

1-15-2018

Practical implications of erythromycin resistance gene diversity on surveillance and monitoring of resistance

Elizabeth L. Rieke

Iowa State University, luby@iastate.edu

Jinlyung Choi

Iowa State University, jinchoi@iastate.edu

Thomas B. Moorman

United States Department of Agriculture


Michelle L. Soupir

Iowa State University, msoupir@iastate.edu

Heather K. Allen

Follow this and additional works at: https://lib.dr.iastate.edu/abe_eng_pubs

United States Department of Agriculture

 Part of the [Agriculture Commons](#), [Bioresource and Agricultural Engineering Commons](#), and the [Genetics Commons](#)

See next page for additional authors

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/abe_eng_pubs/862. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

Practical implications of erythromycin resistance gene diversity on surveillance and monitoring of resistance

Abstract

Use of antibiotics in human and animal medicine has applied selective pressure for the global dissemination of antibiotic-resistant bacteria. Therefore, it is of interest to develop strategies to mitigate the continued amplification and transmission of resistance genes in environmental reservoirs such as farms, hospitals, and watersheds. However, the efficacy of mitigation strategies is difficult to evaluate because it is unclear which resistance genes are important to monitor, and which primers to use to detect those genes. Here we evaluated the diversity of one type of macrolide antibiotic resistance gene (*erm*) in one type of environment (manure) to determine which primers would be most informative to use in a mitigation study of that environment. We analyzed all known *erm* genes and assessed the ability of previously published *erm* primers to detect the diversity. The results showed that all known *erm* resistance genes group into 66 clusters, and 25 of these clusters (40%) can be targeted with primers found in the literature. These primers can target 74–85% of the *erm* gene diversity in the manures analyzed.

Keywords

antibiotic, resistance, target, environment, surveillance, genes

Disciplines

Agriculture | Bioresource and Agricultural Engineering | Genetics

Comments

This is a manuscript of an article published as Choi, Jinlyung, Elizabeth L. Rieke, Thomas B. Moorman, Michelle L. Soupir, Heather K. Allen, Schuyler D. Smith, and Adina Howe. "Practical implications of erythromycin resistance gene diversity on surveillance and monitoring of resistance." *FEMS Microbiology Ecology* (2018). doi: [10.1093/femsec/fy006](https://doi.org/10.1093/femsec/fy006). Posted with permission.

Creative Commons License



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

Authors

Elizabeth L. Rieke, Jinlyung Choi, Thomas B. Moorman, Michelle L. Soupir, Heather K. Allen, Schuyler D. Smith, and Adina Howe

Practical implications of erythromycin resistance gene diversity on surveillance and monitoring of resistance

Authors:

Jinlyung Choi^{1*}, Elizabeth L. Rieke^{1*}, Thomas B. Moorman², Michelle L. Soupir¹, Heather K. Allen³, Schuyler D. Smith², Adina Howe^{1†}

¹Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA, USA

²National Laboratory for Agriculture and the Environment, USDA-ARS, Ames, IA, USA

³Food Safety and Enteric Pathogens Research Unit, National Animal Disease Center, USDA-ARS, Ames, IA, USA

⁴Department of Bioinformatics and Computational Biology, Iowa State University, Ames, IA USA

Keywords: antibiotic, resistance, target, environment, surveillance, genes

**These authors contributed equally*

†corresponding author: Adina Howe, adina@iastate.edu

Abstract

Use of antibiotics in human and animal medicine has applied selective pressure for the global dissemination of antibiotic-resistant bacteria. Therefore, it is of interest to develop strategies to mitigate the continued amplification and transmission of resistance genes in environmental reservoirs such as

farms, hospitals, and watersheds. However, the efficacy of mitigation strategies is difficult to evaluate because it is unclear which resistance genes are important to monitor, and which primers to use to detect those genes. Here we evaluated the diversity of one type of macrolide antibiotic resistance gene (*erm*) in one type of environment (manure) to determine which primers would be most informative to use in a mitigation study of that environment. We analyzed all known *erm* genes and assessed the ability of previously published *erm* primers to detect the diversity. The results showed that all known *erm* resistance genes group into 66 clusters, and 25 of these clusters (40%) can be targeted with primers found in the literature. These primers can target 74-85% of the *erm* gene diversity in the manures analyzed.

Introduction

Antibiotic resistance is a global challenge, with increasing resistance to antibiotics threatening our ability to treat both human and animal diseases (WHO | Global action plan on antimicrobial resistance 2017). Antibiotic use in human medicine and animal agriculture has increased environmental reservoirs of antibiotic resistance genes, which in turn has increased the risk of transmission of antibiotic-resistant bacteria to both humans and animals (McEwen and Fedorka-Cray 2002; Vaz-Moreira *et al.* 2014). This linkage has resulted in the prioritization of understanding how resistance moves from environmental sources to clinical pathogens and the associated influence of human activity. To understand the movement of antibiotic resistance in the environment, we need accessible tools that can provide large-scale surveillance of resistance in diverse environmental samples.

Molecular microbiology advances have allowed us to leverage amplification and subsequent sequencing of DNA that encodes for antibiotic resistance genes, resulting in our awareness of an incredibly diverse global reservoir of environmental “resistomes”. Generally, metagenomic shotgun sequencing is a costly tool for antibiotic gene surveillance as it provides information on *all* genes in an environmental sample. Among these genes, only a fraction (0.01-1%) are related to antibiotic resistance,

resulting in a significant majority of sequences from metagenomes not readily usable for resistance detection (Shi *et al.* 2013; Li *et al.* 2015). A promising alternative to metagenomic sequencing is high-throughput amplicon qPCR assays, such as the Wafergen Smartchip which has been previously used for several resistance surveillance studies (Shi *et al.* 2013; Wang *et al.* 2014; Karkman *et al.* 2016; Muziasari *et al.* 2016; Stedtfeld *et al.* 2017). Unlike the broad scope of metagenomic sequencing, high-throughput qPCR assays target a suite of genes using primers and can quantify hundreds of targeted resistance genes and multiple samples simultaneously (e.g., one Wafergen Smartchip contains 5,184 assays). Consequently, the price per gene or sample of these assays for resistance gene detection is orders of magnitude less than metagenomic sequencing, making it more conducive to large-scale surveillance. A significant limitation of this technology is the need to develop primer-based assays for each targeted gene of interest that are effective for high-throughput amplification conditions.

We are increasingly aware that certain genes may be more related to the risks of the emergence or persistence of resistance than others. For example, integrons and sulfonamide resistance genes have been used to detect anthropogenic contaminants (Wang *et al.* 2014; Gillings *et al.* 2015). Further, specific environments (mammalian gut, manure, wastewater, etc.) have been observed to be enriched in antibiotic resistance genes relative to soil or water environments (Chee-Sanford *et al.* 2009; Koike *et al.* 2010; Garder, Moorman and Soupir 2014; Joy *et al.* 2014; Luby, Moorman and Soupir 2016), suggesting that these environments are potential reservoirs of resistance genes. Among the hundreds of genes associated with antibiotic resistance that are observed in environmental metagenomes, selecting the key targets relevant to the spread of resistance is a significant and important opportunity. In this study, we demonstrate how we have chosen specific genes that are the most effective among previously targeted genes to serve as indicators for antibiotic resistance and to understand resistance hotspots and transmission. This framework, while developed for agriculturally-impacted environments, can be broadly applied to the selection of genes from varying resistance gene classes and environments. Specifically, this effort focuses on understanding the diversity of erythromycin ribosomal methylase (*erm*) gene and

the most relevant gene targets for understanding the spread of *erm*-associated resistance from manure sources to the environment.

Erm genes encode resistance to macrolide antibiotics, which have long been used to treat Gram-positive and certain Gram-negative pathogens infecting humans, swine, and cattle (Roberts *et al.* 2008; Pyörälä *et al.* 2014). Broadly, macrolide antibiotics act by binding to the 23S subunit of the bacterial ribosome, causing premature release of peptides during translation. The *erm* genes cause resistance by methylating rRNA at the active site, reducing the ability of macrolide antibiotics to bind to the ribosome (Weisblum 1998; Vester and Douthwaite 2001). *Erm*-mediated resistance to macrolides has also been observed to confer resistance against other antibiotics, including lincosamide and streptogramin B (MLS_B resistance) (Leclercq and Courvalin 1991). The widespread use of macrolides and their relevance for both animal and human health has resulted in a research emphasis on *erm* genes and their bacterial hosts as key targets for understanding the development of resistance and its spread in agricultural environments. Previously, *erm* genes have been detected in various agricultural settings, including swine manure, lagoon water, soils, surface and subsurface drainage from fields, and groundwater surrounding and underlying animal production facilities (Chen *et al.* 2007; Knapp *et al.* 2010; Koike *et al.* 2010; Joy *et al.* 2013, 2014; Whitehead and Cotta 2013; Fahrenfeld *et al.* 2014; Garder, Moorman and Soupir 2014; Soni *et al.* 2015; Luby, Moorman and Soupir 2016).

Most of our previous knowledge of *erm* genes and their associated amplicon targets have stemmed from the characterization and sequencing of bacterial isolates and their phenotypic resistance to MLS_B antibiotics (Pyörälä *et al.* 2014). A total of 21 unique classes of *erm* genes have been identified based on sequence homology to protein-coding *erm* sequences from cultured bacteria (Roberts *et al.* 1999). More recently, metagenomic analyses of DNA from the total microbial community in environmental samples has expanded what is known about *erm* diversity beyond these 21 classes, showing that the *erm* class of genes is comprised of numerous sequence variants from diverse bacterial hosts (Fang *et al.* 2015; Li *et al.* 2015). These sequence variants are present in a range of abundances depending on their environment of

origin. The focus of this study was to better understand the diversity of *erm* genes and to target the gene variants that could be indicative of resistance originating from manure and spreading to agricultural soil and water environments.

Materials and Methods

Phylogenetic analysis of erm genes

Gene sequencing sharing high similarity to *ermA*, *ermB*, *ermC*, and *ermF* were obtained from publicly available databases. The Ribosomal Database Project Fungene Repository (Fish *et al.* 2013) was used to obtain *ermB* and *ermC*-associated sequences. It was required that sequences share 97% amino acid sequence coverage to established HMM protein models for Fungene gene families “Resfam_ermA”, “Resfam_ermB” and “Resfam_ermC” (Version 8.8). Additionally, *ermF* gene nucleotide sequences were obtained from proteins listed in the ARDB-Antibiotic Resistance Genes Database (version 1.1, July 3, 2009) (Liu and Pop 2009) and associated with the annotation “*ermF*”. All *erm*-associated sequences were combined and clustered at 99% nucleotide similarity using CD-HIT (v4.6.1c) (Li and Godzik 2006; Fu *et al.* 2012), resulting in 66 unique clusters. One representative sequence for each cluster was identified by CD-HIT and were aligned using Muscle (v3.8.31) (Edgar 2004) with the following parameters: gap open -400, gap extend 0, clustering method UPGMB. A maximum-likelihood phylogenetic tree was constructed from this alignment using FastTree (v2.1.8) (Price, Dehal and Arkin) with default parameters. Taxonomy was identified based on annotations in the NCBI non-redundant nucleotide database (NCBI Resource Coordinators 2017).

To consider an *erm* gene sequence to be associated with a previously targeted PCR primer sequence, both forward and reverse primers were required to share 100% nucleotide similarity over a minimum of 17 bp of the primer length.

Manure metagenomic datasets

The presence of *erm* genes was characterized in swine and cattle manures. For swine manure, DNA was extracted from two biological replicates (three technical replicates each) of swine manure originating from Iowa State's Northeast Research and Demonstration Farm, near Nashua, IA (43.0° N, 92.5° W). Metagenomic libraries were prepared and sequenced at Iowa State University DNA Sequencing Facility on a HiSeq 2500 instrument (Illumina, San Diego, CA) according to manufacturer's instructions. These datasets are deposited in the NCBI SRA as project SRP109083 (Supplemental Table 1). Sequences were compared to representatives of *erm* genes described above (BLAST, v2.4.0+) (Camacho *et al.* 2009). Sequences were annotated as *erm* genes if they matched the representative sequence within a cluster with a minimum e-value of 1e-5 and if both paired-end reads matched the same representative target. The abundance of *erm* sequences in each sample was calculated as the total number of reads meeting these criteria.

Cattle manure metagenomes were obtained from a previously published study of antibiotic resistant genes in commercial cattle as they moved through the process of beef production from feedlot entry to slaughter (Noyes *et al.* 2016). The presence of *erm* sequences in these samples was determined by the total number of reads that shared sequence homology (BLAST, v2.4.0+, evalue 1e-5) to the best matched *erm* representative sequence. Similarly, metagenomes from human-impacted (Fitzpatrick and Walsh 2016) and pristine environment (Staley *et al.* 2013) were aligned against *erm* sequences and considered a match if alignment scores resulted in e-value scores of at least 1e-5.

Results

A total of 5,648 *erm* DNA sequences were identified from annotated genes based on sequence similarity to well-characterized *erm* genes and were clustered at 99% nucleotide similarity to identify 66 unique *erm* variant clusters. A representative sequence of each cluster was defined as the longest consensus sequence in each cluster as determined by a greedy incremental clustering algorithm (see

Methods, Table 1). These representative sequences were aligned and used to construct a phylogenetic tree describing the diversity of *erm* genes (Figure 1). Based on sequence homology, the resulting *erm* gene clusters encompass the majority of *erm* genes studied in previous literature: *ermA*, *ermB*, *ermC*, *ermF*, *ermG*, and *ermT* (reviewed in (Roberts *et al.* 2008)). Among the gene clusters, a cluster associated with *ermA* was the most represented in our *erm* gene database (Cluster 15, 3,542 genes), followed by an *ermB* cluster (Cluster 18, 1,387 genes), and then an *ermC* cluster (Cluster 30, 399 genes). These three gene clusters comprise 94% of *erm* genes and are evidence to biases in the previous characterization of *erm* genes towards specific gene variants. Beyond the three most abundant gene clusters, the next most represented cluster (Cluster 11, 50 genes) is not well-characterized (e.g., most similar to unannotated *erm* gene clusters in our database) and is most closely related to genes belonging to *Streptococcus agalactiae* strain TR7 in NCBI GenBank (100% nucleotide identity). Most clusters (53 of 66) are associated with five or less gene sequences, demonstrating that much of what we know of specific *erm* gene families is based on very few characterized representatives.

Next, we evaluated the diversity of bacteria carrying these *erm* genes by identifying the taxonomic origin of potential bacterial hosts associated with each *erm* gene sequence (Table 1, Supp Fig 1). In general, the majority of known *erm* gene sequences were associated with *Firmicutes* (98%), followed by *Proteobacteria* (0.6%) and *Bacteroidetes* (0.6%). While *ermF* and *ermG* genes were observed to be carried by only *Bacteroidetes*, *ermA*, *ermB*, *ermC*, and *ermT* genes were associated primarily with *Firmicutes* (Supp Fig 1). Within the *Firmicutes*, *ermB* genes were associated mainly with the order *Lactobacillales*, while *ermA* and *ermT* genes were associated with members of the *Bacillales* order (Supp Fig 2). These results demonstrate a wide range of potential host diversity for *erm* genes and highlights the impact of the choice of primer gene targets selecting for or against specific host bacteria.

Historically, *erm* genes have been extensively targeted for qPCR quantification of gene abundances in the environment (Table 2), and we evaluated the ability of previously published PCR primers to detect the *erm* gene diversity described above by computationally hybridizing the primer

sequences from the literature with the representative *erm* gene sequences in our database. Overall, published primer pairs were 100% similar to 25 of the representative sequences of *erm* clusters (Figure 1). Generally, well-characterized gene clusters (e.g., containing the most known gene sequences) were observed to be associated with previous primer development. Several clusters were not associated with previously published primer targets, very likely due to the few well-characterized *erm* sequences within these clusters. Previously, observed diversity in natural samples have weak correlations with well-characterized genes (Choi *et al.* 2017), suggesting that primer targets selected based on the most well-studied genes may not be effective in environmental samples.

We next evaluated the diversity of *erm* genes in 12,947 environmental metagenomes (Supplemental Table 1), resulting in the observation that significantly more *erm* genes are present in human-impacted environments (feces and animal-associated soil and water) than in natural environments (Figure 2). We also searched an additional 39 metagenomes originating from relatively pristine freshwaters along the Upper Mississippi River (Staley *et al.* 2013, Supplemental Table 1), resulting in only 3 reads out of 716 million, sharing similarity (evalue < 1e-5) to *erm* genes. Combined, these results demonstrate that *erm* genes are rare in environments with minimal human impact and suggest that *erm* genes associated with feces or manure are ideal for tracking the spread and persistence of resistance through the environment. These results are also consistent with previous observations that manure contains abundant genes related to *erm* resistance and is a source of these genes into the environment (e.g., soil and water) (Chee-Sanford *et al.* 2009; Koike *et al.* 2010; Heuer, Schmitt and Smalla 2011; Joy *et al.* 2013; Luby, Moorman and Soupir 2016).

Consequently, we next identified *erm* genes in manure metagenomes. We aligned *erm* gene sequences against metagenomes derived from two large manure metagenomic studies (requiring $n_{\text{manure}} > 3$): swine manure collected near Nashua, IA (Luby, Moorman and Soupir 2016) and cattle manure from a previously published study (Noyes *et al.* 2016). These manure metagenomes were strategically selected based on the number of biological replicates and sequencing depth. Three *erm* clusters comprised 46%

and 45% of the total abundance of *erm* genes in swine and cattle manure, respectively (Supp. Table 1). The genes associated with these most abundant clusters differed between swine and cattle manures. In swine metagenomes, sequences associated with the *ermB* gene cluster (Figure 1, sharing 93-99% similarity) captured 26% of all *erm* sequences, followed by *ermG*-associated sequences capturing 11% and *ermA*-associated sequences capturing 9%. In cattle metagenomes, sequences associated with *ermF* represented 23.5% of all *erm* abundances, followed by sequences associated with *ermG* capturing 12.4% and sequences associated with *ermB* capturing 9%.

Only a subset of *erm* genes detected in manure are targeted by existing primer sets. Overall, a total of 25 out of the 66 *erm* clusters (40%) could be computationally detected with known primers (Supp. Table 1), and these genes also encompass much of the total *erm* abundances observed in manure metagenomes. Collectively, if all primers were used, 74% and 85% of the total *erm* gene sequence diversity observed in swine and cattle metagenomes, respectively, could be detected, suggesting good coverage of these genes for PCR or qPCR assays. Specifically, in swine manure metagenomes, *ermB* primers could detect 29% of *erm* sequences, followed by *ermF* primers capturing 14% and *ermG* primers capturing 12% (Figure 3). In cattle, *ermF* primers are the most effective, capturing 30% of *erm* sequences, followed by 21% with *ermG* primers, and 15% with *ermB* primers. Consequently, depending on the environmental sample in a study, in this case swine versus cattle manure, the choice of *erm* gene targets can significantly alter *erm* abundance estimations. For example, in swine manure, two times more *erm* gene abundance would be estimated if *ermB* primers were used instead of *ermF* primers. Even within the same gene clade, different primers could result in significant differences in abundance estimations, and this result is observed especially for *ermC* primers where a near two-fold difference in abundance estimations would result based on selection of primers from Patterson et al (Patterson *et al.* 2007) vs. Jensen et al (Jensen, Frimodt-Moller and Aarestrup 1999). The selection of Patterson primers would result in the detection of genes from up to seven *erm* gene clusters over the two to three gene cluster detected with Jensen or Koike et al (Koike *et al.* 2010) primers. Similar results are noted in the

cattle manure, where *ermC* primers designed by Patterson capture 13% of the total abundance of *erm* sequences in the metagenomes, while Koike and Jensen primers only capture 4.4 and 2.1%, respectively. These results emphasize that the targeting of a specific *erm* gene, even within closely related gene variants, can significantly alter estimations of associated resistance in manures.

Thus, overall, for swine manure, the most effective gene target based on abundance in swine metagenomes (26% of *erm* genes) originates from an *ermB* cluster (Cluster 25) and is associated with *Clostridium perfringens*. The next most abundant *ermB* cluster in swine (Cluster 19, most similar to a gene in *Streptococcus pneumoniae* CGSP14) represented only 2% of *erm* abundances. These results indicate that while *ermB* primers can target multiple strains (Supp. Figure 1 and 2), in these swine metagenomes, it is one gene cluster that specifically dominates. This gene cluster is also abundant in cattle manure metagenomes, though comprising less of total *erm* gene abundance (9%). Within our *erm* gene database, this particular sequence cluster is represented by a single gene representative and shares 100% similarity to experimental *Clostridium acetobutylicum* strains in the NCBI non-redundant gene database (mutant HQ683763.1 and clone HQ25744.1). The overall lack of similar homologous genes in NCBI nr suggest that this specific *ermB* gene is abundant in manures but is a gene for which we have few sequenced representatives. We identified this gene during our exploration of the effectiveness of current primers on manure metagenomes, and our observations suggest that this gene would benefit from further study given its prevalence.

Discussion

Over the past twenty years, an abundance of literature has been published quantifying macrolide resistance in agricultural landscapes using qPCR approaches. However, these previous studies often use primers for *erm* genes designed in only a handful of publications (Table 2). Our study found that current published primer sets, used on their own, are effective at capturing only a subset of the *erm* diversity in manure samples. For example, if only one primer set were used, less than one-third of *erm* genes would

be detected. To increase our ability to detect *erm* genes in agricultural systems, we identified the most abundant *erm* clusters in both swine and cattle manures, identifying the best gene targets for future studies. These genes and their associated primers are recommended for high throughput qPCR assays that can scale the detection and quantification of these genes for antibiotic gene surveillance.

In all amplicon assays, quantifying environmental abundances of gene targets is limited by the effectiveness of primer design. The results presented here emphasize that estimates of abundances of a gene of interest cannot simply be based on primers to genes that have previously been successfully detected. Rather, genes appropriate for antibiotic gene surveillance should be indicative of the spread of resistance (e.g., originate from manure but lacking from pristine environments), representative of diverse hosts (especially those with clinical risks), and accurately represent gene abundances in environmental samples. Our specific effort targeted the *erm* gene and evaluating the effectiveness of previously published primers sets. The increasing availability of metagenomes makes these evaluations possible, as demonstrated in this study. Although metagenomic sequencing advances will continue to provide powerful tools to understand the broad diversity of resistance in environments, metagenomes are limited by both detection rate and resolution. Short read lengths, the difficulty of assembling many resistance genes (because of their common association with mobile elements containing repeated sequences), and their presence in multiple bacterial hosts challenges the detection of resistance genes using metagenomics. Going forward, high throughput amplicon assays with strategic gene targets and primer designs are a complementary alternative to help fill these gaps and help us understand the movement of resistance genes among complex environments.

Funding:

This project was supported by the AFRI food safety from the USDA National Institute of Food and Agriculture [grant no. 2016-68003-24604] and the National Pork Board [contract 13-051].

References

- Camacho C, Coulouris G, Avagyan V *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* 2009;**10**:421.
- Chee-Sanford JC, Mackie RI, Koike S *et al.* Fate and transport of antibiotic residues and antibiotic resistance genes following land application of manure waste. *J Environ Qual* 2009;**38**:1086.
- Chen J, Yu Z, Michel FC *et al.* Development and application of real-time PCR assays for quantification of erm genes conferring resistance to Macrolides-Lincosamides-Streptogramin B in livestock manure and manure management systems. *Appl Environ Microbiol* 2007;**73**:4407–16.
- Choi J, Yang F, Stepanauskas R *et al.* Strategies to improve reference databases for soil microbiomes. *ISME J* 2017;**11**:829–34.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;**32**:1792–7.
- Fahrenfeld N, Knowlton K, Krometis LA *et al.* Effect of manure application on abundance of antibiotic resistance genes and their attenuation rates in soil: Field-scale mass balance approach. *Environ Sci Technol* 2014;**48**:2643–50.
- Fang H, Wang H, Cai L *et al.* Prevalence of antibiotic resistance genes and bacterial pathogens in long-term manured greenhouse soils as revealed by metagenomic survey. *Environ Sci Technol* 2015;**49**:1095–104.
- Fish JA, Chai B, Wang Q *et al.* FunGene: the functional gene pipeline and repository. 2013, DOI: 10.3389/fmicb.2013.00291.
- Fitzpatrick D, Walsh F. Antibiotic resistance genes across a wide variety of metagenomes. Smalla K (ed.). *FEMS Microbiol Ecol* 2016;**92**:fiv168.
- Fu L, Niu B, Zhu Z *et al.* CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012;**28**:3150–2.
- Garder JL, Moorman TB, Soupier ML. Transport and persistence of tylosin-resistant enterococci, genes, and tylosin in soil and drainage water from fields receiving swine manure. *J Environ Qual* 2014;**43**:1484.
- Gillings MR, Gaze WH, Pruden A *et al.* Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME J* 2015;**9**:1269–79.
- Heuer H, Schmitt H, Smalla K. Antibiotic resistance gene spread due to manure application on agricultural fields. *Curr Opin Microbiol* 2011;**14**:236–43.
- Jensen LB, Frimodt-Moller N, Aarestrup FM. Presence of erm gene classes in Gram-positive bacteria of animal and human origin in Denmark. *FEMS Microbiol Lett* 1999;**170**:151–8.

- Joy SR, Bartelt-Hunt SL, Snow DD *et al.* Fate and transport of antimicrobials and antimicrobial resistance genes in soil and runoff following land application of swine manure slurry. *Environ Sci Technol* 2013;**47**:12081–8.
- Joy SR, Li X, Snow DD *et al.* Fate of antimicrobials and antimicrobial resistance genes in simulated swine manure storage. *Sci Total Environ* 2014;**481**:69–74.
- Karkman A, Johnson TA, Lyra C *et al.* High-throughput quantification of antibiotic resistance genes from an urban wastewater treatment plant. *FEMS Microbiol Ecol* 2016;**92**:1–7.
- Knapp CW, Doling J, Ehlert PAI *et al.* Evidence of increasing antibiotic resistance gene abundances in archived soils since 1940. *Environ Sci Technol* 2010;**44**:580–7.
- Koike S, Aminov RI, Yannarell AC *et al.* Molecular ecology Of macrolide–lincosamide–streptogramin B methylases in waste lagoons and subsurface waters associated with swine production. *Microb Ecol* 2010;**59**:487–98.
- Leclercq R, Courvalin P. Bacterial resistance to macrolide, lincosamide, and streptogramin antibiotics by target modification. *Antimicrob Agents Chemother* 1991;**35**:1267–72.
- Li B, Yang Y, Ma L *et al.* Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *ISME J* 2015;**9**:2490–502.
- Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006;**22**:1658–9.
- Liu B, Pop M. ARDB--Antibiotic resistance genes database. *Nucleic Acids Res* 2009;**37**:D443–7.
- Luby EM, Moorman TB, Soupier ML. Fate and transport of tylosin-resistant bacteria and macrolide resistance genes in artificially drained agricultural fields receiving swine manure. *Sci Total Environ* 2016;**550**:1126–33.
- McEwen SA, Fedorka-Cray PJ. Antimicrobial use and resistance in animals. *Clin Infect Dis* 2002;**34**:S93–106.
- Muziasari WI, Pärnänen K, Johnson TA *et al.* Aquaculture changes the profile of antibiotic resistance and mobile genetic element associated genes in Baltic Sea sediments. Smalla K (ed.). *FEMS Microbiol Ecol* 2016;**92**:fiw052.
- NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2017;**45**:D12–7.
- Noyes N, Yang X, Linke L *et al.* Resistome diversity in cattle and the environment decreases during beef production. *Elife* 2016.
- Patterson AJ, Colangeli R, Spigaglia P *et al.* Distribution of specific tetracycline and erythromycin resistance genes in environmental samples assessed by macroarray detection. *Environ Microbiol* 2007;**9**:703–15.

- Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately maximum-likelihood trees for large alignments.
- Pyörälä S, Baptiste KE, Catry B *et al.* Macrolides and lincosamides in cattle and pigs: Use and development of antimicrobial resistance. *Vet J* 2014;**200**:230–9.
- Roberts MC, Dueck M, Hoban D *et al.* Update on macrolide-lincosamide-streptogramin, ketolide, and oxazolidinone resistance genes. *FEMS Microbiol Lett* 2008;**282**:147–59.
- Roberts MC, Sutcliffe J, Courvalin P *et al.* Nomenclature for macrolide and macrolide-lincosamide-streptogramin B resistance determinants. *Antimicrob Agents Chemother* 1999;**43**:2823–30.
- Shi P, Jia S, Zhang X-X *et al.* Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. *Water Res* 2013;**47**:111–20.
- Soni B, Bartelt-Hunt SL, Snow DD *et al.* Narrow grass hedges reduce tylosin and associated antimicrobial resistance genes in agricultural runoff. *J Environ Qual* 2015;**44**:895.
- Staley C, Unno T, Gould TJ *et al.* Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. *J Appl Microbiol* 2013;**115**:1147–58.
- Stedtfeld RD, Stedtfeld TM, Waseem H *et al.* Isothermal assay targeting class 1 integrase gene for environmental surveillance of antibiotic resistance markers. 2017, DOI: 10.1016/j.jenvman.2017.04.079.
- Vaz-Moreira I, Nunes OC, Manaia CM *et al.* Bacterial diversity and antibiotic resistance in water habitats: searching the links with the human microbiome. *FEMS Microbiol Rev* 2014;**38**:761–78.
- Vester B, Douthwaite S. Macrolide resistance conferred by base substitutions in 23S rRNA. *Antimicrob Agents Chemother* 2001;**45**:1–12.
- Wang F-H, Qiao M, Su J-Q *et al.* High Throughput Profiling of Antibiotic Resistance Genes in Urban Park Soils with Reclaimed Water Irrigation. *Environ Sci Technol* 2014;**48**:9079–85.
- Weisblum B. Macrolide resistance. *Drug Resist Updat* 1998;**1**:29–41.
- Whitehead TR, Cotta MA. Stored swine manure and swine faeces as reservoirs of antibiotic resistance genes. *Lett Appl Microbiol* 2013;**56**:264–7.
- WHO | Global action plan on antimicrobial resistance. *WHO* 2017.

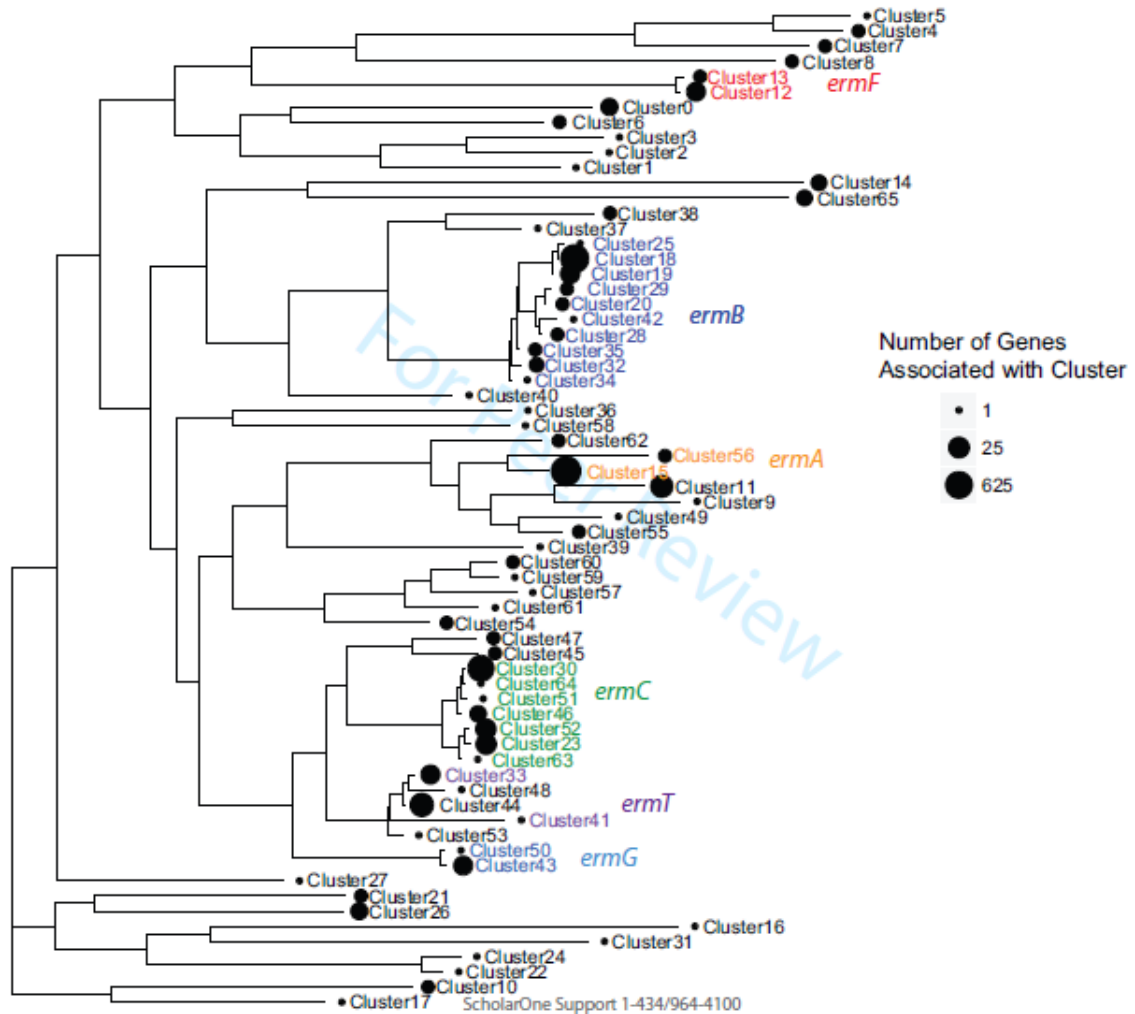


Figure 1. Maximum likelihood phylogenetic tree of 66 *erm* sequence clusters based on 99% nucleotide similarity of 5,648 DNA sequences extracted from known *erm* genes described in existing databases. Clusters that contain gene targets from existing PCR primers (see Table 2) are highlighted in color. The relative number of sequences comprising each cluster among the 5,648 DNA sequences is also shown.

Figure 2. Average number of *erm* genes in metagenomes from various environments (see Supp. Table 1). *For *ermC* gene, the average number of reads in animal-associated soil metagenomes was 1665 ± 659 reads.

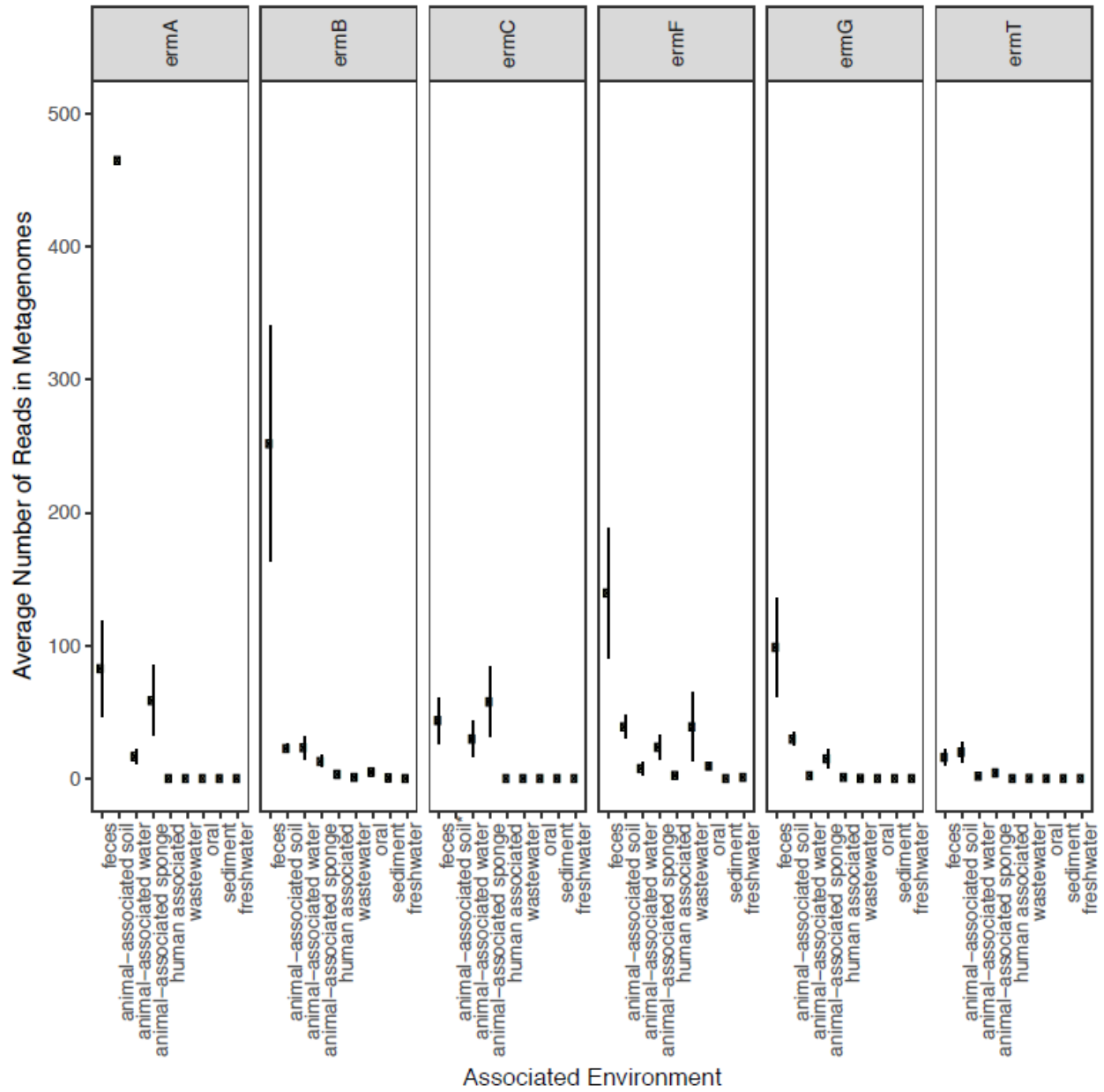


Figure 3. Abundance of DNA sequences homologous to *erm* gene PCR primers as a percent of total *erm* abundance in swine and cattle metagenomes.

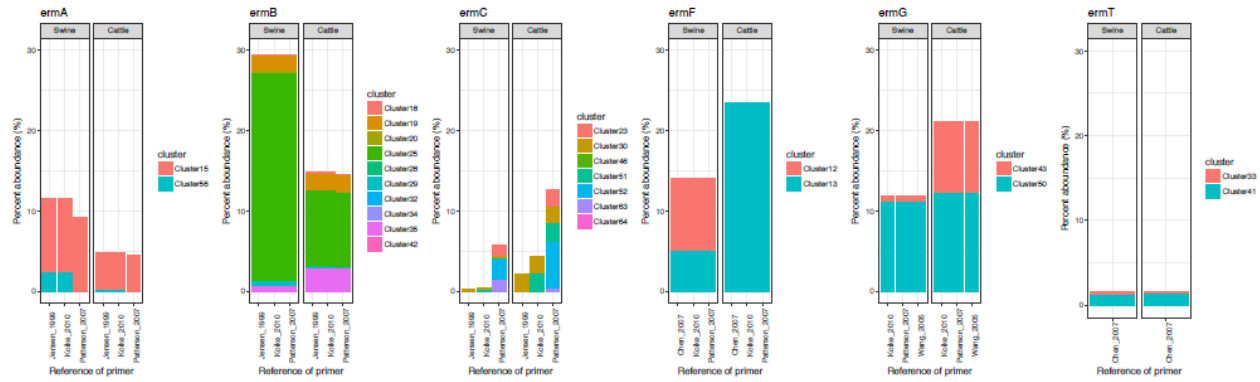


Table 1. *Erm* gene clusters identified from 5,648 *erm* sequences. For each cluster, the most representative gene is referenced by its NCBI accession number in NCBI nucleotide and protein databases.

Cluster (this study)	NCBI Protein Accession No.	NCBI Nucleotide Accession No.	Description in NCBI GenBank	Organism
Cluster 0	BAJ34818	AB601890	erythromycin resistance protein	Photobacterium damsela subsp. piscicida
Cluster 1	KNF08983	LG550100004	rRNA (adenine-N(6)-)-methyltransferase	Clostridium purinilyticum
Cluster 2	AFS78141	CP003326	rRNA (adenine-N(6)-)-methyltransferase	Clostridium acidurici 9a
Cluster 3	ABW20380	CP000853	rRNA (adenine-N(6)-)-methyltransferase	Alkaliphilus oremlandii OHILAs
Cluster 4	KKS60599	LCDU01000003	rRNA (Adenine-N(6)-)-methyltransferase	candidate division WWE3 bacterium GW2011_GWF2_42_42
Cluster 5	KKS35651	LCCU01000032	rRNA (Adenine-N(6)-)-methyltransferase	candidate division WWE3 bacterium GW2011_GWF1_42_14
Cluster 6	AJB79756	CP010391	hypothetical protein	Klebsiella pneumoniae
Cluster 7	EKD94896	AMFJ01010665	hypothetical protein	uncultured bacterium
Cluster 8	KKU26033	LCLY01000007	rRNA (Adenine-N(6)-)-methyltransferase	Microgenomates group bacterium GW2011_GWA2_46_16
Cluster 9	CCQ93859	CARA01000062	rRNA adenine N-6-methyltransferase	Clostridium ultunense Esp
Cluster 10	BAP00917	AP013353	dimethyladenosine transferase	Mycoplasma californicum HAZ160_1
Cluster 11	CNU04734	CQCN01000003	dimethyladenosine transferase	Streptococcus agalactiae
Cluster 12	AAA27431	M17808.1	ermF	Bacteroides fragilis
Cluster 13	AAA63165	M62487.1	ermF	Bacteroides fragilis
Cluster 14	EEO52603	ACAB02000055.1	ermF	Bacteroides sp. D1
Cluster 15	CCJ25599	HE579073	rRNA adenine N-6-methyltransferase	Staphylococcus aureus subsp. aureus ST228
Cluster 16	CCX90994	CAXH01000024	ribosomal RNA small subunit methyltransferase A	Succinatimonas sp. CAG:777
Cluster 17	EGV00599	AFXA01000001	dimethyladenosine transferase rRNA modification enzyme	Mycoplasma columbinum SF7
Cluster 18	EFY03905	AEVND1000118	rRNA adenine N-6-methyltransferase	Phascolarctobacterium succinatutens YIT 12067
Cluster 19	ACB90575	CP001033	erythromycin ribosome methylase	Streptococcus pneumoniae CGSP14
Cluster 20	EJY36237	AMBI01000188	rRNA adenine N-6-methyltransferase	Enterococcus faecium 510
Cluster 21	AKB11102	CP011096	16S rRNA methyltransferase	Mycoplasma synoviae ATCC 25204
Cluster 22	ACD66486	CP001080	dimethyladenosine transferase	Sulfurihydrogenibium sp. YO3AOP1
Cluster 23	EIB96299	AICL01000010	rRNA methylase	Lactobacillus salivarius SMXD51
Cluster 24	EFP06650	ABZS01000069	dimethyladenosine transferase	Sulfurihydrogenibium yellowstonense SS-5
Cluster 25	AFV15157	JQ655732	erythromycin	Clostridium perfringens
Cluster 26	KDE45359	JFKK01000007	16S rRNA methyltransferase	Mycoplasma hyosynoviae
Cluster 27	ADM89794	CP002161	putative dimethyladenosine transferase	Candidatus Zinderia insecticola CARI
Cluster 28	KER55751	JPHP01000035	SAM-dependent methyltransferase	Bacteroides fragilis
Cluster 29	AAR27225	AY357120	N-methyltransferase	Streptococcus pyogenes
Cluster 30	AIU96746	KF831357	ErmC	Staphylococcus aureus
Cluster 31	ACG57739	CP001130	ribosomal RNA adenine methylase transferase	Hydrogenobaculum sp. Y04AAS1

Table 2. Previously published PCR primer and gene targets for *erm* genes.

Gene	Cluster	Primers Design	Papers Citing Primers
<i>ErmA</i>	15	Patterson et al. 2007	– ^b
	n/a ^a	Sutcliffe et al. 1996	Martel et al., 2003a; Jackson et al., 2004; Luthje & Schwarz, 2006; Garofalo et al., 2007; Chenier & Juteau, 2009; Zou et al., 2011; Di Cesare et al., 2012; Hoang et al., 2013; Lerma et al., 2014
	15, 56	Jensen et al. 1999	Aarestrup et al., 2000a; Aarestrup et al., 2000b; Jensen et al., 2002; Petersen & Dalsgaard, 2003; Whitehead & Cotta 2013
	n/a	Chen et al. 2007	Sharma et al., 2009; Just et al., 2011; Alexander et al., 2011; Wang et al., 2012; Holman & Chenier 2013; Wang et al., 2015; Xu et al., 2016
	15, 56	Koike et al. 2010	Ekizoglu et al., 2013
<i>ErmB</i>	n/a	Sutcliffe et al. 1996	Martel et al., 2003a; Martel et al., 2003b; Cauwerts et al., 2007; Ahmad et al., 2011; Hoang et al., 2013
	18, 19, 20, 25, 28, 29, 32, 34, 35, 42	Jensen et al. 1999	De Leener et al., 2005; Whitehead & Cotta, 2013
	18, 19, 20, 25, 28, 29, 32, 34, 35, 42	Chen et al. 2007	Sharma et al., 2009; Chen et al., 2010; Alexander et al., 2011; Just et al., 2011; Kalmokoff et al., 2011; Negreanu et al., 2012; Holman & Chenier, 2013; Beukers et al., 2015; Wang et al., 2015; Sandberg & LaPara, 2016; Xu et al., 2016
	18, 19, 20, 25, 29, 32, 35, 42	Patterson et al. 2007	Knapp et al., 2010
	18, 19, 20, 25, 28, 29, 32, 34, 35, 42	Koike et al. 2010	Ekizoglu et al., 2013; Garder et al., 2014; Joy et al., 2013; Joy et al., 2014; Soni et al., 2015; Luby et al., 2016
<i>ErmC</i>	n/a	Sutcliffe et al. 1996	Martel et al., 2003; Hoang et al., 2013
	30, 46	Jensen et al. 1999	Ekizoglu et al., 2013; Whitehead & Cotta, 2013
	23, 30, 46, 51, 52, 63, 64	Patterson et al. 2007	Knapp et al., 2010; Popowska et al., 2011
	30, 46, 51	Koike et al. 2010	Luby et al., 2016
<i>ermF</i>	12, 13	Chen et al. 2007	Sharma et al., 2009; Chen et al., 2010; Alexander et al., 2011; Kalmokoff et al., 2011; Negreanu et al., 2012; Wang et al., 2012; Hoang et al., 2013; Holman & Chenier 2013; Farenfeld et al., 2014; Garder et al., 2014; Wang et al., 2015; Luby et al., 2016; Xu et al., 2016
	12, 13	Patterson et al. 2007	Knapp et al., 2010
	12, 13	Koike et al. 2010	Ekizoglu et al., 2013; Joy et al., 2013; Joy et al., 2014
<i>ErmG</i>	43, 50	Wang et al. 2005	Wang et al. 2005; Kalmokoff et al., 2011
	43, 50	Patterson et al. 2007	N/A
	43, 50	Koike et al. 2010	Ekizoglu et al., 2013
<i>ErmT</i>	33, 41	Chen et al. 2007	Sharma et al., 2009; Alexander et al., 2011; Kalmokoff et al., 2011; Wang et al., 2012; Hoang et al., 2013; Garder et al., 2014; Wang et al., 2015

^a Primers did not hit any clusters

^b No relevant citing papers