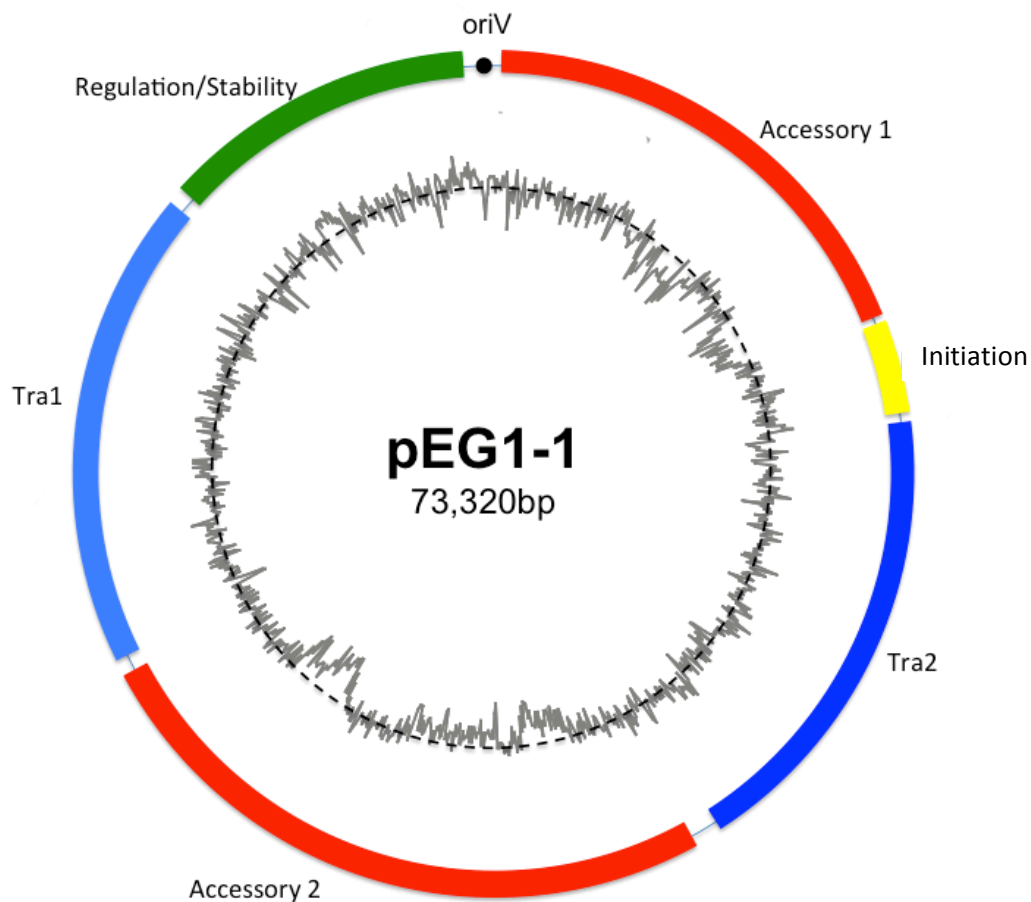
<sup>b</sup>Base pair position from *oriV*.

**pEG1-1 Resistance Phenotype.** A modified Stokes disc diffusion antibiotic susceptibility assay was used by Gehr to determine that plasmid pEG1-1 conferred decreased susceptibility to tetracycline, kanamycin, piperacillin, ticarcillin, tobramycin, piperacillin/tazobactam, and cefepime (Gehr, 2013). This resistance profile was confirmed in the present study. Minimum inhibitory concentrations (MICs) of 22 antibiotics—including ticarcillin, tobramycin, piperacillin/tazobactam, and cefepime—were determined for pEG1-1. Clinical levels of antibiotic resistance, as defined by NCCLS standards for Enterobacteriaceae, to ticarcillin and tobramycin were found: MICs of 4 $\mu$ g/mL and 64 $\mu$ g/mL, respectively. Clinical levels of resistance to the other antibiotics tested, including piperacillin/taxobactam and cefepime were not observed:

2µg/mL for doxycycline and no growth observed on the MIC plate in the presence of all other antibiotics.

**Functional Modules of pEG1-1.** The 78 genes found on pEG1-1 were classified into seven functional modules based on the cellular interactions of the encoded proteins (Figure 7). The initiation module was made up of two genes as well as the origin of vegetative replication (*oriV*). The module designated “regulation/stability”, was comprised of 15 genes encoding proteins with predicted functions involved in maintenance, partitioning, and plasmid control. The Regulation/Stability and Initiation modules are separated by Accessory Module 1, a functional module comprised of a single, large *Tn21*-related transposon. Genes encoding proteins involved in mate-pair formation and conjugative DNA-transfer are contained within two functional modules, Tra1 and Tra2. Tra1 consists of 10 genes and Tra2 was made up of 18 genes. The modules are separated by Accessory Module 2, a functional module comprised of a single complex integron, *in104*.

The modules were divided into two categories: the backbone and accessory regions. The backbone is a conserved set of genes amongst IncP-1β plasmids and consists of the regulation/stability, Tra1 and Tra2 modules (Schlüter *et al.*, 2005). A comparison between other IncP-1β plasmids showed that pEG1-1 possesses a complete set of backbone genes with an organization nearly completely identical to the backbone of pB8 (Schlüter *et al.*, 2005), suggesting pEG1-1 to be a part of the IncP-1β incompatibility group. This was confirmed using the integrated web tool PlasmidFinder (Carattoli *et al.*, 2014). The accessory region consists of the Accessory Modules 1 and 2.

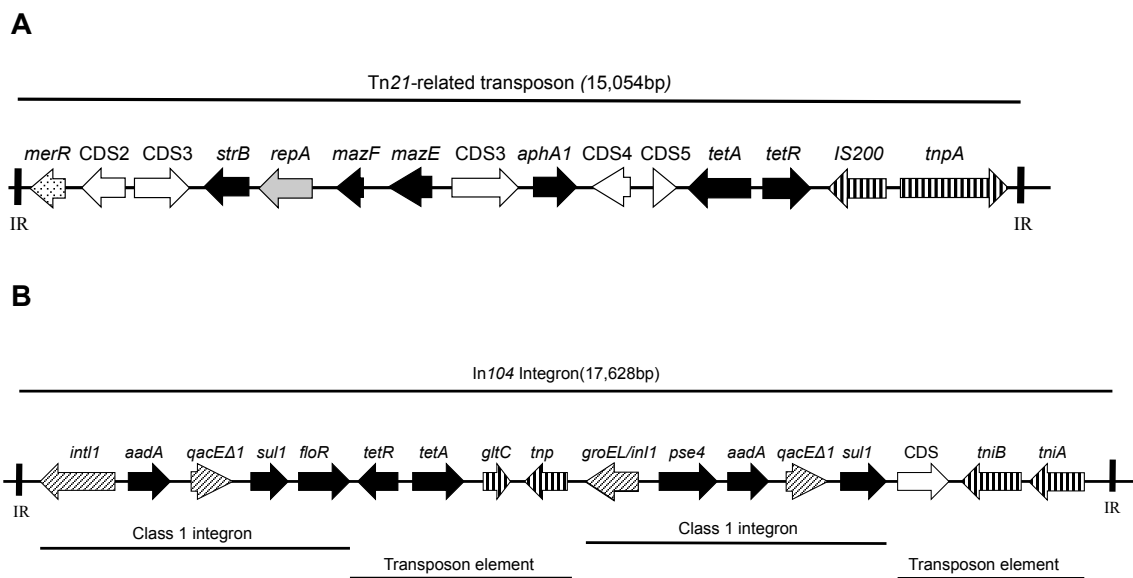


**Figure 7.** The seven functional modules mapped along the multi-drug resistance IncP-1 $\beta$  plasmid pEG1-1 genome. The origin of vegetative replication (*oriV*) is marked as a black circle at the top of the circle. The regulation/stability module (green), Tra1 (light blue), Tra2 (dark blue) and replication module (yellow) make up the IncP-1 $\beta$  backbone genes. Accessory modules 1 & 2 make up the accessory region and are marked in red. The inner gray plot indicates the average GC% throughout the plasmid with the dotted line representing 50% GC content—inside the dotted line <50% GC.

Accessory Modules 1 and 2 contain unique mobile genetic elements that are not observed in any other sequenced IncP-1 $\beta$  plasmid (Schlüter *et al.*, 2007), but have been observed on plasmid of other incompatibility groups (Douard *et al.*, 2010; Nass *et al.*, 2013). Dramatic shifts in the average percent GC of these two modules suggest that these modules may have been recently incorporated into the pEG1-1 genome via recombination events (Figure 7).



Accessory Module 1 (AM1) was made up entirely of a 15,054 bp transposon with structures related to the Tn21 family, a group of transposons associated with the global dissemination of antibiotic resistance (Figure 8a) (Liebert *et al*, 1999). Tn21 transposons



**Figure 8.** Genetic map of Accessory Modules 1 & 2. (A) The Tn21-related transposon of AN1. (B) The In104 integron of AM2. The vertical black bars represent inverted repeat regions (IR). Coding regions are shown by arrows indicating the direction of transcription. Black arrows represent genes conferring antibiotic resistance, arrows with vertical stripes represent conserved transposon genes, arrows with diagonal stripes represent conserved integron genes, and white arrows represent coding sequences with no known function. The *merR* gene and *repA* gene of the AM1 transposon are represented as a dotted and gray arrow, respectively. The genes comprising the class 1 integrons and transposon elements of the In104 integron are marked at the bottom.

are characterized by the presence of multiple antibiotic resistance genes, a *tnpA* transposase, a *tnpR* relaxase, a mercury resistance (*mer*) operon and a pair of Tn21-specific inverted repeats flanking the transposon (Liebert *et al*, 1999). A *tnpA* transposon and a pair of Tn21 inverted repeats were observed within the AM1 transposon, but it lacked a *tnpR* relaxase and full *mer* operon—only the *merR* gene of the operon was found on the AM1 transposon. Also unique was the presence of the *repA* gene, encoding the

replicase enzyme of IncU plasmids (Nass *et al.*, 2013). Six genes related to antibiotic resistance were found on the AM1 transposon: the genes *strB* and *aphA1* confer resistance to ampicillin and streptomycin, *tetC* and *tetR* confer resistance to tetracycline, and the *mazEF* toxin-antitoxin system has been associated with decreased susceptibility to  $\beta$ -lactam antibiotics (Schuster *et al.*, 2015).

Accessory Module 2 (AM2) was made up entirely of a 17,628bp mobile genetic element classified as an *in104* complex integron (Figure 8b). *In104* complex integrons are found within the *Salmonella* genomic island 1 (SGI1) and contain two class 1 integrons. Also characteristic of *in104* complex integrons is the presence of transposase enzymes and inverted repeats flanking the entire element. All of these characteristics were observed in the AM2 element, as well as eight antibiotic resistance genes: the two *aada* genes confer resistance to aminoglycoside antibiotics such as streptomycin and spectomycin, the two *sulI* genes confer resistance to sulfonamide antibiotics, both *tetR* and *tetA* confer tetracycline resistance, the *floR* gene confers resistance to chloramphenicol and florfenicol and the *pse4* gene confers resistance to a broad range of  $\beta$ -lactam antibiotics.

## Discussion

**Sequencing, Assembly and Annotation.** Genome studies of multi-drug resistance plasmids can provide critical insight into the reservoir of antibiotic resistance genes present in stream sediment. Unfortunately, the plasmid miniprep previously developed in our laboratory produced low-yields of plasmid DNA that were contaminated with RNA and chromosomal DNA. These factors made this protocol initially an impractical option to prepare plasmid DNA for sequencing. In this study, we have shown that Plasmid-Safe™ ATP-Dependent DNase and RNase A enzymes can successfully remove the RNA and chromosomal DNA contaminants and that the plasmid miniprep protocol can be scaled up by a factor of 20 to obtain larger quantities of plasmid DNA. These methods allow one to obtain large quantities of pure plasmid DNA for sequencing on NGS platforms. In addition, we were able to develop a fragmentation method using partial restriction digestion to prepare plasmid DNA for MinION sequencing.

The sequence data obtained from the PGM and MinION were typical of each platform (Acinas et al., 2005, Wei & Williams, 2015). The PGM reads were short (~200 bp), but had high phred scores indicative of over 99% base-call accuracy (Figure 3). The MinION reads were long (mean ~1.6 kb and 4 kb for Run 1 and 2, respectively), but had lower phred scores indicative of only 90% base-call accuracy (Figure 4).

The relatively complex Ion Torrent sequence protocol led to shorter read-lengths than anticipated and an uneven ISP Density across the sequencing chip used (Figure 3a), but the reads produced on the MinION were consistent with the DNA preparation protocols employed (Figure 4a,b). This suggests that the MinION sequencing protocol

was more reproducible compared to PGM sequencing. This is important to note if this protocol is to be adjusted for instructional-laboratory purposes. PGM sequencing requires a number of skillsets not commonly mastered by undergraduate students, such as DNA size selection via E-Gel® SizeSelect™ Gels (Thermo Fisher, Waltham, MA), precise dilutions to a picogram scale, making quality assessments through use of qPCR, emulsion PCR, and more (Ion Torrent, 2014). MinION sequencing, on the other hand, requires little more than proper pipetting technique (Sequencing kit SQK-005, Oxford Nanopore, Oxford UK).

Sequencing of multi-drug resistance plasmids can be conducted on the PGM and MinION using the methods developed in this study. Despite the more difficult protocol, the PGM run produced over 4 Mbp of data. On the MinION — between both runs — over 11 Mbp of 2D data were generated.

The assembly generated in this study was validated by both depth-of-coverage and the percent identity to a reference sequence. Length-summary statistics on contig sequences, such as the N50 — a weighted median variable — are commonly employed to assess the quality of a *de novo* assembly (Earl *et al.*, 2011). Unfortunately, these metrics are dependent on an assembly that has generated several contigs of various lengths. In this study, our sequence reads were assembled into a single contig, forcing us to employ depth-of-coverage and %ID to a reference genome to assess the assembly generated. For the pEG1-1 genome, these metrics were sufficient to allow us to assess the quality of the generated assembly, but our approach was dependent on high depth-of-coverage and an extant, related reference genome. These limitations may mean that this method of assessing assembly quality is particularly ill-suited for metaplasmidome studies in which

a mixed sample of multiple plasmids are sequenced and assessed in a single run (Li *et al.*, 2015). Having multiple plasmids in a single sample may lower the average depth-of-coverage below 50X and if a novel plasmid is found within the metaplasmidome, we would not be able to validate the assembly. Instead, this approach is more appropriate for single-plasmid studies and may even benefit from the addition of plasmid incompatibility-group identification prior to sequencing so that reference sequences can be confirmed. However, if the goal is to isolate and assemble novel plasmids, alternative metrics to assess such an assembly will need to be established.

**Accessory regions of plasmid pEG 1-1.** The isolation of an IncP-1 $\beta$  plasmid from an agriculturally-impacted stream sediment sample was consistent with other studies: IncP-1 $\beta$  plasmids are typical in bacteria isolated from agricultural sites (Bahl, 2009) and their broad-host range makes exogenous capture into an *E. coli* recipient a logical outcome. However, the presence of the *inI04* complex integron within the pEG1-1 genome was surprising. The *inI04* integron is a critical component of the *Salmonella* genomic island (SGI1), a multi-drug resistance chromosomal element that is largely responsible for the global dissemination of multiply-antibiotic-resistant *Salmonella enterica* serovar Typhimurium, especially the widely-disseminated multiresistance clone dT104, after which it was named (Levings *et al.*, 2005). All of the resistance genes associated with SGI1 are contained within the *inI04* complex integron (Levings *et al.*, 2005). Most *inI04* complex integrons—including the one found on pEG101—carry the antibiotic resistance genes *aadA*, *sulI*, *floR*, *tetA/R*, and *blaP1* and confer pentaresistance to aminoglycosides, sulfonamides, florfenicol, tetracyclines and  $\beta$ -lactam antibiotics, but *inI04* variants, due to insertion and transposition events, can differ in genotype

(Seibor & Neuwirth, 2013). Unlike typical integrons, which are not mobile elements, the *in104* complex integron contains genes coding for transposase enzymes that allow it to be transposed into a host's chromosome or another plasmid (Levings *et al.*, 2005). An *in vitro* study of SGI1 demonstrated that the mobilization of SGI1 resistance depended specifically on the presence of IncA and IncC plasmids (Douard *et al.*, 2010). Douard *et al.*, explicitly demonstrated that IncP plasmids were unable to mobilize SGI1 resistance in repeated mating experiments (Douard *et al.*, 2010) yet our analysis of the pEG1-1 genome clearly indicates the presence of the *in104* integron.

This discovery suggests that *in104* has undergone recent recombination events that enable mobilization through a broader range of plasmids. The first *in104* integron described contained a *tnpA* transposase at the 3' end of the element (Boyd *et al.*, 2001), but the *tnpA* gene was absent on the *in104* integron of pEG1-1. Instead, *tniA* and *tniB* genes were found at the 3' end of the integron (Figure 8b). This exchange between the *tnpA* and *tniA* transposase enzymes may be what has allowed the newly acquired mobilization of *in104* on an IncP plasmid.

The *in104* complex integron also provides insight on what the original bacterial host of pEG1-1 may have been. The exogenous capture method used to isolate pEG1-1 was limited in that it provides no evidence on the origin of the plasmid (Gehr, 2013). However, since the *in104* complex integron has only been documented in *Salmonella enterica* (Boyd *et al.*, 2001) and *Proteus mirabilis*, this suggests that pEG1-1 may have originated in one of these two members of the Enterobacteriaceae (Seibor & Neuwirth, 2013).

Tn21 transposons are one of the most well-studied transposon families, yet the gene composition and order of genes within the tn21-related transposon of pEG1-1 was unlike any of the tn21 reference sequences available (Liebert *et al.*, 1999). The unique arrangement of this tn21-related transposon suggests it may be a derivative of the tn21 family that, similarly to the *in104* integron found on pEG1-1, has undergone a number of recombination events. For example, the presence of the *repA* gene suggest that this transposon may have interacted with the backbone of a different plasmid. Other IncP-1 $\beta$  plasmids use the replicase enzyme encoded by the *trfA* gene and do not contain the *repA* gene within their backbone (Popowska & Kraqczyk-Balska, 2013). The *repA* gene on pEG1-1 may have originated from the backbone of an IncU plasmid, another plasmid incompatibility group associated with the dissemination of antibiotic resistance that relies on the *repA* gene for replication (Cattoir *et al.*, 2008).

The *mazEF* gene was another distinguishing feature that is not characteristic of Tn21 transposons. This was a particularly interesting feature since, as described above, the *mazEF* toxin-antitoxin (TA) system is attributed to decreased susceptibility to  $\beta$ -lactam antibiotics (Schlüter *et al.*, 2015). Additionally, TA systems are involved in plasmid-addiction systems (PAS), a phenomenon in which successful bacteria propagation is influenced by the vertical transmission of a plasmid (Van Melderen & De Bast, 2009). If a bacterium carries a plasmid with a TA system, both the toxin and antitoxin are passed on to daughter cells. However, if the plasmid is not also replicated and passed onto the daughter cell, the unstable antitoxin degrades and the stable toxin persists, killing the daughter cell. Thus the vertical transfer of the plasmid with the complete TA system is selected for. The *mazEF* TA system found on pEG1-1 may induce

a plasmid-addiction system that acts as a major contributing factor to the persistence of multi-drug resistance plasmids in stream sediment.

The antibiotic resistance profile of pEG1-1 can be attributed to the antibiotic resistance genes within the AM1 and AM2 modules. Resistance to tetracycline was most likely conferred by the multiple *tetA/R* complexes, kanamycin & tobramycin resistance may be conferred by the *aada*, *strB*, or *aphA1* genes, and resistance to ticarcillin and piperacillin was most likely conferred by the *pse4* gene. The observed decreased susceptibility – though not reaching clinical levels of resistance as judged by MICs – to piperacillin/tazobactam, and cefepime conferred by pEG1-1 may be explained by the presence of the *mazEF* TA system. In *Staphylococcus aureus*, the *mazEF* TA system is upregulated under stress conditions and is associated with the formation of small colony variants and persisters that reside in a drug-tolerant state (Schuster *et al.*, 2015). Schuster *et al.* demonstrated that mutations of the *S. aureus mazEF* TA system led to increased  $\beta$ -lactam susceptibility (Schuster *et al.*, 2015) but to this author's knowledge, the *mazEF* TA has not been directly associated with clinical levels of  $\beta$ -lactam resistance, particularly in Gram-negative bacteria. The presence of genes associated with resistance to sulfonamide and florfenicol drugs suggest that pEG1-1 may also confer resistance to additional antibiotics that we have not yet assessed.

**Conclusion.** The pEG1-1 genome demonstrates how genes conferring heavy metal resistance, agricultural antibiotic resistance, and human antibiotic resistance can be linked. This may be concerning for public health since current efforts to mitigate the emergence of antibiotic resistance focus heavily on the regulation of human antibiotics in agriculture (The White House Office of the Press Secretary, 2015). However, even in the



absence of direct selection, clinical antibiotic resistance may be spreading as a consequence of coselection in response to three different factors: agricultural antibiotic use—such as tetracycline—heavy metals in the soil—such as mercury—and the plasmid addiction phenomenon perpetuated by the *mazE/F* TA system.

Additionally, the presence of the *in104* complex integron and the Tn21-related transposon suggests that multi-drug resistance plasmids are not only capable of transmission between mature bacteria within stream sediment, but are also actively recombining. Thus, the increased incidence and new combinations of antibiotic resistance genes observed in clinical isolates may be only the tip of the vast “iceberg” of resistance genes actually found in populations of bacteria – and particularly their plasmids – found in natural environments.

### Literature Cited

- Aarestrup, F.M. (1999).** Association between the consumption of antimicrobial agents in animal husbandry and the occurrence of resistant bacteria among food animals. *International Journal of Antimicrobial Agents* 12:279-85.
- Acinas, S., Sarma-Rupavtarm, R., Klepac-Ceraj, V., & Polz, M. (2005).** PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied Environmental Microbiology* 71:8966–9.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., & Lipman, D.J. (1990)** Basic local alignment search tool. *Journal of Molecular Biology* 215:403-10.
- Aminov, R.I. (2010).** A brief history of the antibiotic era: lessons learned and challenges for the future. *Frontiers in Microbiology* 1:134-41.
- Anderson, K.L., Whitlock, J.E., & Harwood, V.J. (2005).** Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. *Applied Environmental Microbiology* 71:3041-8.
- Andrews S. (2010).** FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed June 14, 2015.
- Bahl, M. I. (2009).** All IncP-1 plasmid subgroups, including the novel e subgroup, are prevalent in the influent of a Danish wastewater treatment plant. *Plasmid* 62:134–9.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A, Dvorkin, M., Kulikov, A.S., Lesin, V., Nikolenko, S., Pham, S., Prjibelski, A., Pyshkin, A., Sirotkin, A., Vyahhi, N., Tesler, G., Alekseyev, M., & Pevzner, P. (2012).** SPAdes: a new genome assembly algorithm and its application to single-cell sequencing. *Journal of Computational Biology* 19:455-77.
- Bennett, P.M. (2008).** Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *British Journal of Pharmacology* 153:S3470357.
- Bradley J. S., Garau, J., Lode, H., Rolston, K.V., Wilson, S.E., & Quinn, J.P. (1999).** Carbapenems in clinical practice: a guide to their use in serious infection. *International Journal of Antimicrobial Agents* 11:93–100.
- Boyd, D., Cloeckert, A., Chaslus-Dancla, E., & Mulvey, M.R. (2002).** Characterization of variant *Salmonella* genomic island 1 multidrug resistance regions from serovars Typhimurium DT104 and Agona. *Antimicrobial Agents and Chemotherapy* 6:1714-22.

- Brooks, J.M.** (2005). M.S. Thesis: Exogenous isolation and characterization of tetracycline resistant plasmid from poultry litter and litter amended soil. James Madison University.
- Bushnell, B.** (2014). BMAP short read aligner and other bioinformatics tools. Available at: <https://sourceforge.net/projects/bbmap/files/>. Accessed: February 21, 2016.
- Carattoli, A., Zankari, E., García-Fernández, A., Larsenc, M.V., Lundc, O., Villaa, L., Aarestrup, F.M., & Hasman, H.** (2014). *In Silico* detection and typing of plasmids using PlasmidFinder and plasmid Multilocus Sequence Typing. *Antimicrobial Agents and Chemotherapy* 58:3895-903.
- Carnerio, M.O., Russ, C., Ross, M.G., Gabriel, S.B., Nusbaum, C., & DePristo, M.A.** (2012). Pacific Biosciences sequencing technology for genotyping and variation discovery in human data. *BioMed Central* 13:375.
- Centers for Disease Control and Prevention, Office of Infectious Disease.** (2013). Antibiotic resistance threats in the United States, 2013. Available at: <http://www.cdc.gov/drugresistance/threat-report-2013>. Accessed March 09, 2016.
- Chee-Sanford, J. C., Aminov, R. I., Krapac, I. J., Garrigues-Jeanjean, N., & Mackie, R. I.** (2001). Occurrence and Diversity of tetracycline resistance genes in lagoons and groundwater underlying two swine production facilities. *Applied and Environmental Microbiology* 67:1494–1502.
- Chen, Y., Liu, T., Yu, C., Chiang, T., & Hwang, C.** (2013). Effects of GC bias in next-generation-sequencing data on *de novo* genome assembly. *PLoS ONE* 8:e62856.
- Claverys, J.P., Prudhomme, M., Mortier-Barriere, I., & Martin, B.** (2000). Adaptation to the environment: *Streptococcus pneumoniae*, a paradigm for recombination-mediated genetic plasticity? *Molecular Microbiology* 35:251-9.
- Darling, A., Mau, B., & Perna, N.T.** (2010). progressiveMauve: multiple genome alignment with gene gain, loss, and rearrangement. *PLoS One* 5:e11147.
- Davies, J. & Davies, D.** (2010). Origins and evolution of antibiotic resistance. *Microbial and Molecular Biology Reviews* 74:417–33.
- Delcher, A.L., Harmon, D., Kasif, S., White, O., & Salzberg, S.L.** (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Research* 27:4636-41.
- Desai, A., Marwah, V., Yadav, A., Jha, V., Dhaygude, K., Bangar, U., Kulkarni, V., & Jere, A.** (2013). Identification of optimum sequencing depth especially for *de novo* genome assembly of small genomes using next generation sequencing data. *PLoS One* 8:e60204.

- Dobrindt, U., Chowdary, M.G., Krumbholz, G., & Hacker, J.** (2010). Genome dynamics and its impact on evolution of *Escherichia coli*. *Medical Microbiology and Immunology* 199:145–54.
- Douard, G., Praud, K., Cloeckert, A., & Doublet, B.** (2010). The *Salmonella* genomic island 1 is specifically mobilized in *trans* by the IncA/C multidrug resistance plasmid family. *PLoS ONE* 5:e15302.
- Dröge, M., Pühler, A., & Selbitschka, W.** (2000). Phenotypic and molecular characterization of conjugative antibiotic resistance plasmids isolated from bacterial communities of activated sludge. *Molecular Genetics and Genomics* 263:471-82.
- Ewing, B. & Green, P.** (1998). Base-calling of automated sequencer traces using Phred II error probabilities. *Genome Research* 8:186–94.
- Figuroa-Bossi, N., Uzzau, S., Maloriol, D., & Bossi, L.** (2001). Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. *Molecular Microbiology* 39:260-71.
- Gehr, E.** (2013). M.S. Thesis: The potential for replication and transmission of antibiotic resistance plasmids in an *E. coli* population in agriculturally impacted stream sediment. James Madison University.
- Gordon, D.M., Bauer, S., & Johnson, J.R.** (2002). The genetic structure of *Escherichia coli* populations in primary and secondary habitats. *Microbiology* 148:1513-22.
- Gould, I.M. & Bal, A.M.** (2013). New antibiotic agents in the pipeline and how they can overcome microbial resistance. *Virulence* 4:185–91.
- Gross, M.** (2013). Antibiotics in crisis. *Current Biology* 23(24):R1063-5.
- Heather, J.M. & Chain, B.** (2016). The sequence of sequencers: the history of sequencing DNA. *Genomics* 107(1):1-8.
- Herrick, J.B., Haynes, R., Heringa, S., Brooks, J.M., & 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. *Journal of Applied Microbiology* 117:380-9.
- Heuer, H., Szczepanowski, R., Schneiker, S., Pühler, A., Top, E.M., & Schlüter, A.** (2004). The complete sequences of plasmids pB2 and pB3 provide evidence for a recent ancestor of the IncP-1beta group without any accessory genes. *Microbiology* 150:3591-9.
- Ion Torrent.** (2014). Ion Xpress™ plus gDNA fragment library preparation, user guide rev B.0. Publication Number MAN0009847.

- Ion Torrent.** (2015). Ion PGM™ Hi-Q™ Sequencing Kit, user guide rev. D.0. Publication Number MAN0009816.
- Jain, M., Fiddes, I., Miga, K.H., Olsen, H.E., Patern, B., & Akeson, M.** (2015). Improved data analysis for the MinION Nanopore sequencer. *Nature Methods* 12:351-6.
- Kelly, B.G., Vesperman, A., & Bolton, D.J.** (2009). Horizontal gene transfer of virulence determinants in selected bacterial foodborne pathogens. *Food and Chemical Toxicology* 47:969-77.
- Khachtourians, G. G.** (1998). Agricultural use of antibiotics and the evolution and transfer of antibiotic-resistant bacteria. *Canadian Medical Association Journal* 159:1129-36.
- Koren, S., Schatz, M.C., Walenx, B.P., Martin, J., Howard, J., Ganapathy, G., Wang, Z., Rasko, D.A., McCombie, W.R., Jarvis, E.D., & Phillippy, A.M.** (2012). Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nature Biotechnology* 30:693-700.
- Lee, K., Yong, D., Yum, J. H., Kim, H. H., & Chong, Y.** (2003). Diversity of TEM-52 extended-spectrum beta-lactamase-producing nontyphoidal *Salmonella* isolates in Korea. *The Journal of Antimicrobial Chemotherapy* 52:493–6.
- Levings, R.S., Lightfoot, D., Partridge, S.R., Hall, R.M., & Djordjevic, S.P.** (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. *Journal of Bacteriology* 187:4401-9.
- Li, B., Yang, Y., Ma, L., Ju, F., Guo, F., Tiedje, J.M., & Zhang, T.** (2015). Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *The International Society for Microbial Ecology Journal* 9: 2490-502.
- Liebert, C.A., Hall, R.M., & Summers, A.O.** (1999). Transposon Tn21, flagship of the floating genome. *Microbiology and Molecular Biology Reviews* 63:507-22.
- Loman, N., Quick, J., & Simpson, J.T.** (2015). A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nature Methods* 12:733-5.
- Lonardi, S., Mirebrahim, H., Wanamaker, S., Alpert, M., Ciardo, G., Duma, D., & Close, T.J.** (2015). When less is more: 'slicing' sequencing data improves read decoding accuracy and *de novo* assembly quality. *Bioinformatics* 31:2972-80.
- Mardis, E.** (2008). The impact of next-generation sequencing technology on genetics. *Trends in Genetics* 24:133–141.

**McArthur, A., Waglechner, N., Nizam, F., Yan, A., Azad, M., Baylay, A., Bhullar, K., Canova, M.J., De Pascale, G., Ejim, L., Kalan, L., King, A.M., Koteva, K., Morar, M., Mulvey, M., O'Brien, J., Pawlowski, A., Piddock, L., Spanogiannopoulos, P., Sutherland, A., Tang, I., Taylor, P., Thaker, M., Wang, W., Yan, T., & Wright, G.** (2013). The comprehensive antibiotic resistance database. *Antimicrobial Agents and Chemotherapy* 57:3348-57.

**Morey, M., Fernandez-Marmiesse, A., Castineiras, D., Fraga, J.M., Couce, M.L., & Cocho, J.A.** (2013). A glimpse into past, present, and future DNA sequencing. *Molecular Genetics and Metabolism* 110:2-34.

**Moura, A., Soares, M., Pereira, C., Leitão, N., Henriques, I., & Correia, A.** (2009) INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. *Bioinformatics* 25:1096–8.

**Naas, T., Bonnin, R.A., Cuzon, G., Villegas, M.V., & Nordmann, P.** (2013). Complete sequence of two KPC-harboring plasmids from *Pseudomonas aeruginosa*. *Journal of Antimicrobial Chemotherapy* 68:1757-62.

**Nealson, K.H.** (1997). Sediment bacteria: who's there, what are they doing, and what's new? *Annual Review of Earth and Planetary Sciences* 24:403-34.

**Norberg, P., Bergstrom, M., Jethava, V., Dubhashi, D., & Hermansson, M.** (2011). The IncP-1 plasmid backbone adapts to different host bacterial species and evolves through homologous recombination. *Nature Communications* 2:1–11.

**Ochman, H., Lawrence, J.G., & Groisman, E.A.** (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304.

**Pallen, M.J. & Wren, B.W.** (2007). Bacterial pathogenomics. *Nature* 449:835-42.

**Perreten, V., Schwarz, F., Cresta, L., Boeglin, M., Dasen, G., & Teuber, M.** (1997). Antibiotic resistance spread in food. *Nature* 389:801-2.

**Popowaska, M., & Krawczyk-Balska, A.** (2013). Broad-host-range IncP-1 plasmids and their resistance potential. *Frontiers in Microbiology* 4:44.

**Revilla, C., Garcillan-Barcia, M.P., Fernandez-Lopez, R., Thomson, N.R., Sanders, M., Cheung, M., Thomas, C.M., & De La Cruz, F.** (2008). Different pathways to acquiring resistance genes illustrated by the recent evolution of IncW plasmids. *Antimicrobial Agents and Chemotherapy* 52:1472-80.

**Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S. R., Connor, T.R., Bertoni, A., Swerdlow, H.P., & Gu, Y.** (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BioMed Central* 13:341.

- Sanger, F. & Coulson A.R.** (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* 94:441–8.
- Sanger, F., Nicklen, S., & Coulson, A.R.** (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* 74:5463–7.
- Schuster, C.F., Mechler, L., Nolle, N., Krismer, B., Zelder, M., Götz, F., & Bertram, R.** (2015). The *mazEF* toxin-antitoxin system alters the  $\beta$ -lactam susceptibility of *Staphylococcus aureus*. *PLoS ONE* 10:e0126118.
- Schlüter, A., Heuer, H., Szczepanowski, R., Schneiker, S., Pühler, A., & Top, E.** (2005). Plasmid pB8 is closely related to the prototype IncP-1beta plasmid R751 but transfers poorly to *Escherichia coli* and carries a new transposon encoding a small multidrug resistance efflux protein. *Plasmid* 54:135-48.
- Schlüter, A., Szczepanowski, R., Pühler, A., & Top, E.M.** (2007). Genomics of IncP-1 antibiotic resistance plasmids isolated from wastewater treatment plants provides evidence for a widely accessible drug resistance gene pool. *Microbiology Reviews* :449-77.
- Seemann, T.** (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068-9.
- Shintani, M., Takahashi, Y., Yamane, H., & Nojiri, H.** (2010a). The behavior and significance of degradative plasmids belonging to Inc groups in *Pseudomonas* within natural environments and microcosms. *Microbes and Environments* 25:253–65.
- Shintani, M., Yamane, H., & Nojiri, H.** (2010b). Behaviour of various hosts of the IncP-7 carbazole-degradative plasmid pCAR1 in artificial microcosms. *Bioscience, Biotechnology and Biochemistry* 74:343–9.
- Seibor, E. & Neuwirth, C.** (2013). Emergence of salmonella genomic island 1 (SGI1) among *proteus mirabilis* clinical isolates in Dijon, France. *The Journal of Antimicrobial Chemotherapy* 68:1750-6.
- Smith, D. L., Harris, A. D., Johnson, J. A., Silergeld, E. K., & Morris, J. G. Jr.** (2002). Animal antibiotic use has an early but important impact on the emergence of antibiotic resistance in human commensal bacteria. *Proceedings of the National Academy of Science* 99:6434-9.
- Stephenson, G.R., & Rychert, R.C.** (1982). Bottom sediment: a reservoir of *Escherichia coli* in rangeland streams. *Journal of Range Management* 35, 119-23.

**Tauch, A. Schlüter, A., Bischoff, N., Goesmann, A., Meyer, F., & Pühler, A.** (2003). The 79,370-bp conjugative plasmid pB4 consists of an IncP-1 $\beta$  backbone loaded with a chromate resistance transposon, the *strA-strB* streptomycin resistance gene pair, the oxacillinase gene *bla<sub>NPS-1</sub>*, and a tripartite antibiotic efflux system of the resistance-nodulation-division family. *Molecular Genetics and Genomics* 268:570–84.

**The White House Office of the Press Secretary.** (2015). National Action Plan to Combat Antibiotic-Resistant Bacteria.. Available at: [https://www.whitehouse.gov/sites/default/files/docs/national\\_action\\_plan\\_for\\_combating\\_antibiotic-resistant\\_bacteria.pdf](https://www.whitehouse.gov/sites/default/files/docs/national_action_plan_for_combating_antibiotic-resistant_bacteria.pdf). Accessed March 13, 2016.

**United States Food and Drug Administration, Department of Health and Human Services.** (2012). Summary report on antimicrobials sold or distributed for use in food-producing animals. Available at: <http://www.fda.gov/downloads/ForIndustry/UserFees/AnimalDrugUserFeeActADUFA/UCM440584.pdf>. Accessed March 09, 2016.

**Upshur, R.** (2008). Ethics and infectious disease. *Bulletin of the World Health Organization* 86:8.

**Van Meervenne, E., Van Coillie, E., Kerckhof, F.-M., Devlieghere, F., Herman, L., & De Gelder L. S. P.** (2012). Strain-specific transfer of antibiotic resistance from an environmental plasmid to foodborne pathogens. *Journal of Biomedicine & Biotechnology* 2012:834598.

**Van Melderren, L. & Saavedra De Bast, M.** (2009). Bacterial toxin-antitoxin systems: more than selfish entities? *PLoS Genetics* (3):e1000437.

**Ventola, C.L.** (2015). The antibiotic resistance crisis part 1: causes and threats. *Pharmacy and Therapeutics* 40(4):277-83.

**Virginia Department of Environmental Quality.** (2014). Draft 2014 305(b)/303(d) Water Quality Assessment Integrated Report. Available at: [http://www.deq.virginia.gov/Programs/Water/WaterQualityInformationTMDLs/WaterQualityAssessments/2014305\(b\)303\(d\)IntegratedReport.aspx](http://www.deq.virginia.gov/Programs/Water/WaterQualityInformationTMDLs/WaterQualityAssessments/2014305(b)303(d)IntegratedReport.aspx). Accessed March 14, 2016.

**Whiteford, N., Haslam, N., Weber, G., Prügel-Bennett A., Essex J.W., Roach P.L., Bradley M., & Neylon C.** (2005). An analysis of the feasibility of short read sequencing. *Nucleic Acids Research* 33:e171.

**World Health Organization.** (2014). Antimicrobial resistance: global report on surveillance. Available at: [http://apps.who.int/iris/bitstream/10665/112642/1/9789241564748\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/112642/1/9789241564748_eng.pdf). Accessed March 15, 2016.



**Zur Wiesch, P.A., Kouyos, R., Engelstadter, J., Regoes, R.R., & Bonhoeffer, S.** (2011). Population biological principles of drug-resistance evolution in infectious diseases. *The Lancet Infectious Diseases* 11:236-47.