A health metadata-based management approach for comparative analysis of high-throughput genetic sequences for quantifying antimicrobial resistance reduction in Canadian hog barns

Samuel M. Chekabab [a,b], John R. Lawrence [c], Alvin Alvarado [b], Bernardo Predicala [b], Darren R. Korber [a,*]

[a] *Food and Bioproduct Sciences, University of Saskatchewan, Saskatoon, SK S7N 5A8, Canada*
[b] *Prairie Swine Centre Inc., Box 21057, 2105 – 8th Street East, Saskatoon, SK S7H 5N9, Canada*
[c] *Environment and Climate Change Canada, 11 Innovation Blvd., Saskatoon, SK S7N 3H5, Canada*

ARTICLE INFO

ABSTRACT

New Canadian regulations have required that all use of antibiotics in livestock animal production should be under veterinary prescription and oversight, while the prophylactic use and inclusion of these agents in animal feed as growth promoters are also banned. In response to this new rule, many Canadian animal producers have voluntarily implemented production practices aimed at producing animals effectively while avoiding the use of antibiotics. In the swine industry, one such program is the 'raised without antibiotics' (RWA) program.

In this paper, we describe a comprehensive investigative methodology comparing the effect of the adoption of the RWA approach with non-RWA pig production operations where antibiotics may still be administered on animals as needed. Our experimental approach involves a multi-year longitudinal investigation of pig farming to determine the effects of antibiotic usage on the prevalence of antimicrobial resistance (AMR) and pathogen abundance in the context of the drug exposures recorded in the RWA versus non-RWA scenarios.

Surveillance of AMR and pathogens was conducted using whole-genome sequencing (WGS) in conjunction with open source tools and data pipeline analyses, which inform on the resistome, virulome and bacterial diversity in animals and materials associated with the different types of barns. This information was combined and correlated with drug usage (types and amounts) over time, along with animal health metadata (stage of growth, reason for drug use, among others). The overarching goal was to develop a set of interconnected informatic tools and data management procedures wherein specific queries could be made and customized, to reveal statistically valid cause/effect relationships.

Results demonstrating possible correlations between RWA and AMR would support the Canadian pig industry, as well as regulatory agencies in new efforts, focused on reducing overall antibiotics use and in curbing the development and spread of AMR related to animal agriculture.

1. Introduction

Antimicrobial drug pollution in aquatic ecosystems and food webs constitute a serious concern in worldwide public health, agriculture, aquaculture and environmental biodiversity and sustainability. Worldwide, antibiotic usage exceeds 100,000 tons per year [1]. Levels of antimicrobial use varies dramatically from country to country, for example, some countries in Europe reportedly use lower quantities of drugs to treat animals compared with humans [2]. In Canada, however, 82% of total antimicrobial active ingredients ($\approx 1.5 \times 10^6$ kg) sold in 2014 was distributed for use in food-producing animals [3]. This amount is consistent globally, where an estimated 73% of all antimicrobials are administered on farm animals [4]. This extensive and increasing usage of antibiotics in agriculture and aquaculture has been on-going for around 75 years, with usage trends correlating with antimicrobial drug discoveries made over the period [5,6,7]. The main consequence of this is the appearance of antibiotic-resistant bacterial strains that have undergone selection in various human/animal and environmental microbial communities in what is commonly referred to as antimicrobial resistance (AMR). Over the past decade, concerns

---

* Corresponding author.
  E-mail address: darren.korber@usask.ca (D.R. Korber).

about AMR have increasingly been raised by various scientific communities; in future years, it will certainly be one of the major health challenges that will need to be addressed in order to maintain a supply of clinically-relevant drugs for human and animal therapeutic use.

The industry of pig farming is a major economic player in the Canadian animal agriculture sector. Overall, the antibiotics commonly and extensively utilized by pig producers belong to the tetracycline and macrolide classes [8]. Historically, approximately only 25% of the total antibiotics used are for treating diseases. The remaining quantity of drugs was used for disease prevention and growth promotion. This implies that, even when pathogenic bacteria are the intended targets of the antibiotics, many non-pathogenic bacterial species are also affected and thus may constitute a continuing reservoir of AMR. Since simply using antibiotics may contribute to AMR, it follows that any drugs used should strictly be applied to treat infections in sick animals. In response to increasing public apprehension regarding the use of antibiotics in livestock production along with general concerns about the spread of AMR, various measures, such as a total ban on the use of antibiotics in livestock feed as growth promoters and strict regulations on the use of any antibiotic for the treatment of sick animals, are being implemented in Canada. Indeed, these regulations were implemented by the Public Health Agency of Canada (PHAC) in December 2018, which mandate that the use of all medically important antimicrobials (MIAs) in animal agriculture production now fall under veterinary prescription and oversight and that these agents can no longer be provided prophylactically, nor included as growth promoters in animal feed [9]. In advance of this new rule, some Canadian producers voluntarily implemented procedures wherein animals are effectively raised without the use of antibiotics (RWA) from birth to slaughter.

In this paper, we describe an integrated investigation methodology comparing the effect of the adoption of the RWA approach with non-RWA conventional operations where antibiotics may be administered to any animal as needed, as opposed to the RWA approach wherein any sick animals that require antibiotics are taken out of the program and marketed as conventional non-RWA pigs. Here, we employed a longitudinal study of pig farms to determine the effects of the RWA program on the prevalence of antimicrobial resistance genes (ARGs) as well as pathogen abundance. The approaches employed in this study are intended to facilitate the quantification of the effectiveness of reducing the total on-farm use of antibiotics on the frequency of detection of both ARGs and the pathogens that frequently carry them. Accordingly, we present an experimental design based on multi-year surveillance monitoring of conventional farms still using antibiotics as usual (except in feed) and farms that have entered the RWA program. This monitoring strategy focuses on 3 key parameters: antibiotic usage, antibiotic resistance, and prevalence of pathogens. The workflow proposed here covers a multi-step study experimental design performed: i) over time, ii) with the collection and transcription of data records (drugs and treatment reasons) from participating barns, iii) sampling farm animals and their environment, iv) subjecting samples to whole-genome sequencing (WGS), and v) processing WGS data via comprehensive metagenomic analyses through the CosmosID platform (taxonomy and resistome profiling, and pathogens prevalence).

WGS is a widely available tool with demonstrated potential for AMR surveillance; its utilization can help decipher many complex microbiota-related questions including possible exploratory correlations between various available microbially-related databases [10,7]. In this study, we illustrate several examples of AMR surveillance and determining pathogen prevalence from WGS data analyzed through user-friendly pipelines that provide information on the resistome profiles, as well as bacterial diversity. Such information collected and combined with drug-use and animal health metadata can be used to reveal a potential correlation between RWA and AMR, and thus support Canadian pig farmers in their efforts to reduce antibiotics use in pig production as well as health and agriculture agencies in their governance plans on antibiotic regulations.

2. Methodology

2.1. Experimental design

Our methodology included the periodic collection of metadata comprised of all records of administered drugs and illnesses or treatment reasons reported in the barn. Samples from the animals and the barn environment were also periodically collected then sequenced and metagenomically-compared.

2.2. Sample collection

A 2.5-year longitudinal surveillance schedule mainly based on a 6-month sampling interval, from both animals and barn environments, was conducted. Animal samples included fecal and nasal swabs obtained from 6-week old piglets and 3rd-parity sows. At each time point, both the piglets and the sows were sampled to examine the age-dependent effect. Fecal samples were collected aseptically from 3 animals of each age (piglets or sows) and stored in 50 mL sterile tubes, whereas the nasal swabs were collected from 3 animals using the Swab Collection and DNA Preservation System (Norgen Biotek Corp. Thorold, ON). The nasal swab samples were aimed to detect potential subsets of respiratory viruses along with other microorganism categories and their associated AMR that may not be well represented in the manure and fecal samples. Environmental samples included manure and soil collected from inside and around the barns and stored in 50 mL sterile tubes. Manure samples were collected from the manure lagoon or the manure collection pit, while soil samples were collected from the ground surface near the entrance of the barns. Sample handling was conducted in accordance with CDC's Biosafety in Microbiological and Biomedical Laboratories (BMBL) manual for level 1 materials [11].

2.3. WGS Shotgun-Illumina HighSeq

To identify the total ARGs (resistome) and bacterial-related diversity as well as the prevalence of pathogens in the collected samples, random shotgun next-generation sequencing (NGS) was performed via a third-party service using an Illumina HiSeq platform (Omega-Bioservices, Norcross, GA, USA). Samples were handled according to the sequencing service procedures and shipped to Omega-Bioservices for DNA extraction, data quality determination, and NGS. Accordingly, 1 g of sample material was used for DNA extraction using the Mag-Bind Universal Pathogen DNA Kit (Omega Bio-tek, Inc. Norcross, GA, USA), and the purity and yield of the DNA were checked using the Quant-iT™ PicoGreen™ ds DNA System kit (ThermoFisher Scientific, Pittsburgh, PA, USA). Shotgun NGS libraries were constructed from DNA using Kapa Biosystems Prep Kit following the manufacturers' protocols (Roche®, KK2103 Pleasanton, CA, USA). Samples representing a distinct time point were run on 1 lane of a HiSeq4000/X Ten (Illumina), generating a total of 100–120 GB of 150-bp paired-end data reads. Eight samples per run produced an average minimum of ~30 Million reads (MReads) per sample with each sample generating 2 FASTQ files (R1 Forward read and R2 Reverse read) shared through the BaseSpace Sequence Hub. Sequences were then subjected to quality control processes (i.e., denoising and trimming

the adaptors) and reported with the MultiQC tool (https://multiqc.info/) prior to uploading onto the CosmosID platform for metagenomic analysis (CosmosID Inc., Rockville, MD, USA). CosmosID focuses on the rapid gene markers characterization of microorganisms, pathogens and anti-microbial resistance for infectious disease identification, food safety inspections, pharmaceutical discovery, public health surveillance and microbiome analysis [12,13].

2.4. CosmosID pipeline: data normalization and analyses

Eight samples per run produced between 20 and 100 Million reads (MReads) per sample, for a total number of reads ranging from 1000 to 1200 Mreads. For comparative analyses among the non-RWA and RWA groups, all data sets were subsampled to a fixed 20 MRead depth to ensure uniform population diversity and reduce bias in the data analyses arising from variation in read depth. This subsampling method was obtained by rarefying randomly sampled reads without replacement from each of the 8 samples up to the common count of 20Mreads using the seqtk tool package available at https://github.com/lh3/seqtk. The rarefying depth from each run was set to the lowest Mreads/Sample with a total coverage while the rest of the reads were discarded [14]. Thereby, the difference in the number of reads obtained from different samples reflects biological differences in the samples.

In our described workflow, the metagenomic shotgun analysis employed a functionality-based strategy wherein functional gene products were identified regardless of which bacterial/microbial species the genetic material originated from. This read-based profiling method allowed multiple profiling targets, including those based on taxonomic, resistome and virulome criteria. In this procedure, all our unassembled sequencing reads were analyzed using the CosmosID software package that utilizes data mining algorithms and curated databases that provide fine resolution for organism identification and discrimination at the strain-level, as well as genes of interest and accurate measurement of their relative abundances. CosmosID algorithm is based on GENIUS Software that uses data-mining K-mer and highly-curated dynamic comparator databases, such as GenBook [15], that disambiguates Mreads from a metagenomic sample into discrete microorganisms (or potential gene products) engendering the particular sequences. A complete description of the k-mer-based profiling assignment algorithms used is provided elsewhere [16,17,14,18].

2.5. CosmosID comparators

GenBook comparator databases comprise nearly 160,000 phylogenetically organized genomes and gene sequences. The results include tables and visualizations for genome databases (i.e., bacteria, fungi, protists, viruses, and respiratory viruses) as well as for gene databases (i.e., antimicrobial resistance and virulence factors). The CosmosID platform constitutes both a subset of genomes sequenced by CosmosID and its collaborators, in addition to publicly available genomes or gene sequences through NCBI- RefSeq/WGS/SRA/nr, PATRIC, M5NR, IMG, ENA, DDBJ, CARD, ResFinder, ARDB, ARG-ANNOT, MVIRDB, and VFDB. The CosmosID pipeline has been optimized for processing unmapped/unaligned sequence reads of lengths <100 bp and offers a flexible tool used to compare whole genomes with enhanced discriminatory power. It uses a first precomputation phase that outputs a reference microbial database to a whole-genome phylogeny tree, with sets of fixed-length k-mer fingerprints identified with distinct nodes of the tree [12,19]. A second per-sample computation phase searches short sequence reads against the fingerprint sets and gives fine-grain composition and relative abundance estimates at all nodes of the tree; it uses edit distance-scoring techniques to compare a target sample with a reference set [12,19]. The first comparator finds reads with an exact match with an n-mer uniquely identified with a set of reference strains, with the second statistically scoring the entire read against the reference to verify that the read is uniquely identified with that set. The analysis turnaround time was 20–30 min from uploading sequences to metagenomic reporting that included four profiling variables for each hit detected (bacterium, virus, or ARG... etc); i) *unique match frequency* which is the number of unique k-mer found in the queried sample and equivalent to the number of reads that match the hit. ii) *unique match %* is the number of unique matches divided by the number of total unique patterns for that hit in the reference database. iii) *total matches %* is the total number of matches (unique + shared) divided by the total number of patterns for that hit in the reference database. iv) *relative abundance* is calculated based on the number of hit-specific k-mers and their observed frequency in the sample and then normalized to represent the abundance of each detected hit. In addition, CosmosID provides a filtering method that allow calls to meet a filtering threshold for high confidence that these calls are in the sample. The filtering threshold which determines if results are considered significant is based on internal statistical scores determined by analyzing a large number of diverse metagenomes. Organisms listed in the filtered results are likely to be present in the sample. Unfiltered (total) results include those in the filtered results and usually additional organisms. Organisms listed in the unfiltered (total) results but not in the filtered results need further validation to determine if they are actually present in the sample – either by deeper sequencing of the sample followed by re-analysis or by orthogonal validation using targeted PCR or other method.

2.6. Exploration of possible readouts

The profiling of taxonomy is achieved through microbial identification to the species, subspecies, and strain level along with quantification of identified organism's relative abundance at each taxonomic level through GenBook comparators and the GENIUS software implemented within the CosmosID algorithm. CosmosID databases currently contain over 15,000 bacterial, 5000 viral, 250 protists and 1500 fungal species. These include curated public databases NCBI- RefSeq, PATRIC, M5NR, IMG, ENA, and DDBJ.

The "Pathobacteriome" profile was identified using the bacterial pathogens subset of the taxonomy profiles. This subset includes only pathogenic bacteria identified at the strain level. The table listing of bacterial species obtained from taxonomy profiling was further used for manual assignment of risk groups (RG1, RG2, etc.). The bacterial species lists were first queried against the animal and human RG database https://health.canada.ca/en/epathogen. This database comprises 1479 RG2 entries, including 789 RG2 for humans, 227 RG2 for animals and 463 RG2 for both. Unclassified species were further checked against the Bacterial Diversity Meta-database, available in the public domain https://bacdive.dsmz.de. The pathobacteriome profiles represented all identified organisms classified as risk groups other than the RG level 1.

Similarly, the resistome, the collection of ARGs in the microbiome, was also profiled by querying unassembled sequence reads against CosmosID's curated ARG database, generating a table list of identified and quantified ARGs. The CosmosID ARG database is organized as a phylogenetic tree, which avoids the potential problems of highly similar sequences affecting abundance estimates. It is also the result of combining multiple ARG databases, including NCBI- RefSeq, PATRIC, M5NR, ENA, DDBJ, CARD, ResFinder, ARDB, and ARG-ANNOT. Altogether, these databases comprise over 3600 ARGs. ARGs were identified based on percent of gene coverage for each gene as a function of the gene-specific read frequency in each sample. For each reference gene, sets of unique k-mers that

span the entire gene are interrogated through the data sets and the average frequency of all k-mers is recorded. This approach circumvents the need for read-assembly for each gene. The resultant ARG profile table was then clustered into 16 classes of drug resistance and 7 mechanisms of resistance, and based on a classification combined from 2 pipelines with a focus on antimicrobial resistance (https://megares.meglab.org/ and https://card.mcmaster.ca/).

The "Virome" profile identifies and quantifies the total viruses present in the microbiome, with the ability to separate subsets of respiratory viruses and bacteriophages.

The collection of virulence genes (VG) in the microbiome (the virulome) can also be derived by querying unassembled sequence reads against CosmosID-curated VG databases. The CosmosID VG database is curated by combining mvirdb, VFDB and PATRIC databases.

2.7. Statistical analysis

The diversity, ordination and differential abundance testing, and both multivariate and univariate analyses, were applied to the resultant taxa/ARG/VF abundance tables. These were used to calculate observed and expected species richness (Shannon alpha diversity indices and beta diversity distance matrices). We sampled 5 farms (2 non-RWA and 3 RWA). For statistical analysis, and for equal distributed comparisons, we included data from 4 farms (2 of each type of farm considered as 2 biological replicates). Samples were repeated over time (to date, 3 time points have been collected with a 6-month interval, with 2 future time point samples planned as this is an ongoing study). At each time point, both the piglets and the sows were sampled to examine the age-dependent effect. For all features (pathogen species, respiratory virus, ARG class and virulence factors) Principle Coordinate Analysis (PCoA) was performed to cluster samples based on abundance (Jaccard distance matrix; community structure). Differentially abundant features were identified using permANOVA, with PC-ORD 7 software, and the estimation of the different microbial communities' relationships were calculated to generate PCoA plots representing a mathematical evaluation of the distance and correlation among the microbial communities. Individual features such as individual ARG classes of drug resistance were compared using 2-way parametric ANOVA with 'non-RWA' and 'RWA' as barn groups, and 'fecal', 'manure' and 'nasal swab' or 'timepoint repeated measurements' as sub-groups.

2.8. Data availability

DNA metagenomics sequencing data are available in Sequence Read Archive (SRA) (https://submit.ncbi.nlm.nih.gov/subs/sra/), Accessions PRJNA633402 (Fecal-piglet 1), PRJNA633385 (Fecal-piglet 2), PRJNA633399 (Fecal-piglet 3) and PRJNA633392 (total Manure).

3. Results and discussion

3.1. Examples in determining on-farm antibiotic usage patterns and total use of drugs

To obtain accurate documentation of all drugs used in participating farms, we monitored and recorded the inventory and usage of drugs on the farm, from August 2018 until December 2019 (16 months). These records included the type of drug/vaccines, dosage, animal stage of development, reason for treatment, as well as the date of drug administration. This information was extracted from the routine barn treatment data sheets that are currently required for all farms as part of the Canadian Quality Assurance (CQA) of the Canadian Pork Excellence (CPE) program. Table 1 shows a list of drugs and dosages used in the participating barns. The list of drugs comprises 6 antibiotics, 3 anti-inflammatories, 6 vaccines, 2 anti-parasitics, and 2 supplements. The information and amounts of drugs given were cross-referenced against the Provincial Veterinary Services and swine drug treatment databases (https://www.drugs.com/vet/swine-a.html). The antibiotics mostly given under the non-RWA production operation and belonged to four classes: Antifolates (Trimidox), β-lactams (Penicillin G, Ampicillin, Ceftiofur), Tetracyclines (Biomycin) and Amphenicols (Nuflor) (Table 2). For the 16-month period indicated above, more than 15.5 kg of antibiotics were used in the non-RWA barns whereas less than 1 kg was used in the RWA barns (Table 2). In non-RWA barns, a total of 22,499 treated animals were comprised of 82% piglets, 16% grow-finishers and 2% sows; in these barns, a combined 9410 g of Antifolates, 2048 g of β-lactams and 4113 g of Tetracyclines were used over the period. In RWA barns, a total of 1338 animals (91% piglets, 5% grow-finishers and 4% sows) received a combined total of 160 g of Antifolates, 477 g of β-lactams, 99 g of Tetracyclines, and 208 g of Amphenicols. Overall, the antibiotic usage patterns reported here corroborate the Canadian Integrated Program for Antimicrobial Resistance Surveillance (CIPARS) 2016 report that collects, analyses, and communicates

Table 1
Compilation of all types (and corresponding dosage) of drugs recorded in the various participating barns.

Type of drugs	Drug name and concentration	Dosage
Antibiotics	Trimidox (trimethoprim & sulfadoxine 40 mg:200 mg/mL)	1 mL/15 kg/day
	Biomycin (Oxytetracycline 200 mg/mL)	1 mL/10 kg/day
	Penicillin G (1 mL = 3.105 IU = 297 mg)	6000 IU per kg (1 mL/50 kg)
	Polyflex (ampicillin) (broad-spectrum penicillin 25 g)	6 mg/kg/day
	Cevaxel RTU/Excenel (ceftiofur 50 mg/mL)	3 mg/kg/day
	Nuflor (Florfenicol 300 mg/mL)	15 mg/kg/day
Anti-inflammatory	Predef (isoflupredone acetate 2 mg/mL)	5 mg/140 kg/day
	Anafen (Ketoprofen 100 mg/mL)	9 mg/kg/day
	Metacam (Meloxicam 5 mg/mL)	0.4 mg/kg/day
Vaccines	Porcilis Ileitis vaccine	
	Suvaxyn E-Oral (Haemophilus parasuis Bacterin)	2 mL/pig
	FarrowSure B (Parvovirus vaccine)	
	Fostera PCV (Type 1-Type 2 Chimera)	
	LitterGuard Rotavirus Vac (C. perfringens/E. coli Bacterin-toxoid)	
	Circo/MycoGard (Porcine Circovirus Vaccine, Type 2)	
Anti-parasitic	Bimectin (22,23-dihydroavermectin B)	200 mg/kg
	Oxyto-sure (oxytocin)	0.25–1.0 mL (5–20 units)
Supplements	Vitamaster (Mix vitamins/amino acids/microminerals)	1 mL/45 kg/day
	Iron Dexafer 200 (ferric hydroxide dextran 200 mg/mL)	3 mL/piglet

Table 2
Quantification of antibiotics given in the barns for the 3 production areas - P: Piglets, G-F: Grow-Finishers and S: Sows.

		non-RWA				RWA			
Stage and % of animals		*82% P*	*16% GF*	*2% S*	*Total*	*91% P*	*5% GF*	*4% S*	*Total*
Number of treated animals		18,397	3646	456	22,499	1222	63	53	1338
Antibiotics	Antifolates (Trimidox)	4062 g	2170 g	3178 g	9410 g	160 g	0 g	0 g	160 g
	β-lactams	633 g	1401 g	14 g	2048 g	177 g	68 g	232 g	477 g
	Tetracyclines (Biomycin)	1971 g	1586 g	556 g	4113 g	29 g	2 g	68 g	99 g
	Amphenicols/Nuflor	0 g	0 g	0 g	0 g	206 g	2 g	0 g	208 g
	Total of Antibiotics				**15571 g**				**944 g**

trends in antimicrobial use and in antimicrobial resistance for select bacteria from humans, animals, and retail meat across Canada [20]. Indeed the CIPARS [20] reported a comparable (i.e. tetracyclines 46%) relative proportions of antimicrobial classes utilized in pigs. This, however, was focused on reporting data from grower-finisher pig herds between 2012 and 2016 and reported the antibiotic drugs as DDDvetCA = Canadian Defined Daily Doses (average labelled dose) in milligrams per kilogram grower-finisher pig weight per day (mg drug/kg animal/day). Our approach here reported the exact amount (in grams) of the drugs used in all development stages of pig production. Thus, our approach is more comprehensive because it will allow examination of the possible age-dependent effect of RWA measures. In addition, the vaccination and other non-antibiotic drug data were systematically recorded and kept as metadata to possibly support further comparative analyses between barns (data not shown).

3.2. Examples in clustering and reporting symptoms, diseases, and treatment reasons

Barn health metadata regarding the reason for drug treatments included 33 separate treatment categories that were clustered based on similarities (and synonyms) of the clinical symptoms and on their respective frequencies. This metadata standardization allows monitoring and comparison of treatment reasons in RWA approach with non-RWA conventional operations where antibiotics may still administered to any animal as needed, as opposed to the RWA approach wherein any sick animals requiring antibiotics are taken out of the program and marketed as conventional non-RWA pigs. Overall, the most prevalent illnesses and treatment reasons recorded during the observation period were limping, scours, respiratory impairment and infections. Scours symptoms were frequently observed in non-RWA piglets and were usually treated for 3 days with an anti-inflammatory combined with either Antifolates or Tetracyclines combination of either Antifolates or Tetracyclines with an anti-inflammatory (Fig. 1A). Limping symptoms were the 2nd most important reason for antibiotic treatment in non-RWA barns, and were also treated with Antifolates or Tetracyclines in piglets and sows, β-lactams were also used for this condition in grow-finishers (Fig. 1A, C, E). Limping was the main reason for the use of antibiotics in RWA barns. RWA piglets were either treated with Antifolates or Amphenicols whereas RWA grow-finishers and sows received β-lactams (Fig. 1B, D, F).

3.3. Taxonomy profiling and prevalence of pathogens

Bacterial sequences accounted for 99.6% of all reads that matched k-mer markers. Our CosmosID analysis resulted in the identification of 726 bacterial strains belonging to 406 species, 94 genera, and 12 phyla. To determine the prevalence of pathogens present in the samples, the initial focus of the analysis was on the metagenomic taxonomy displaying the relative abundance and frequencies of bacteria at the class level (Fig. 2) along with species/strain level determinations (data not shown). Fig. 2 shows an

example of taxonomy profiling in both non-RWA versus RWA for both piglet fecal and manure samples. At the phylum level, 80–90% of all the profiles were comprised of Firmicutes and Bacteriodetes. The stack-bars in the graph represent averaged data from biological duplicates of the core bacteriomes from each type of barn. The core microbiomes were comprised of up to 4 to 5 bacterial classes with >3% relative abundance, over 3 time points of each group sampled for the 16-month period. The piglet-fecal core bacteriome was comprised of Bacteroidia (5–36%), Negativicutes (7–24%), Bacilli (9–51%) and Clostridia (20–26%). The manure core bacteriome comprised of γ-Proteobacteria (5–13%), Bacteroidia (9–30%), Negativicutes (3.5–13%), Bacilli (29–53%) and Clostridia (13–25%). This taxonomy profiling does not show any significant RWA-dependent differences or shifts in bacterial class. Indeed, the Permutation-based MANOVA (PerMANOVA) Bray-Curtis test, comparing 8 bacteriome classes, showed no significant difference between RWA and non-RWA treatments (p = 0.445). However, we observed a significant difference in the taxonomic diversity between piglet-fecal and manure samples that was independent to the RWA effect (p = 0.0076).

Pathogen prevalence data was extracted from the subset of human and/or animal RG2 species/strain abundances. In total, the piglet fecal and manure "Pathobacteriomes" were comprised of 154 and 161 pathogenic strains, respectively (See Supplemental Table S1). This "Pathobacteriome" profiling showed an overall reduction in the abundance of pathogens in RWA compared to non-RWA barns (Fig. 3). In non-RWA piglet fecal samples, we found 53, 124 and 63 different pathogens that respectively represented 13%, 9% and 16% of species profiles over the 3 time points #1 (Nov 2018), #2 (May 2019) and #3 (Nov 2019). RWA piglet fecal samples included 55, 63 and 74 pathogens that represented 9%, 3% and 14%, respectively, of the 3 taxonomy profiles obtained over the same respective time points.

In non-RWA manure, 65, 78 and 92 pathogens were identified which represented 14%, 6% and 11% of the species profiles over the respective time intervals. In contrast, the RWA piglet feces had 82, 69 and 103 pathogens respectively representing 10%, 11% and 8% of the 3 taxonomy profiles over time. In addition, we observed that the prevalence of pathogens was significantly reduced in both RWA and non-RWA fecal and non-RWA manure samples collected during spring (May) compared to the fall season (November).

3.4. Examples of resistome profiling

A key objective of this study was to determine and quantify the effect of the RWA production approach on the prevalence of ARG over time. To achieve this, we conducted comparative resistome profiling which showed that ARGs present in the samples belonged to a core resistome comprised of 4 main resistance classes: Tetracyclines (32–50%), Aminoglycosides (18–36%), Macrolides (12–32%) and β-Lactams (3–13%). The remaining ARGs genes were part of the multi-drug resistant group (MDR) or were unclassified (Fig. 4). Beta-diversity analysis of the resistome profiles of a total
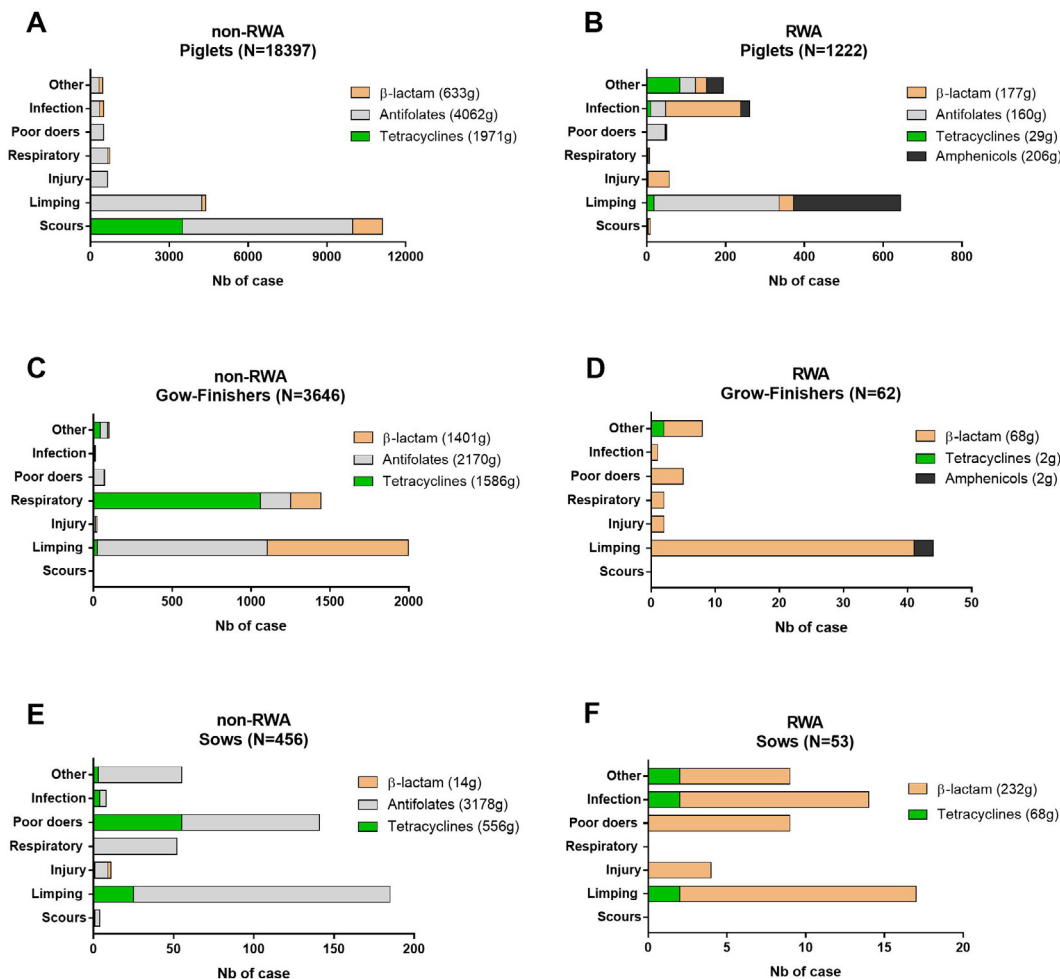
Fig. 1. Cumulative data of clustered disease treatment reasons recorded for 16 months (August 2018–December 2019) from 2 non-RWA barns (A, C and E) and 2 RWA barns (B, D, and F). Stacked bars represent up to 4 different administered antibiotic drugs.
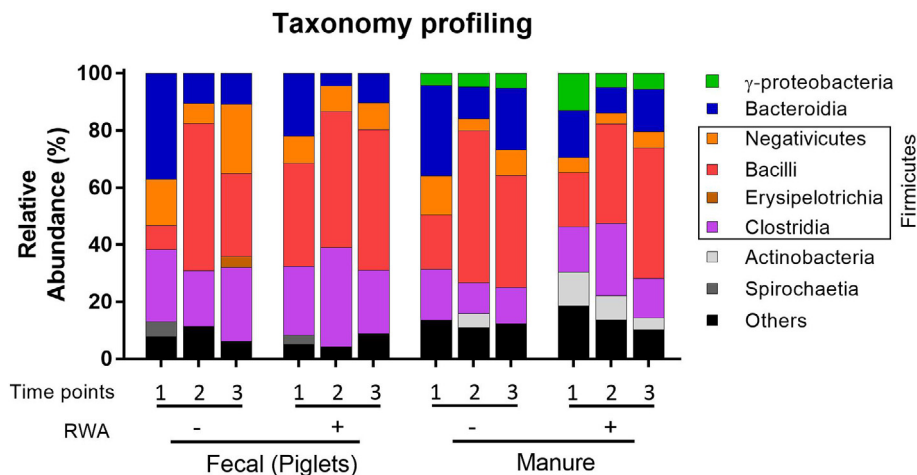

Fig. 2. Metagenomic taxonomy profiling at the class-level from piglet feces and manure samples collected from non-RWA – (minus sign) and RWA + (plus sign). The stacked bars represent averaged relative abundance of the core bacteriomes from each type of barn.

of 24 samples sequenced (3-time points and 2 barns per group) revealed 2 clusters of clearly separate groups of type of samples – Fecal and Manure – with respect to ARGs abundance (Fig. 5). The manure group had two close but distinct, sub-groups that included RWA and non-RWA data. Alpha diversity ordination

showed a significant permANOVA distance ($p < 0.01$) between the groups of RWA resistomes and non-RWA resistomes in manure data (Fig. 6A), which was time-independent ($p > 0.05$) (Fig. 6B). An example of 3D ordination of manure resistomes illustrated that ARG classes of drug resistance can be superimposed on the PCoA
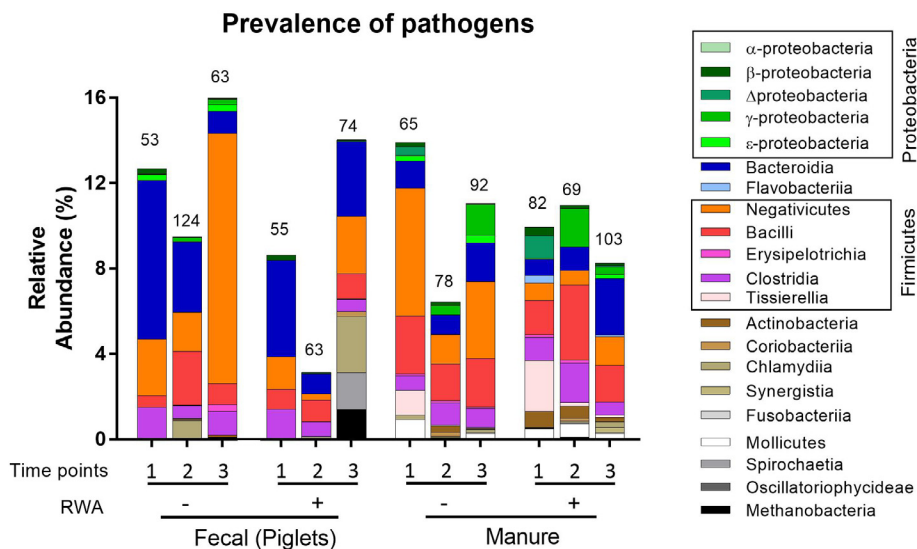
Prevalence of pathogens

Fig. 3. Pathogen prevalence in piglet feces and manure samples. The stacked bars represent the averaged relative abundance of the pathogen lists obtained from all bacteriome classes by extracting the subset of human and/or animal RG2 species/strain.
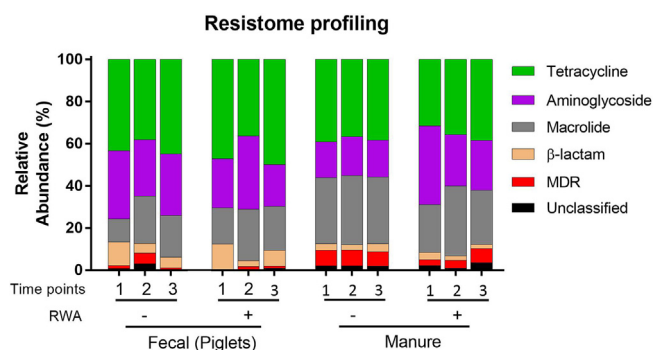

Fig. 4. Metagenomic resistome: A Relative abundance of antibiotic resistance genes (ARG) clustered in 5 classes: Tetracyclines, Aminoglycosides, Macrolides, β-Lactams, and MDR and collected from non-RWA – (minus sign) and RWA + (plus sign).

of the plotted-resistome groups (Fig. 6C). Interestingly, using 2-way parametric ANOVA to compare individual ARG classes of drug resistance in manure samples showed a significant ($p < 0.05$)

decrease in the relative abundance of MDR-ARGs for the 1st time point of the RWA resistome (Fig. 7A). However, a significant ($p < 0.05$) increase in the relative abundance of Aminoglycoside-ARGs was also observed in the RWA resistome (Fig. 7B). On the other hand, the observed RWA effects were not significant with respect to the piglet fecal resistomes (Fig. 7).

In addition to relative abundance normalization readouts shown in the results above, the frequencies of taxa and features can also be examined using an alternative comparison strategy. The frequency readout represents the number of unique k-mer occurrences in the queried sample which is roughly equivalent to the number of reads that matched an identified organism or gene. This readout thus displays the raw number of k-mer hits to unique regions in a genome, rather than the normalized representation that we see with relative abundance. In our data analysis, using frequency as a parameter for comparative resistome analysis, we were able to show that piglet fecal resistomes have significantly decreased Tetracycline-ARGs and MDR-ARGs frequencies in RWA compared to non-RWA resistomes (data not shown). For instance, combining both readouts (abundance and frequency) helps to significantly differentiate between RWA and non-RWA effects.
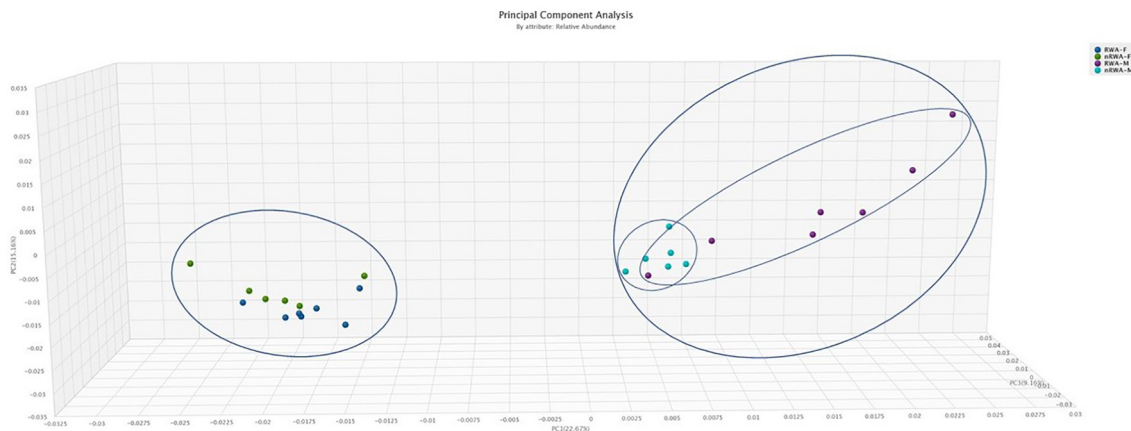

Fig. 5. Resistome comparative beta diversity analysis prepared using the CosmosID pipeline showing 2 distinct groups: the ellipse on the left represents the Fecal group (containing a non-distinct sub-group of RWA (RWA-F) and non-RWA (nRWA-F)). The large ellipse on the right represents the Manure group containing 2 close, but distinct, sub-groups of RWA (RWA-M) and non-RWA (nRWA-M).
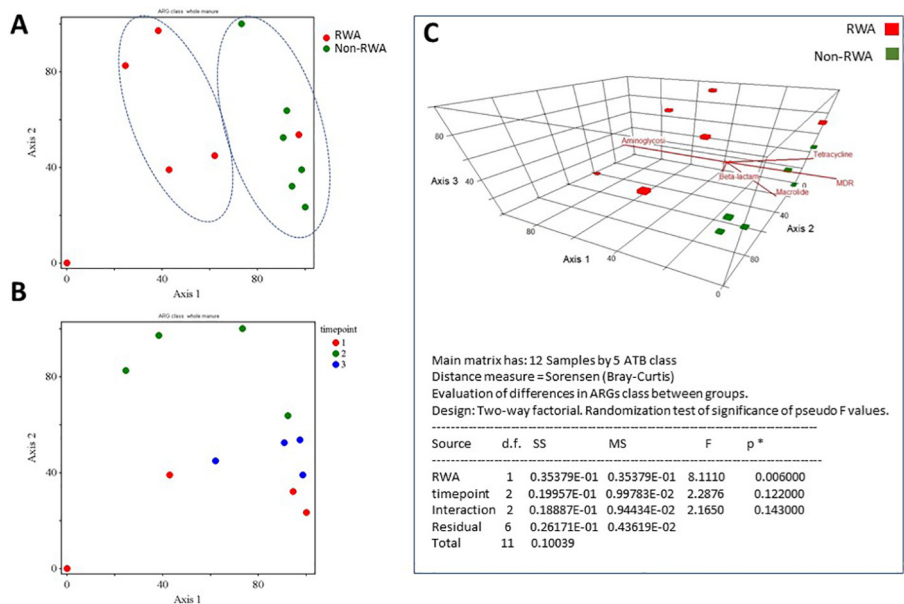
Fig. 6. Ordination permANOVA analysis of the resistome profiling: (A). Ordination manure resistome RWA vs. non-RWA. (B). Type dependent ordination in manure resistomes. (C). 3D ordination of manure resistomes showing drug resistance classes that co-occurred with PCoA of the plotted-resistome groups and the results of permANOVA tests.
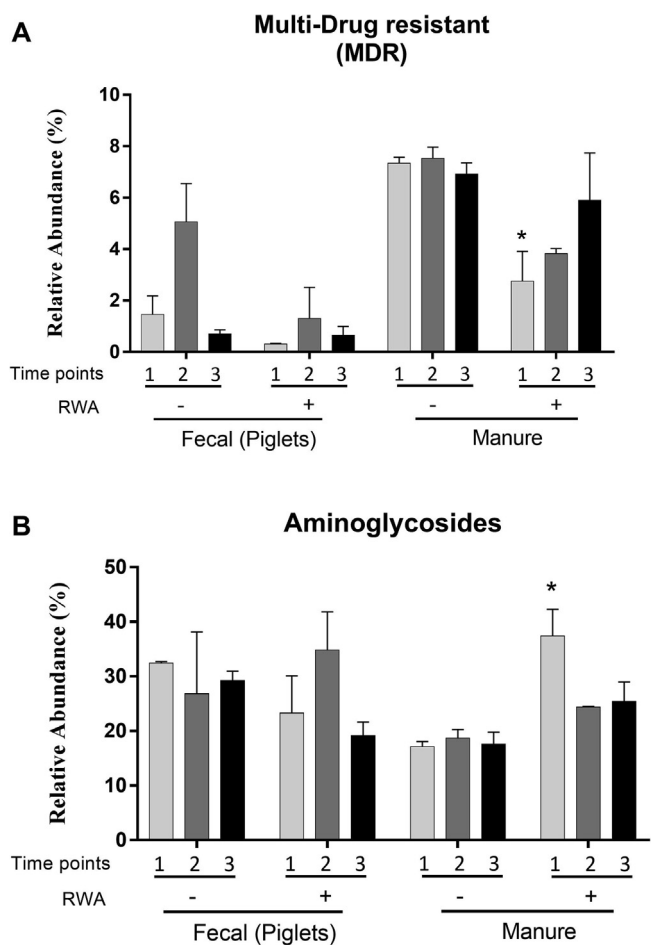

Fig. 7. Comparison of individual ARG classes of drug resistance between non-RWA and RWA in piglet feces and manure samples. (A). Comparisons of relative abundance of MDR-ARGs. (B). Comparison of relative abundance of Aminoglycoside-ARGs. ANOVA 2-way analysis with repeated measures (3-time points) comparing RWA vs. non-RWA. * $p < 0.05$.

The comparison of genetic sequence data from various samples obtained in this study provides valuable context to address the question of AMR in the animals, barn and surrounding environment. Individual hogs, and their health, are a key focus in our study and of paramount importance to the hog production continuum and thus nasal swabs provide a "value-added" screen for agents of respiratory disease that might not be well represented in fecal/environmental samples. Together, these different data sets can offer information about the status of AMR in the hog production systems, as well as early-warning data regarding pathogens of concern to hog and human health.

3.5. Additional readouts and potential correlations

In addition to resistomes, microbiome diversity, and the 'Path-omicrobiome' subset, our WGS analyses also generated 'Virulome' as well as 'Virome' data sets and a 'Respiratory Virome' subset (data not shown). Once abundance readouts are sorted into respective virulome and virome profiles, they could be used to query possible correlations with either the drug/vaccine data, clinical symptoms/illnesses, the resistome and pathogens. This would further generate, for example, possible relationships or co-occurrences of pathogens with virulence factors and/or bacteriophages that could eventually help differentiate between barns using an RWA approach and conventional barns. Further, comparative work on such data sets could help determine whether certain illnesses/symptoms are related to antibiotic classes or whether correlations exist between specific sets or patterns of resistance genes or pathogenicities.

Similar correlations have recently been tested between the resistome and the 'Virulome' in the human gut and in different combined environmental biomes [7]. The authors of this transversal comparative study used environmental and human gut microbes from distinct human populations from across the world to demonstrate that great metagenome protein family richness existed, greater than for the Resistome and Virulence diversities. Furthermore, that virulence and pathogenicity indeed correlated in and co-occurred across all types of samples (human gut and environmental biomes). However, these correlations concerned

the metagenomes richness (diversity) and not the relative abundance or frequency of each feature hits as we describe here. Indeed, the power of our designed workflow investigation lies in its potential for revealing greater insights into the effect of RWA with quantitatively measurable readouts and specific measurements.

Among the first 3 time points of the WGS analyzed thus far, we have observed a substantial reduction in the Tetracycline-ARGs class in RWA barns as compared to non-RWA barns. Since tetracyclines are among the drugs still heavily used in conventional barns, this suggests that RWA measures could possibly reduce resistance to tetracyclines. In addition, RWA barns also have a significantly reduced group of MDR-ARGs. This could be linked to lower amounts of various drug classes recorded as being used in RWA barns compared to conventional (non-RWA) barns, such as the Antifolates and β-lactams. Indeed, the MDR mechanism essentially uses the efflux-resistance strategy (efflux pumps that expel structurally-unrelated drugs), which include the Antifolate and β-lactams efflux transporters [21,22].

Our experimental approach would also allow comparisons and correlations of the type of animal samples (fecal vs. nasal swab), environmental samples (manure vs. soil) and animal age samples (piglet vs. sow), in addition to the longitudinal time-dependent effects.

4. Contribution to the field statement

Antibiotics are used in livestock to treat illnesses and thereby protect animal welfare. This use has been associated with the emergence, selection and spread of resistant bacteria. Next-generation sequencing (NGS) can be an alternative to phenotypic susceptibility testing for surveillance of the prevalence of antimicrobial resistance. Our investigative methodology also allows comparison of the effect of RWA and non-RWA pig production operations on overall drug consumption, pathogen loads, and the prevalence of antimicrobial resistance genes.

We described a comprehensive NGS-WGS workflow analysis that maps the reads directly to reference sequences which determined the taxonomy and resistome profiles as well as the pathogen prevalence in animals and their environment over time. Furthermore, additional metadata inputs from pig farms allowed a quantitative assessment of the effect of RWA on AMR. This methodological approach can be used to investigate the potential correlation between RWA practices and AMR to support the pig industry, as well as regulatory agencies in new efforts focused on reducing overall antibiotics use and in curbing the development and spread of AMR related to animal agriculture.

Author contributions statement

SMC, JRL, AA, BP and DRK contributed to the experimental design, the analyses and interpretation of the results, and the writing and editing of this manuscript.

Conflict of interest statement

Authors declare no Conflicts of Interest. SMC, AA and BP declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2020.09.012.

References

[1] Danner MC, Robertson A, Behrends V, Reiss J. Antibiotic pollution in surface fresh waters: occurrence and effects. Sci Total Environ 2019;664:793–804.

[2] ESVAC (2019). Sales of veterinary antimicrobial agents in 31 european countries in 2017: Trends from 2010 to 2017. https://www.ema.europa.eu/en, Centers for Disease Control and Prevention.

[3] PHAC (2016). Canadian antimicrobial resistance surveillance system report 2016. https://www.canada.ca/en/public-health/services/publications, Public Health Agency of Canada.

[4] Van Boeckel TP, Pires J, Silvester R, Zhao C, Song J, Criscuolo NG, et al. Global trends in antimicrobial resistance in animals in low- and middle-income countries. Science 2019;365(6459).

[5] Dibner JJ, Richards JD. Antibiotic growth promoters in agriculture: history and mode of action. Poult Sci 2005;84(4):634–43.

[6] Castanon JIR. History of the use of antibiotic as growth promoters in european poultry feeds. Poult Sci 2007;86(11):2466–71.

[7] Escudeiro P, Pothier J, Dionisio F, Nogueira T. Antibiotic resistance gene diversity and virulence gene diversity are correlated in human gut and environmental microbiomes. mSphere 2019;4(3).

[8] CIPARS (2015). The canadian integrated program for antimicrobial resistance surveillance (cipars) 2015 annual report. https://www.canada.ca/en/public-health/services/surveillance/canadian-integrated-program-antimicrobial-resistance-surveillance-cipars.html, Public Health Agency of Canada.

[9] PHAC. (2017). Responsible use of medically important antimicrobials in animals. from https://www.canada.ca/en/public-health/services/antibiotic-antimicrobial-resistance/animals/actions/responsible-use-antimicrobials.html.

[10] Zaheer R, Noyes N, Ortega Polo R, Cook SR, Marinier E, Van Domselaar G, et al. Impact of sequencing depth on the characterization of the microbiome and resistome. Sci Rep 2018;8(1):5890.

[11] CDC-BMBL (2018). Biosafety in microbiological and biomedical laboratories (bmbl) 5th edition. https://www.cdc.gov/labs/BMBL.html, Centers for Disease Control and Prevention.

[12] Hourigan SK, Subramanian P, Hasan NA, Ta A, Klein E, Chettout N, et al. Comparison of infant gut and skin microbiota, resistome and virulome between neonatal intensive care unit (nicu) environments. Front Microbiol 2018;9:1361.

[13] Al-Hebshi NN, Baraniya D, Chen T, Hill J, Puri S, Tellez M, et al. Metagenome sequencing-based strain-level and functional characterization of supragingival microbiome associated with dental caries in children. J Oral Microbiol 2019;11(1):1557986.

[14] Connelly S, Bristol JA, Hubert S, Subramanian P, Hasan NA, Colwell RR, et al. SYN-004 (ribaxamase), an oral beta-lactamase, mitigates antibiotic-mediated dysbiosis in a porcine gut microbiome model. J Appl Microbiol 2017;123(1):66–79.

[15] CosmosID. (2017). "The cosmosid bioinformatics platform (genbook database)." from cosmosid.com/blog-cosmosid/benchmarking-genome-biology-2017.

[16] Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, et al. Longitudinal analysis of microbial interaction between humans and the indoor environment. Science 2014;345(6200):1048–52.

[17] Clausen PT, Zankari E, Aarestrup FM, Lund O. Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole genome data. J Antimicrob Chemother 2016;71(9):2484–8.

[18] Junqueira ACM, Ratan A, Acerbi E, Drautz-Moses DI, Premkrishnan BNV, Costea PI, et al. The microbiomes of blowflies and houseflies as bacterial transmission reservoirs. Sci Rep 2017;7(1):16324.

[19] Yan Q, Wi YM, Thoendel MJ, Raval YS, Greenwood-Quaintance KE, Abdel MP, et al. Evaluation of the cosmosid bioinformatics platform for prosthetic joint-

associated sonicate fluid shotgun metagenomic data analysis. J Clin Microbiol 2019;57(2).

[20] CIPARS (2016). Canadian integrated program for antimicrobial resistance surveillance (cipars) annual report. https://www.canada.ca/en/public-health/services/surveillance/canadian-integrated-program-antimicrobial-resistance-surveillance-cipars/cipars-reports/2016-annual-report-summary.html, Public Health Agency of Canada.

[21] Assaraf YG. Molecular basis of antifolate resistance. Cancer Metastasis Rev 2007;26(1):153–81.

[22] Nikaido H, Pagès JM. Broad-specificity efflux pumps and their role in multidrug resistance of Gram-negative bacteria. FEMS Microbiol Rev 2012;36(2):340–63.

Glossary & definitions

AMR: Antimicrobial resistance

ARDB: Antibiotic Resistance Genes Database. Manually curated and unifying most of the publicly available information on antibiotic resistance

ARG: Antimicrobial resistance gene

ARG-ANNOT: Antibiotic Resistance Gene-ANNOTation. Detects existing and putative new ARGs in bacterial genomes

CARD: The Comprehensive Antibiotic Resistance Database. Compiles resistance genes, their products and associated phenotypes

CIPARS: Canadian Integrated Program for Antimicrobial Resistance Surveillance

DDBJ: The DNA Data Bank of Japan is a biological database that collects DNA sequences. A member of the International Nucleotide Sequence Database Collaboration or INSDC

DDDvetCA: Canadian Defined Daily Doses

ENA: The European Nucleotide Archive is a repository providing free and unrestricted access to annotated DNA and RNA sequences

Frequency: A kmer is a nucleotide sequence of a certain length. It is common in genomics to select all possible kmers of a fixed length for each read in a sample, for example. The number of unique kmer occurrences in the queried sample. This is roughly equivalent to the number of reads that matched to the organism identified

IMG: Integrated Microbial Genomes. A genome browsing and annotation platform with all the draft and complete microbial genomes sequenced by the DOE-JGI integrated with other publicly available genomes

Kmer: a kmer is a nucleotide sequence of a certain length. It is common in genomics to select all possible kmers of a fixed length for each read in a sample, for example

M5NR: non-redundant protein database (MD5nr) based on the use of MD5 checksums

MDR: Multi-drug resistance

MIAs: Medically Important Antimicrobials

MVIRDB: a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications

NCBI-RefSeq: NCBI Reference Sequences; The Reference Sequence database is open access, annotated and curated collection of publicly available nucleotide sequences and their protein products

PATRIC: the Pathosystems Resource Integration Center. Provides integrated data and analysis tools to support biomedical research on bacterial infectious diseases

PCoA: Principal Coordinate Analysis

PHAC: Health Agency of Canada

Relative Abundance: is calculated based on the number of organism-specific kmers and their observed frequency in the sample and then normalized to represent the abundance of each organism

ResFinder: Resistance finder is a database that captures antimicrobial resistance genes from whole-genome data sets. The database uses BLAST in order to accomplish this. The database allows inputs of full sequences, partial sequences, or short sequence reads from other sequencing platforms

RWA: raised without the use of antibiotics

SRA: Sequence Read Archive

VFDB: Virulence Factor Database. Provides scientist quick access to virulence factors in bacterial pathogens

WGS: Whole Genome Sequencing