Source tracking and global distribution of the mobilized tigecycline resistant gene *tet*(X)

Rong-min Zhang[1,2,3†], Jian Sun[1,2,3†], Ruan-yang Sun[1,2,3], Min-ge Wang[1,2,3], Chao-yue Cui[1,2,3], Liang-xing Fang[1,2,3], Mei-na Liao[1,2,3], Xiao-qing Lu[1,2,3], Yong-xin Liu[4], Xiao-Ping Liao[1,2,3]*, Ya-Hong Liu[1,2,3]*

[1]National Risk Assessment Laboratory for Antimicrobial Resistance of Animal Original Bacteria, South China Agricultural University, Guangzhou 510642, China.

[2]Laboratory of Veterinary Pharmacology, College of Veterinary Medicine, South China Agricultural University, Guangzhou 510642, China.

[3]Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, P. R. China

[4]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing, China.

National Risk Assessment Laboratory for Antimicrobial Resistance of Animal Original Bacteria, South China Agricultural University, Guangzhou 510642, China.

[†]These two authors contributed equally to this work.

*Address correspondence to Xiaoping-Liao and Yahong-Liu, E-mail: xpliao@scau.edu.cn and lyh@scau.edu.cn

Corresponding author at: College of Veterinary Medicine, South China Agricultural University, Guangzhou, Guangdong province, PR China.

Abstract

The emergence of *tet*(X) genes has compromised the clinical use of the last-line antibiotic tigecycline. We identified 322 (1.21%) *tet*(X) positive samples from 12,829 human microbiome samples distributed in four continents (Asia, Europe, North America and South America) using retrospective data from worldwide. These *tet*(X) genes were dominated by *tet*(X2)-like orthologs but we also identified 12 samples carrying novel *tet*(X) genes, designed *tet*(X15) and *tet*(X16), that were resistant to tigecycline. The metagenomic analysis revealed these *tet*(X) genes distributed in anaerobes dominated by *Bacteroidaceae* (78.89%) of human-gut origin. The transmission of these *tet*(X2)-like orthologs between *Bacteroidaceae* and *Riemerella anatipestifer* was primarily promoted by the mobile elements IS*Bf11* and IS*4351*. *tet*(X2)-like orthologs was also developed during transmission by mutation to high-level tigecycline resistant determinants *tet*(X15) and *tet*(X16). Further tracing these *tet*(X) in single bacterial isolate from public repository indicated that *tet*(X) genes were present as early as 1960s in *R. anatipestifer* that was the primary *tet*(X) carrier at early stage (before 2000). The *tet*(X2) and non-*tet*(X2) orthologs were primarily distributed in humans and food animals respectively, and non-*tet*(X2) were dominated by *tet*(X3) and *tet*(X4). Genomic comparison indicated these *tet*(X) genes were likely to be generated during *tet*(X) transmission between *Flavobacteriaceae* and *E. coli*/*Acinetobacter* spp.., and IS*CR2* played a key role in the transmission. These results suggest *R. anatipestifer* was the potential ancestral source of *tet*(X) gene. Additionally, *Bacteroidaceae* of human-gut origin was an important hidden reservoir and mutational incubator for the mobile *tet*(X) genes that enabled spread to facultative anaerobes and aerobes.

Keywords: Tigecycline resistance, *Tet*(X), Source tracking, Human microbiome, *Riemerella*

anatipestifer, *Bacteroidaceae*

Introduction

The first generation of tetracycline antibiotics consisted of tetracycline, chlortetracycline and oxytetracycline and were put into clinical practice in 1952 (1) while the second generation derivatives doxycycline and minocycline were put into use in 1976 (2). These antibiotics have been incorporated into animal feed to improve growth and feed efficiency (3). However, bacterial resistance to the tetracyclines was observed from the very beginning of their usage. To date, more than 65 specific resistant determinants and 9 MDR efflux pump genes of the root nodulation-division (RND) superfamily have been confirmed including AdeABC, AcrAB-TolC and MexAB-OprM (2). These determinants confer resistance to first and second generation tetracyclines and are widely distributed among 130 Gram-negative and Gram-positive bacteria (2).

A third-generation tetracycline (tigecycline) was approved in the United States in 2005 and its use in the EU and China was authorized in 2006 and 2010, respectively (4, 5). Tigecycline has a robust treatment range and includes bacteria resistant to first and second generation tetracyclines (6), and is a 'last resort' antibiotic used to treat severe infections caused by carbapenem- and colistin-resistant pathogens (5). Thus, this antibiotic was classified as a critically important antimicrobial by the World Health Organization and its usage is restricted (7). However as early as 1884, the transferable gene *tet*(X) displaying tigecycline insusceptibility was discovered on an R plasmid from a *B. fragilis* isolate of human origin. This was the earliest occurrence of an antibiotic resistance gene (ARG) that directly inactivated tetracyclines (8). The *tet*(X) gene was only functional under aerobic growth conditions because it is a flavin dependent monooxygenase that requires FAD, NADPH, Mg^{2+} and O_2 to inactive almost all of the tetracycline class (9). In 2001

the existence of *tet*(X1) and *tet*(X2) were confirmed on a transposon from *Bacteroides thetaiotaomicron* of human origin and shared 61.7 and 99.5 % amino acid identity with *tet*(X), respectively. To date, *tet*(X)/*tet*(X2) genes have already spread to 16 countries/regions covering five continents (Asia, Europe, North America, South America and Africa) (2). A small comfort was that these ARGs displayed low level tigecycline resistance (MIC $\leq$ 2ug.ml^{-1}) (2).

The emergence of plasmid-mediated high-level tigecycline resistance encoded by the *tet*(X3) and *tet*(X4) genes in 2019 posed a severe threat to public health (7, 10). Additionally, 10 more *tet*(X) orthologs have been identified and include *tet*(X5) - *tet*(X14)) (11-14). These orthologs were primarily found in food animals especially swine, including *tet*(X3), *tet*(X4), *tet*(X6) and *tet*(X14) firstly detected in *Acinetobacter baumanii*, *Escherichia coli*, *Myroides phaeus* and *Empedobacter stercoris*, respectively. The *tet*(3.2) and *tet*(X5) gene were identified from an *Empedobacter brevis* isolate of shrimp origin and a *Acinetobacter baumannii* isolate of human origin, respectively. All the *tet*(X7) - *tet*(X13) orthologs were identified directly from gut-derived metagenomic libraries, but their host bacteria were unknown. Epidemiological studies (2, 15, 16) indicated the dissemination of these *tet*(X) orthologs were dominated by *tet*(X3) and *tet*(X4) that were primarily detected from *Acinetobacter* spp. and *E. coli*, respectively. Furthermore, *tet*(X3)/*tet*(X4) samples from humans, animals and meat for consumption revealed a prevalence of 0.3 - 66.7 % and the highest level of 66.7 % was detected from pig caecum samples from abattoirs (7). Compared with the *tet*(X3)/*tet*(X4) in animal isolates (6.9%, 73/1060) (7), lower prevalence from human (0.32%, 4/1250) were observed in a retrospective screening of *tet*(X)-carrying clinical isolates (5).

The *tet*(X) genes have been primarily identified using traditional cultural methods and this imposed limitations on their identification including the loss of uncultured bacteria, low-throughput

and long processing times. Although the *tet*(X7) - *tet*(X13) orthologs were found directly from gut-associated samples using metagenomic sequencing, the bacterial hosts, relative abundance and propagative characteristics were absent (14). Nonetheless, public repositories are a promising high-throughput resource for exploring antibiotic resistomes. For instance, a retrospective epidemiological study based on the available public bacterial gene datasets revealed that the food chain was a potential dissemination pathway for *mcr-1* (17). Additionally, a metagenomic screening study based on public metagenome datasets revealed a high detection rate of *tet*(X3) (25.4 %) in poultry samples (18). However, there are few studies that utilize data mining for *tet*(X) in public databases (18, 19). In addition, public data repositories including GenBank are a valuable resource for the exploration of novel bacterial species. For instance, a recent study utilized 9,428 metagenomes to reconstruct 154,723 microbial genome bins that generated 4,390 species level genome bins (SGB) including 77 % of which were not present in public repositories (20). Identification of *tet*(X) genes from these SGBs and tracing their distribution in assembly isolates from public repository may provide a new perspective for source tracking of the global spread of *tet*(X).

In the current study, we utilized these data mining techniques and discovered that *tet*(X) had emerged as early as 1960 and the *Riemerella anatipestifer* was its potential ancestral source. Additionally, *Bacteroidaceae* of human gut origin were a hidden reservoir and mutational incubator for mobile *tet*(X) genes that enabled spread to facultative anaerobes and aerobes.

Methods

Collection of microbial genomic sequences from human microbiome in retrospective data

A total of 26,728 metagenomic samples of human-microbiome origin deposited in public repositories were downloaded and reconstructed microbial genomic bins (MGBs) to explore new

bacterial species in previous studies (20-22). We removed the duplicative samples from these metagenomic samples according their accession number and found a total of 12,829 non-duplicate metagenomic samples from 31 countries. These samples were reconstructed into 202,265 MGBs in previous studies (20-22) (Table S1 and S2) and the online released data were screened for the presence of all known *tet*(X) orthologs using BLAST using an 80 % identity and 70 % hit length cut-off in current study. The prevalence of *tet*(X) in 31 countries were plotted using R version 3.5.3. Phylogenetic analysis for amino acid sequences of all Tet(X) gene products was constructed using neighbor joining with the default parameters in Mega X Version 10.0.5 (23) and alignments were constructed using ESPript 3 (24).

Functional identification of Tet(X)s

Tigecycline resistance for these gene products was assessed by synthesis of full-length nucleotide sequences of all detected *tet*(X) genes. *EcoR*I and a *Sal*I sites were then added 5′ and 3′ respectively (Tsingke Biological Technology, Beijing, China). The synthesized *tet*(X) genes were cloned into plasmid vector pBAD24 and transformed into competent *E. coli* JM109 as described in our previous study (25). The transconjugants *E. coli* JM109+pBAD24-*tet*(X4) and *E. coli* JM109+pBAD24, were used as positive and negative controls, respectively, as previously described (25). The minimum inhibitory concentration (MIC) for tetracycline, doxycycline, minocycline, tigecycline, eravacycline and omadacycline were determined by the broth microdilution method in accordance with Clinical and Laboratory Standards Institute (CLSI) guidelines. Tetracycline, doxycycline and minocycline breakpoints were interpreted according to the European Committee on Antimicrobial Susceptibility Testing (EUCAST) guidelines (http://www.eucast.org/clinical_breakpoints). The United States FDA criteria was employed to

interpret tigecycline breakpoints for *E. coli* and MIC ≥4 mg L^{-1} was considered non-susceptible while eravacycline and omadacycline were uninterpreted with no breakpoint. *E. coli* ATCC 25922 was used as the quality control strain.

Taxonomic assignment and phylogenetic analysis of *tet*(X)s-carrying MGBs

We obtained 322 *tet*(X)-carrying MGBs from three previous studies (20-22). This group included taxonomic assignments for 196 that had been previously annotated (20) while the remaining 126 were annotated using metaWRAP-Annotate-bins module using the MetaWRAP pipeline and default parameters (26). Briefly, the assembly contigs from each *tet*(X)-carrying MGB was taxonomically profiled using Kraken2 (27) and then this entire metagenomic bin could conservatively and accurately estimate the taxonomic profiles (26).

The phylogenetic structure for the *tet*(X)-carrying MGBs were performed using an automatic PhyloPhlAn (3.0) pipeline (20, 28), through which the phylogeny in Figure 1b was built using 400 universal PhyloPhlAn markers with parameter: "--diversity high --accurate --min_num_markers 80." This pipeline integrates diamond (version 0.9.32) , mafft (version 7.464) (29), trimal (version 1.4.rev15) (30) and RAxML (version 8.2.12) (31), and the parameters of these software were set as described previously (20). The phylogenetic tress in Figure 1b were plotted using GraPhlan (version 1.1.3) (32).

Phylogenetic analyses of the *tet*(X) carrying isolates and evolutionary timescale for the *tet*(X)s from isolates

To further trace the spread of all *tet*(X)s in culturable bacteria isolates, a total of 774,435 bacteria assembled whole genome sequences (WGS) were downloaded from the NCBI database as of November 7th, 2020. *tet*(X)-like OFRs were determined using BLATX against all the *tet*(X)s

variants mentioned in current study with a minimum similarity of 70% and 100% coverage. The collection date, origin, countries and the bacterial host of the *tet*(X)-positive isolates were retrieved according to their Biosample Number. The phylogenetic structure for the *tet*(X)-carrying isolates were also performed using PhyloPhlAn (3.0) pipeline mentioned above.

To determine the evolutionary history of *tet*(X)s, the earliest emergency of *tet*(X) variants (with collection date) with a 388 amino acid (aa) length were applied to generate a chronogram using Bayesian evolutionary analysis version 1.10 (33). For all model combinations, three independent chains of 100 million generations each were run to ensure convergence with sampling every 1,000 iterations. Tracer v1.7.1 was used to assess convergence using all parameter effective sampling sizes of > 200 (34). LogCombiner v2.6.1 was used to combine tree files and a maximum clade credibility tree was created using TreeAnnotator v2.6.0 (34). Tree annotation was visualized using iTOL (35) and FigTree version 1.4.2.

Annotation and comparison of the genomic region flanking the *tet*(X) gene

The *tet*(X)-carrying contig were extracted from the MGBs of metagenomic analysis and isolates from public repository. CD-HIT was employed to group *tet*(X)-carrying full length contigs using a cutoff with a minimum similarity of 97 % over 97 % of the query coverage (36). These *tet*(X)-carrying contigs were annotated using Prokka (37) and in conjunction with standalone BLAST analyses against the ResFinder (38) and ISfinder (39) databases to cross-validate ARGs and mobile genetic elements, respectively.

Results

Identification of Tet(X) orthologs

We accessed a total of 202,265 MGBs constructed from 12,829 metagenomic samples of

179 human-microbiome origin and found 322 (1.21 %) of them encoded 388 aa proteins with 96.13 -
180 100% similarity with Tet(X) orthologs reported in a previous study (11) (Table S2). In particular,
181 96.27 % (310/322) of them shared 98.71 - 100 % identity with Tet(X2) (Acc. No. AJ311171) and
182 phylogenetically could be grouped into 15 *tet*(X2)-like clades (Figure S1 and S2, Table S3). The
183 remining 3.73 % (12/322) shared < 98.20 % identical with the known Tet(X). Since there was not
184 a criterion for assignment of Tet(X) orthologs in previous studies (11, 14), we temporarily
185 designated these *tet*(X2)-like orthologs as *tet*(X2.2) to *tet*(X2.16). These orthologs represented a
186 large proportion of samples (n=12,829) from 19 countries and were therefore likely represented a
187 relatively comprehensive assessment of *tet*(X2)-like orthologs. We therefore tentatively used the
188 lowest cutoff of 98.20 % between *tet*(X2)-like orthologs (*tet*(X2.15) *vs tet*(X2.16)) for assignment
189 of novel *tet*(X) orthologs. Based on this, we found two new *tet*(X) orthologs and their subtypes
190 shared less than 98.20% amino acid identity with their closest neighbors and most resembled *tet*(X)
191 orthologs in the phylogeny designed *tet*(X15), *tet*(X15.2), *tet*(X15.3), *tet*(X16), *tet*(X16.2) and
192 *tet*(X16.3) (Figure S1, Table S3). Most of these *tet*(X) orthologs found from metagenomic analysis
193 were not present in the NCBI database with the exceptions of *tet*(X2) (Acc. No. AJ311171),
194 *tet*(X2.4) (Acc. No. JQ990987) and *tet*(X16.2) (Acc. No. KU547718.1) (Table S3).

195 **Resistance phenotypes of *tet*(X) orthologs**

196 Both the *tet*(X15) and *tet*(X16) groups found from metagenomic analysis of human-gut origin
197 (Figure 1c) in the *E. coli* JM109 were resistant to tigecycline (MICs 8 to 16 ml/liter), and
198 exhibited high MIC to the fourth-generation antibiotics omadacycline (MICs 16 to 32 ml/liter) and
199 eravacycline (MICs 2 to 4 ml/liter) (Table S4). In addition, all the non-*tet*(X2) exhibited resistance
200 to tetracycline (MICs 128 to 256 ml/liter), doxycycline (MICs 32 to 64 ml/liter) and minocycline

(MICs 16 to 32 ml/liter) (Table S4). All the *tet*(X2) like orthologs were susceptible to tigecycline (MICs 0.25 to 1 ml/liter) but resistant to tetracycline (MICs 8 to 256/liter). Almost all of them were resistant to doxycycline (MICs 8 to 64 ml/liter) and minocycline (MICs 2 to 32 ml/liter), excluding *tet*(X2.10) that intermediate to doxycycline (1 ml/liter) and minocycline (MICs 0.5 ml/liter). Most of these *tet*(X2) like orthologs exhibited MIC of omadacycline from 0.5 to 4 ml/liter, and eravacycline from 0.0125 to 1 ml/liter (Table S4). Three *tet*(X2) like orthologs *tet*(X2), *tet*(X2.4) and *tet*(X2.15) showed relatively higher MICs to omadacycline (8 to 16 ml/liter) and eravacycline (1 to 4 ml/liter).

Global distribution and taxonomic assignment of *tet*(X) carrying MGBs

The 322 *tet*(X)s carrying MGBs were detected in 19 countries including Europe (n=12), Asia (n=4) and America (n=3) but were absent in Oceanica and Africa (Figure 1a, Table S5). The prevalence of *tet*(X) in European countries (0 - 9.40 %) was more complex than Asian (0 - 3.17 %) and American countries (0 - 2.44 %) although the positive rates for *tet*(X) in these three continents were not significantly (P>0.05) different (Figure S3). The prevalence of *tet*(X) was highest in Spain (9.40 %), followed by Germany (5.14 %), France (3.82 %), Denmark (3.34 %) and China (3.17 %) and the remaining countries were < 3 % (Figure 1a and Table S5).

The bacterial taxonomy assignment indicated that all the 322 *tet*(X)-carrying MGB could be assigned to five phyla and was dominated by *Bacteroidetes* (78.89 %, 254/322) followed by *Firmicutes* (16.78 %, 47/322), *Proteobacteria* (1.24 %, 4/322), *Candidatus Melainabacteria* (1.24%, 4/322) and *Fusobacteria* (0.62%, 2/322) (Figure 1b). Furthermore, 96.68% (301/322) were classified to the family level and most of them were also belonging to *Bacteroidaceae* (70.10%, 211/301) (Figure 1b and Table S6). Furthermore, 72.30% (154/211) of these *tet*(X)

carrying *Bacteroidaceae* MGBs could be further divided into 14 *Bacteroides* species that were dominated by *Bacteroides vulgatus* (37.67 %, 58/154), *Bacteroides uniformis* (20.78 %, 32/154), *Bacteroides dorei* (14.94 %, 23/154), *Bacteroides ovatus* (5.84 %, 9/154), *Bacteroides fragilis* (3.90 %, 6/154) and *Bacteroides caccae* (3.25 %, 5/154) (Table S7).

The *tet*(X2)-like positive MGBs were distributed across these 19 *tet*(X) positive countries. *tet*(X2.4) (55.02 %, 170/309), *tet*(X2) (26.86 %, 83/309), *tet*(X2.5) (9.32 %, 30/309) and *tet*(X2.2) (4.53 %, 14/309) totaled 96.11 % (297/309) and prevailed over other *tet*(X2)-like orthologs (≤ 0.6 %, 2/309) (Figure S4). Only 81.08% (240/296) of these four predominant *tet*(X) carrying MGBs could be assigned at genus level and most of them were also identified in *Bacteroides* (71.67%,172/240), followed by *Prevotella* (6.67%, 16/240) and *Alistipes* (5%, 12/240) (Table S8). Additionally, the non-*tet*(X2) orthologs included 75 % (9/12) that were distributed in China and USA. Of which, *tet*(X15.2) was the most prevalent ortholog (Figure 1c). The remaining three non-*tet*(X2) carrying MGBs were distributed in Europe and the single *tet*(X16.2) ortholog was carried by an *Enterococcaceae* from Italy. Almost all *tet*(X)-carrying MGBs were collected from human stools excluding only one *tet*(X2.4)-carrying MGB identified as an *Enterococcus* spp. from an oral cavity sample (Table S8).

The 322 *tet*(X) carrying MGBs included 196 from the 4,390 MGBs and their average abundance in human microbiomes has been calculated in previous study (Table S8) (20). The average abundance of these 196 *tet*(X)-positive MGBs was 5.97 ± 3.89 and significantly higher than that of the total 4,390 MGBs (1.76 ± 3.74) (20).

Culturable isolates from public repository insight into the distribution and evolutionary timescale of Tet(X)s

To further trace the distribution of these *tet*(X)s in culturable isolates, we examined 774,435 WGS of bacterial isolates present in GenBank and only 0.12 % (896/774,435) carried an ORF with > 70 % amino acid identity with the known *tet*(X) genes including the novel ones found in the current study (Table S9). The PhyloPhlAn analysis indicated that the facultative anaerobe clade was phylogenetically distinct between anaerobes and aerobes (Figure 2a). These *tet*(X) genes were found in 17 bacterial families that were dominated by aerobes including *Moraxellaceae* (279/896, 31.17 %), *Enterobacteriaceae* (208/896, 23.24 %) and *Weeksellaceae* (16.20 %, 145/896). The *Bacteroidaceae* (20.67 %, 185/896) were also an important anaerobic carrier for *tet*(X2) like orthologs (Figure 2a and Table S9). Three *tet*(X) orthologs were most prevalent and included *tet*(X2)-like (35.71 %, 320/896), *tet*(X3) (27.34 %, 245/896) and *tet*(X4) (21.99 %, 199/896) (Table S9). Interestingly, different bacterial families from variant hosts were preference for carrying specific *tet*(X) ortholog (Figure 2a). Almost all the *Bacteroidaceae* (98.91 %, 182/184) and most *Weeksellaceae* (63.45 %, 92/145) isolates carried *tet*(X2)-like orthologs and were primarily from human samples (38.64 %, 114/295). Additionally, all *tet*(X3) were detected from *Acinetobacter* spp. and almost all *tet*(X4) (88.89 %, 176/197) were carried by *E. coli* and these were primarily from food animals (66.59 %, 295/443) including pigs, chickens, ducks, cattle and geese.

The *tet*(X2)-like orthologs were detected prior to 8.41 ± 6.17 years ago and earlier than *tet*(X3) (4.00 ± 1.12) and *tet*(X4) (2.38 ± 1.34). To be noted, the collection dates of three *R. anatipestifer* isolates from UK duck samples were prior to 1980. One isolate (BioSample: SAMN09912225) was collected in 1966 and was positive for *tet*(X12) and a *tet*(X2)-like gene that only differed in a single amino acid from *tet*(X2) was designed *tet*(X2.17). The two other isolates

were collected in 1976. One (BioSample: SAMN09912224) carried a *tet*(X2.17) gene and another (BioSample: SAMN09912221) carried a *tet*(X) gene shared 98.20% similarity with *tet*(X16), designed *tet*(X16.4).

Date for lineage divergence of the earliest occurred *tet*(X) orthologs were produced by Bayesian phylogenetic inference (Figure 2b). The analysis indicated a mean rate of 0.28 SNP per year for these *tet*(X) during 1966-2018. The most-recent common ancestor (MRCA) of all *tet*(X) from this study was approximately from 1939 to1965, and the tracer analysis revealed the presence of *tet*(X) most likely occurred in AD 1953 (95% highest posterior distributor (HPD) AD). Two main lineages that originated from *tet*(X2.17) and *tet*(X12), respectively, were observed in this phylogeny and both *tet*(X) orthologs were collected from the *Riemerella anatipestifer* of duck origin from the UK in 1966 year.

Annotation and comparison of the *tet*(X) genomic environment

A total of 1218 *tet*(X)-carrying contigs ranging from 1,190 to 931,600 bp were retrieved from the metagenome and bacterial-isolate data. These contigs were grouped into 455 clusters that carried a range of 1 - 48 contigs (Table S10). In each cluster, longer contig shared more than 97% coverage and more than 97% similarity with shorter contig. The high coverage and similarity of these contigs indicated that these *tet*(X) could spread among each cluster (Table 10). This indicated *tet*(X) orthologs could spread among a great diversity of host including human, animal and environment (Figure 3). Both humans and pigs were the primary *tet*(X) hosts. *tet*(X2)-like *tet*(X3), *tet*(X4), *tet*(X7) and *tet*(X16.3) have been found in humans as well as *tet*(X2)-like, *tet*(X3), *tet*(X4) and *tet*(X6) in pigs. However, only *tet*(X2)-like and *tet*(X3) orthologs could transfer between these two hosts (Figure 3a). Interestingly, *tet*(X2)-like orthologs could hitch a great

289 diversity of vehicles to spread between humans and pigs and these included *Bacteroides* spp., *R.*
290 *anatipestifer* and *Chryseobacterium* spp.. The remaining *tet*(X) genes were spread only *via* special
291 species between different hosts. For instance, the *tet*(X3) gene could only be transited by
292 *Acinetobacter spp*. and spread between pigs and other hosts including pigeons, cattle, geese, ducks
293 and humans (Figure 3a). Additionally, the *tet*(X4) in the genomic array *rdm*C-*tet*(X4)-△IS*CR2*
294 could spread among wild birds, humans, pigs, chickens and the environment (Figure 3b).
295 Genomic annotation and comparisons indicated that other ARGs frequently flanked the *tet*(X)
296 gene and these included *erm*F, *aad*K, *tet*(Q) and bla_{OXA347} that conferred resistance to erythrocin,
297 streptomycin, tetracycline and ampicillin, respectively (Figure 4a and 4b). Of which, the *erm*(F)
298 gene was the most frequently flanking the *tet*(X) gene (n=152) and their genomic environment
299 were clustered into two types according to their relative position: *erm*F located upstream (15.79 %,
300 24/152) and downstream (80.92 %, 123/152) of *tet*(X) gene (Figure 4a and 4b). The upstream
301 *erm*F always formed a conserved structure *tnpF*-*erm*F-*tet*(X1)-*tet*(X2)/*tet*(X2.2)-*aad*K (n=24)
302 (Figure 4a). This structure was also present in a conjugative transposon CTnDOT of *Bacteroides*
303 origin (40), but the *aad*K (930 bp) was replaced by another aminoglycoside ARG *aad*S (903 bp) in
304 CTnDOT and these two ARGs shared a 96.77 % identity at the nucleotide level. Interestingly, the
305 *tet*(X) orthologs and their genomic contexts were more diverse when *erm*F was located
306 immediately downstream of *tet*(X). These *tet*(X) genes included 12 orthologs that were dominated
307 by *tet*(X2.4) and *tet*(X2.5) (Figure 4b, S1 text A1). In addition, these *tet*(X) genes were able to
308 spread among anaerobes, aerobes and facultative anaerobes (text S1).

309 **Discussion and conclusion**

310 **Source tracking of the *tet*(X) orthologs**

One of the most significant findings in current study was that *tet*(X) emerged as early as 1960s in *R. anatipestifer* of duck origin, which was earlier than that had been previously reported in 1980s (8) but we found *tet*(X) was present. A previous epidemiological study indicated the MRCA of these *tet*(X)s orthologs likely occurred 9900 years ago (7887 BC) (15). However, we also estimated the MRCA of these *tet*(X) genes but determined it occurred only 68 years ago (1953 AD) that was only one year following the introduction of tetracycline in clinics. This was consistent with the streptomycin, erythrocin and florfenicol, in which their resistance also emerged over the ensuing years following their introduction in clinic [4]. In this analysis, we screened for the presence of *tet*(X) orthologs from public repositories on a large scale (774,435 isolates) and the earliest emergence of 28 *tet*(X) variants was selected to perform MRCA estimations (Figure 2b). The time of the earliest occurrences of *tet*(X) were an important factor for the chronogram phylogeny construction using the Bayesian evolutionary analysis. Thus, these data should be validated to confirm these results.

The *Flavobacteriaceae* have been recognized as a potential ancestral source of the tigecycline resistance gene *tet*(X) (5). We found that both the two earliest (1966) occurring *tet*(X) orthologs (*tet*(X2.17) and *tet*(X12)) harbored by *R. anatipestifer* that is also a bacterial species member belonging to *Flavobacteriaceae* family. Meanwhile, before 2000 only a total of seven isolates were confirmed as *tet*(X) gene carrier and five of them were also identified as *R. anatipestifer* (Table S9). Among the 896 *tet*(X)-carrying isolates from public repository, 71 *tet*(X)-carrying were *Flavobacteriaceae* including *R. anatipestifer* that accounted for a large number (64.79%, 46/71) versus all other *Flavobacteriaceae* members (Figure 2a). Additionally, *R. anatipestifer* harbored a great diversity of *tet*(X) variants and only 23.94 % (17/71) carried the

known *tet*(X)s including *tet*(X12), *tet*(X14), *tet*(X2.17) and *tet*(X16.4) (Figure 2a). The remaining carried other *tet*(X) orthologs that shared 94.9 - 99.7 % similarity with their most closely related *tet*(X) ortholog (Table S9). A recent study indicated the poultry pathogen *R. anatipestifer* appears to be a reservoir for Tet(X) tigecycline resistance (41).These indicated that *R. anatipestifer* was most likely the ancestral source of the tigecycline resistance gene *tet*(X).

A great diversity for *tet*(X) and their flaking genomic contexts was observed when the *erm*F gene was present downstream of *tet*(X) (Figure 4b and text S1). A comparison of the *tet*(X) genomic contexts from MGBs and culturable isolates yielded 11 genomic backbones that were associated with the formation of non-*tet*(X2) orthologs found in the current study (Figure 4b). The non-*tet*(X2) orthologs were likely to be generated from *tet*(X2)-like orthologs during their transmission between anaerobes or between anaerobes and facultative anaerobes (Figure 4b and text S1). IS*Bf11* and IS*4351* played important roles in their transmission between anaerobes that was dominated by *Bacteroides* spp. where a mobile cyclic structure was speculated based on genomic Types I - IV (Figure 4b and text S1), and the *tet*(X15)-like and *tet*(X16)-like genes were likely to be generated during the transmission of this mobile structure. In addition, a self-transmissible cyclic structure lacking IS elements was also hypothesized based on the genomic Types V - XI. This structure likely transferred between *Bacteroides* spp. and *R. anatipestifer* and *tet*(X16.4) was most likely generated during the transmission. In addition, genomic comparisons indicated that these *tet*(X) genes were also able to spread between *Flavobacteriaceae* and *E. coli* as well as between *Flavobacteriaceae* and *Acinetobacter* sp. (Figure 4c and text S1). We have demonstrated that *R. anatipestifer*, a *Flavobacteriaceae* family member, was a potential ancestral source of *tet*(X) and the new *tet*(X) orthologs were likely to be

355 produced during their transmission. The high similarity of the nucleotide sequences flanking
356 *tet*(X3) and *tet*(X4) (Figure 4c and text S1) suggested that these two genes were also derived from
357 *Flavobacteriaceae* and IS*CR2* played a key role in this process (Figure 4c and text S1).

358 **Global distribution of *tet*(X) orthologs**

359 The human microbiome plays an important role in public health. Here, we first determine the
360 *tet*(X) prevalence in the human microbiome using a large-scale survey of 12,829 samples (20-22).
361 A total of 16 *tet*(X2)-like and two new non-*tet*(X2) orthologs have been identified directly in the
362 human stool samples. Since there was not standard for assignment of the new found *tet*(X)
363 orthologs, and it was necessary to distinguish the different *tet*(X) orthologs in current study. Thus,
364 we temporarily set up a criterion for the assignment of *tet*(X) orthologs. This maybe not
365 comprehensive as the assignment for mobile colistin resistance (*mcr*) genes (42) which have
366 established a platform in NCBI (pd-help@ncbi.nlm.nih.gov) to confirm and allocate the allele
367 numbers for new *mcr* gene. An allele numbers assignment for *tet*(X)s should be established
368 urgently.

369 We found a prevalence for *tet*(X) at 1.21 % (322/26,548) that was higher than for *E. coli* and
370 *K. pneumoniae* from hospital isolates (0.32%, 4/1520) (5) indicating that traditional culture
371 methods have underestimated the prevalence of *tet*(X). Our results were similar to an
372 epidemiological study that detected the bla_{NDM} and *mcr-1* genes directly from samples that was
373 higher than for the *E. coli* isolates (43).

374 The *tet*(X2)-like genes carrier in human microbiomes were dominated by the *Bacteroidaceae*
375 in contrast to previous epidemiological studies where *tet*(X3) and *tet*(X4) were primarily carried
376 by *A. baumannii* and *E. coli*, respectively (2, 15). The *Bacteroides* are predominant anaerobes

377 estimated to account for 25 - 30% of human gut microflora (44) while the *Enterobacteriaceae*
378 normally constitutes only 0.1 - 1 % (45). We also found that the average abundance of
379 *tet*(X)-carrying MGBs (5.97 ± 3.89) annotated as *Bacteroidaceae* prevailed over species-level
380 genome bins (1.76 ± 3.74) in the human microbiome. This was likely the reason for the absence of
381 *tet*(X3) and *tet*(X4) in our human microbiome analyses. Although *tet*(X) genes are inactive in
382 anaerobes, the high abundance of *tet*(X2)-carrying MGBs and a variety of non-*tet*(X2)-like
383 orthologs found in the current study indicated that the *Bacteroidaceae* were an important reservoir
384 and mutational incubator for the mobile *tet*(X) orthologs in the human microbiome (Figure 5).
385 Furthermore, the *Bacteroidaceae* could generate new non-*tet*(X2) orthologs with tigecycline
386 inactivation functions, and the comparison of the *tet*(X) genomic environment suggested that these
387 non-*tet*(X2) enabled transfer to facultative anaerobes and aerobes (Figure 5).

388 We have demonstrated *R. anatipestifer* as potential ancestral source of *tet*(X) genes, and the
389 earliest emergence of high-level tigecycline resistance genes *tet*(X3) and *tet*(X4) were likely to be
390 1971 and 1982, respectively, both of which were earlier than the clinical introduction of
391 tigecycline in 2005. This was further evidence that the use of even older antibiotic tetracycline
392 may contribute to the resistance to newer antibiotics (4). The *tet*(X) distributions from culturable
393 isolates indicated that *tet*(X2)-like and non-*tet*(X2) orthologs were prevalent in anaerobes and
394 aerobes respectively (Figure 5). Since Tet(X) is active only in an aerobic environment, the
395 non-*tet*(X2)-like orthologs with tigecycline inactivate function tended to be captured by aerobes
396 under tetracycline selective pressure (Figure 5). This was likely to be the reason that non *tet*(X2)
397 orthologs primarily distributed in aerobes but the high prevalence of *tet*(X2)-like orthologs in
398 *Bacteroidaceae* from human microbiome need to be further explored.

Taxonomic assignments for *tet*(X)-carrying MGBs were estimated using bioinformatic methods that may not be as precise as cultural methods but such an approach has proved feasible (20, 46). Another limitation of this approach was challenging to combine the chromosome with their respective plasmid sequences (47). Therefore, these *tet*(X)-carrying contexts from MGBs in the current study were likely to be chromosome-borne. This differed with *tet*(X3)/*tet*(X4) that are present in a variety of plasmids and IS*CR2* was an essential element for their mobilization (2, 15). Transmission between *Bacteroidaceae* of these *tet*(X)s orthologs was primarily mediated by the CTnDOT-like conjugative transposon and *erm*F-related IS elements including IS*Bf11* and IS*4351*. CTnDOT has been reported to harbor *erm*F, *tet*(X1) and *tet*(X2) in *Bacteroides* (40) and conjugative transposons can also insert into co-resident plasmids in addition to the chromosome (48). Therefore, conjugative transposons have been found in numerous genera including *Enterococcus*, *Streptococcus*., *Lactococcus*, *Butyrivibrio*, *Clostridium*, *Salmonella*, *Pseudomonas*, *Mezorhizobium* and *Vibrio* (40). The erythromycin resistance *erm*F gene is frequently reported in *R. anatipestifer* and *Bacteroides* spp. isolates (49, 50). The IS *Bf11* and IS*4351* flanking *tet*(X) in the Type III genomic contexts have also been previously identified (51) and reveals that this mobile structure has spread to China, the USA, France, Denmark, Sweden and Belgium. In addition, the IS*CR2* element belonging to the IS*91* family has been described in the first report of *tet*(X3) and *tet*(X4), both downstream and upstream of *tet*(X3). This IS element could form a mobile amplicon and this was demonstrated using inverse PCR experiments (7) and in an *Acinetobacter towneri* isolate flanking region of *tet*(X6) (52). We also identified this IS element located upstream of *tet*(X6) and indicated that this IS element can play an important role in transmission of non-*tet*(X2) orthologs with tigecycline inactivation functions in the

421 *Flavobacteriaceae*, *E. coli* and *Acinetobacter*.

422 In conclusion, we conducted an analysis integrated human gut metagenome and global
423 bacterial isolates to trace the origin and distribution of *tet*(X) gene. The *tet*(X) gene emerged as
424 early as 1960 and the *R. anatipestifer* was an ancestral source of *tet*(X). The *tet*(X3)-carrying
425 *Acinetobacter* spp. and *tet*(X4)-carrying *E. coli* were prevalent in food animals and these two
426 *tet*(X)s were likely formed during the transmission of *tet*(X)s between *Flavobacteriaceae* and *E.*
427 *coli*/*Acinetobacter*. The IS*CR2* element played a key role in the transmission. The *tet*(X2)-like
428 orthologs enriched in the anaerobes that was dominated by *Bacteroidaceae* of human-gut origin
429 and could transfer between these anaerobes. The mobile elements CTnDOT, IS*Bf11* and IS*4351*
430 played important roles in the transmission. The low-level tigecycline resistance *tet*(X2)-like gene
431 could mutate to high-level tigecycline resistant determinants that could spread to facultative
432 anaerobes and aerobes. *Bacteroidaceae* present in the human gut were an important reservoir and
433 mutational incubator for *tet*(X) genes.

434

435 Acknowledgments

436 This work was supported by the National Natural Science Foundation of China (31730097), the
437 Program for Changjiang Scholars and Innovative Research Team in University of Ministry of
438 Education of China IRT_17R39, the Local Innovative and Research Teams Project of Guangdong
439 Pearl River Talents Program (2019BT02N054), and the 111 Project (D20008).
440 The authors are grateful to the previous studies that provide a large-scale of microbial genomic
441 bins for the construction of the metagenomic analysis in current study.

442

Reference

1. Nelson ML, Levy SB. 2011. The history of the tetracyclines. Ann N Y Acad Sci 1241:17-32.
2. Fang LX, Chen C, Cui CY, Li XP, Zhang Y, Liao XP, Sun J, Liu YH. 2020. Emerging High-Level Tigecycline Resistance: Novel Tetracycline Destructases Spread via the Mobile Tet(X). Bioessays 42(8):e2000014.
3. Sarmah AK, Meyer MT, Boxall AB. 2006. A global perspective on the use, sales, exposure pathways, occurrence, fate and effects of veterinary antibiotics (VAs) in the environment. Chemosphere 65:725-59.
4. Aminov RI. 2013. Evolution in action: dissemination of *tet(*X) into pathogenic microbiota. Front Microbiol 4:192.
5. Zhang R, Dong N, Shen Z, Zeng Y, Lu J, Liu C, Zhou H, Hu Y, Sun Q, Cheng Q, Shu L, Cai J, Chan EW, Chen G, Chen S. 2020. Epidemiological and phylogenetic analysis reveals *Flavobacteriaceae* as potential ancestral source of tigecycline resistance gene *tet*(X). Nat Commun 11:4648.
6. Bertrand X, Dowzicky MJ. 2012. Antimicrobial susceptibility among gram-negative isolates collected from intensive care units in North America, Europe, the Asia-Pacific Rim, Latin America, the Middle East, and Africa between 2004 and 2009 as part of the Tigecycline Evaluation and Surveillance Trial. Clin Ther 34:124-37.
7. He T, Wang R, Liu D, Walsh TR, Zhang R, Lv Y, Ke Y, Ji Q, Wei R, Liu Z, Shen Y, Wang G, Sun L, Lei L, Lv Z, Li Y, Pang M, Wang L, Sun Q, Fu Y, Song H, Hao Y, Shen Z, Wang S, Chen G, Wu C, Shen J, Wang Y. 2019. Emergence of plasmid-mediated high-level tigecycline resistance genes in animals and humans. Nat Microbiol 4(9):1450-1456.
8. Guiney DG, Jr., Hasegawa P, Davis CE. 1984. Expression in *Escherichia coli* of cryptic tetracycline resistance genes from bacteroides R plasmids. Plasmid 11:248-52.
9. Yang W, Moore IF, Koteva KP, Bareich DC, Hughes DW, Wright GD. 2004. TetX is a flavin-dependent monooxygenase conferring resistance to tetracycline antibiotics. J Biol Chem 279:52346-52352.
10. Sun J, Yang RS, Zhang Q, Feng Y, Fang LX, Xia J, Li L, Lv XY, Duan JH, Liao XP, Liu YH. 2016. Co-transfer of $bla_{\text{NDM-5}}$ and *mcr-1* by an IncX3-X4 hybrid plasmid in *Escherichia coli*. Nat Microbiol 1:16176.
11. Cheng Y, Chen Y, Liu Y, Guo Y, Zhou Y, Xiao T, Zhang S, Xu H, Chen Y, Shan T, Xiao Y, Zhou K. 2020. Identification of novel tetracycline resistance gene *tet*(X14) and its co-occurrence with *tet*(X2) in a tigecycline-resistant and colistin-resistant *Empedobacter stercoris*. Emerg Microbes Infect 9(1):1843-1852.
12. Wang L, Liu D, Lv Y, Cui L, Li Y, Li T, Song H, Hao Y, Shen J, Wang Y, Walsh TR. 2019. Novel plasmid-mediated *tet*(X5) gene conferring resistance to tigecycline, eravacycline and omadacycline in clinical *Acinetobacter baumannii*. Antimicrob Agents Chemother 64(1):e01326-19.
13. Liu D, Zhai W, Song H, Fu Y, Schwarz S, He T, Bai L, Wang Y, Walsh TR, Shen J. 2020. Identification of the novel tigecycline resistance gene *tet*(X6) and its variants in *Myroides*, *Acinetobacter* and *Proteus* of food animal origin. J Antimicrob Chemother 75(6):1428-1431.

14. Gasparrini AJ, Markley JL, Kumar H, Wang B, Fang L, Irum S, Symister CT, Wallace M, Burnham CD, Andleeb S, Tolia NH, Wencewicz TA, Dantas G. 2020. Tetracycline-inactivating enzymes from environmental, human commensal, and pathogenic bacteria cause broad-spectrum tetracycline resistance. Commun Biol 3:241.
15. Chen C, Cui C-Y, Yu J-J, He Q, Wu X-T, He Y-Z, Cui Z-H, Li C, Jia Q-L, Shen X-G, Sun R-Y, Wang X-R, Wang M-G, Tang T, Zhang Y, Liao X-P, Kreiswirth BN, Zhou S-D, Huang B, Du H, Sun J, Chen L, Liu Y-H. 2020. Genetic diversity and characteristics of high-level tigecycline resistance Tet(X) in *Acinetobacter* species. Genome Medicine 12:111.
16. Li R, Lu X, Peng K, Liu Z, Li Y, Liu Y, Xiao X, Wang Z. 2020. Deciphering the Structural Diversity and Classification of the Mobile Tigecycline Resistance Gene *tet*(X)-Bearing Plasmidome among *Bacteria*. mSystems 5(2):e00134-20.
17. Hu Y, Liu F, Lin IY, Gao GF, Zhu B. 2016. Dissemination of the *mcr-1* colistin resistance gene. Lancet Infect Dis 16:146-7.
18. Wang Y, Liu F, Zhu B, Gao GF. 2020. Metagenomic data screening reveals the distribution of mobilized resistance genes *tet*(X), *mcr* and carbapenemase in animals and humans. J Infect 80:121-142.
19. Zeng J, Pan Y, Yang J, Hou M, Zeng Z, Xiong WJEi. 2019. Metagenomic insights into the distribution of antibiotic resistome between the gut-associated environments and the pristine environments. Environ Int 126:346-354.
20. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P, Collado MC, Rice BL, DuLong C, Morgan XC, Golden CD, Quince C, Huttenhower C, Segata N. 2019. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell 176:649-662.e20.
21. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC. 2019. New insights from uncultivated genomes of the global human gut microbiome. Nature 568:505-510.
22. Forster SC, Kumar N, Anonye BO, Almeida A, Viciani E, Stares MD, Dunn M, Mkandawire TT, Zhu A, Shao Y, Pike LJ, Louie T, Browne HP, Mitchell AL, Neville BA, Finn RD, Lawley TD. 2019. A human gut bacterial genome and culture collection for improved metagenomic analyses. Nat Biotechnol 37:186-192.
23. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35:1547-1549.
24. Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42:W320-4.
25. Sun J, Chen C, Cui CY, Zhang Y, Liu X, Cui ZH, Ma XY, Feng Y, Fang LX, Lian XL, Zhang RM, Tang YZ, Zhang KX, Liu HM, Zhuang ZH, Zhou SD, Lv JN, Du H, Huang B, Yu FY, Mathema B, Kreiswirth BN, Liao XP, Chen L, Liu YH. 2019. Plasmid-encoded *tet*(X) genes that confer high-level tigecycline resistance in *Escherichia coli*. Nat Microbiol 4:1457-1464.
26. Uritskiy GV, DiRuggiero J, Taylor J. 2018. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 6:158.
27. Wood DE, Salzberg SLJGb. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:R46.
28. Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, Zhu Q, Bolzan M, Cumbo F, May U, Sanders JG, Zolfo M, Kopylova E, Pasolli E, Knight R, Mirarab S,

Huttenhower C, Segata N. 2020. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat Commun 11:2500.

29. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-80.

30. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972-3.

31. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-3.

32. Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N. 2015. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ 3:e1029.

33. Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214.

34. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol 67:901-904.

35. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256-w259.

36. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150-2.

37. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-9.

38. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo AR, Florensa AF, Fagelhauer L, Chakraborty T, Neumann B, Werner G, Bender JK, Stingl K, Nguyen M, Coppens J, Xavier BB, Malhotra-Kumar S, Westh H, Pinholt M, Anjum MF, Duggett NA, Kempf I, Nykäsenoja S, Olkkola S, Wieczorek K, Amaro A, Clemente L, Mossong J, Losch S, Ragimbeau C, Lund O, Aarestrup FM. 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother 75(12):3491-3500.

39. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32-6.

40. Whittle G, Hund BD, Shoemaker NB, Salyers AA. 2001. Characterization of the 13-KilobaseermF Region of the *Bacteroides* Conjugative Transposon CTnDOT. Appl Environ Microbiol 67:3488-3495.

41. Umar Z, Chen Q, Tang B, Xu Y, Wang J, Zhang H, Ji K, Jia X, Feng Y. 2021. The poultry pathogen Riemerella anatipestifer appears as a reservoir for Tet(X) tigecycline resistance. Environ Microbiol doi:10.1111/1462-2920.15632.

42. Partridge SR, Di Pilato V, Doi Y, Feldgarden M, Haft DH, Klimke W, Kumar-Singh S, Liu JH, Malhotra-Kumar S, Prasad A, Rossolini GM, Schwarz S, Shen J, Walsh T, Wang Y, Xavier BB. 2018. Proposal for assignment of allele numbers for mobile colistin resistance (*mcr*) genes. J Antimicrob Chemother 73:2625-2630.

43. Wang Y, Zhang R, Li J, Wu Z, Yin W, Schwarz S, Tyrrell JM, Zheng Y, Wang S, Shen Z, Liu Z, Liu J, Lei L, Li M, Zhang Q, Wu C, Zhang Q, Wu Y, Walsh TR, Shen J. 2017. Comprehensive resistome analysis reveals the prevalence of NDM and MCR-1 in Chinese poultry production. Nat Microbiol 2:16260.

44. Moore WE, Cato EP, Holdeman LV. 1978. Some current concepts in intestinal bacteriology. Am J Clin Nutr 31:S33-42.

45. Consortium HMP. 2012. Structure, function and diversity of the healthy human microbiome.

Nature 486:207-14.

46. Zhang R-M, Liu X, Wang S-L, Fang L-X, Sun J, Liu Y-H, Liao X-P. 2020. Distribution patterns of antibiotic resistance genes and their bacterial hosts in pig farm wastewater treatment systems and soil fertilized with pig manure. Sci Total Environ 758:143654.

47. Sczyrba A, Hofmann P, Belmann P, Koslicki D, Janssen S, Dröge J, Gregor I, Majda S, Fiedler J, Dahms E, Bremges A, Fritz A, Garrido-Oter R, Jørgensen TS, Shapiro N, Blood PD, Gurevich A, Bai Y, Turaev D, DeMaere MZ, Chikhi R, Nagarajan N, Quince C, Meyer F, Balvočiūtė M, Hansen LH, Sørensen SJ, Chia BKH, Denis B, Froula JL, Wang Z, Egan R, Don Kang D, Cook JJ, Deltel C, Beckstette M, Lemaitre C, Peterlongo P, Rizk G, Lavenier D, Wu YW, Singer SW, Jain C, Strous M, Klingenberg H, Meinicke P, Barton MD, Lingner T, Lin HH, Liao YC, et al. 2017. Critical Assessment of Metagenome Interpretation-a benchmark of metagenomics software. Nat Methods 14:1063-1071.

48. Salyers AA, Shoemaker NB, Li LY. 1995. In the driver's seat: the Bacteroides conjugative transposons and the elements they mobilize. J Bacteriol 177:5727-31.

49. Xing L, Yu H, Qi J, Jiang P, Sun B, Cui J, Ou C, Chang W, Hu Q. 2015. *ErmF* and *ereD* are responsible for erythromycin resistance in *Riemerella anatipestifer*. PLoS One 10:e0131078.

50. Johnsen BO, Handal N, Meisal R, Bjørnholt JV, Gaustad P, Leegaard TM. 2017. *erm* gene distribution among *Norwegian Bacteroides* isolates and evaluation of phenotypic tests to detect inducible clindamycin resistance in *Bacteroides* species. Anaerobe 47:226-232.

51. Pan Y, Awan F, Zhenbao M, Zhang X, Zeng J, Zeng Z, Xiong W. 2020. Preliminary view of the global distribution and spread of the *tet*(X) family of tigecycline resistance genes. J Antimicrob Chemother 75(10):2797-2803.

52. Cheng Y, Chen Y, Liu Y, Song J, Chen Y, Shan T, Xiao Y, Zhou K. 2021. Detection of a new *tet*(X6)-encoding plasmid in *Acinetobacter towneri*. J Glob Antimicrob Resist 25:132-136.

Figure legends

Figure 1. Global distribution of *tet*(X)s from human microbiome. A. World map showed the positive rates of *tet*(X) gene in 19 countries and the colored countries represented the positive rates of *tet*(X) according to the hot map (>0% - 9.40%) at the bottom. The gray countries indicated they were negative for *tet*(X) gene. The size of the pie charts represented the numbers of *tet*(X)-positive MGBs and the colors in the pie charts indicated the composition of *tet*(X) variants. B. PhyloPhlAn analysis of the *tet*(X) carrying MGBs. The taxonomic assignments of the *tet*(X)-carrying MGBs were depicted with colored circles in the phylogenetic tree. The *tet*(X) variants carried by the MGBs, as well as oxygen demand, gram stain, age category and the

607 countries of the *tet*(X)-carrying MGBs were showed in the five colored rings surrounding the
608 phylogenetic tree. C. Distribution of the 12 MGBs carried non-*tet*(X2) genes with tigecycline
609 inactivate function.
610
611 Figure 2. Culturable isolates insight into *tet*(X) distribution patterns. A. PhyloPhlAn analysis of
612 the *tet*(X)-carrying isolates from the public repository. The species of the *tet*(X)-carrying isolates
613 were depicted with colored circles in the phylogenetic tree. The information of the *tet*(X) carrying
614 isolates including *tet*(X) variants, oxygen demand, host, collection date and country were showed
615 in the six colored rings surrounding the phylogenetic tree. B. Dates of lineage divergence of the
616 earliest *tet*(X) orthologs as determined using Bayesian phylogenetic inference. The *tet*(X) variants,
617 countries, host and species of these isolates were shown at the right region.
618
619 Figure 3. Possible transmission routes of the *tet*(X) genes. The colored circles surrounding the
620 hosts represented the *tet*(X) carrying bacterial families. The dotted lines represented the possible
621 transmission routes of the *tet*(X) genes between different hosts.
622 Figure 4. Comparison of *tet*(X) genomic environment. A. The genomic comparison of *erm*(F) gene
623 located upstream of *tet*(X2)-like genes. The proportions of the *tet*(X2) and *tet*(X2.2) located
624 downstream *tet*(X1) were showed in the pie chart. B. The genomic comparison of *tet*(X) genes
625 located downstream *erm*(F). The proportions of the *tet*(X) variants located downstream of *erm*(F)
626 were showed in the pie chart. The possible mechanisms of non-*tet*(X2) formations were showed in
627 the two circles plotted with dotted line. C. Genomic comparison of the regions flanking *tet*(X3)
628 and *tet*(X4) among *Flavobacteriaceae*, *Acinetobacter* and *E. coli*. Arrows indicate the directions

629 of transcription of the genes, and different genes are shown in different colors. Regions of ≥ 99.0%
630 nucleotide sequence identity are shaded light grey. Regions of 77% - 91% nucleotide sequence
631 identity are shaded dark grey. The *Δ* symbol indicates a truncated gene. IS, insertion sequence. See
632 Table S11 for genomic Type I - XVII definitions.
633 Figure 5. Potential origin and main transmission routes of the *tet*(X) genes.
634
635 Supplementary files
636 Text S1. Description of the comparison of *tet*(X) genomic-environment in Figure 4 and S5.
637 Figure S1. Phylogenetic analysis of the Tet(X) at amino acid level.
638 Figure S2. Comparative analysis of the Tet(X2)-like orthologs at the amino acid level.
639 Figure S3. Significance analysis of *tet*(X2) distributions in Europe, Asia and America.
640 Figure S4. Distribution of *tet*(X2)-like orthologs among different bacterial genus, countries and
641 age groups from microbial genomic bins of human-gut origin.
642 Figure S5. Comparative analysis of the *tet*(X2) genomic context among *E. coli*, *R. anatipestifer*,
643 *Phocaeicola vulgatus* and *Odoribacter laneus*. Arrows indicate the directions of transcription of
644 the genes, and different genes are shown in different colors. Regions of ≥ 99.0% nucleotide
645 sequence identity are shaded light grey. The *Δ* symbol indicates a truncated gene. IS, insertion
646 sequence. See Table S11 for genomic Type XVIII – XXII definitions.
647 Table S1. A total of 12,829 non-duplicate metagenomic samples that derived from previous studies
648 (see PMID of Publication).
649 Table S2. A total of 202,265 MGBs that was constructed from the 12,829 non-duplicate
650 metagenomic samples.

651 Table S3. The *tet*(X2)-like and non *tet*(X2) orthologs designed in current study.

652 Table S4. Minimum inhibitory concentration of the *tet*(X)s from metagenomic analysis. TET:

653 tetracycline; DOX: doxycycline; MIN; minocycline; TIG: tigecycline; ERA: eravacycline; OMA:

654 omadacycline.

655 Table S5. Positive rates of *tet*(X) gene in 31 countries.

656 Table S6. Positive rates of *tet*(X) carrying MGBs annotated at family level.

657 Table S7. Positive rates of *tet*(X) carrying MGBs annotated at species level.

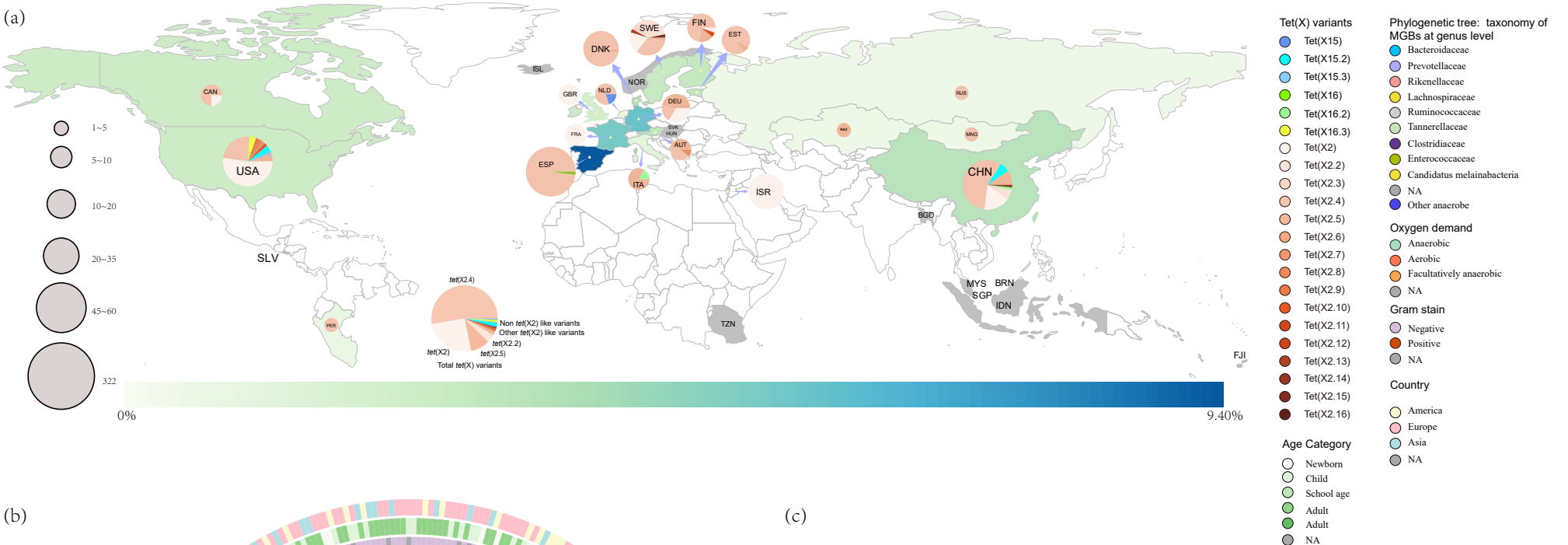
658 Table S8. Detail information of the 322 *tet*(X) carrying MGBs.

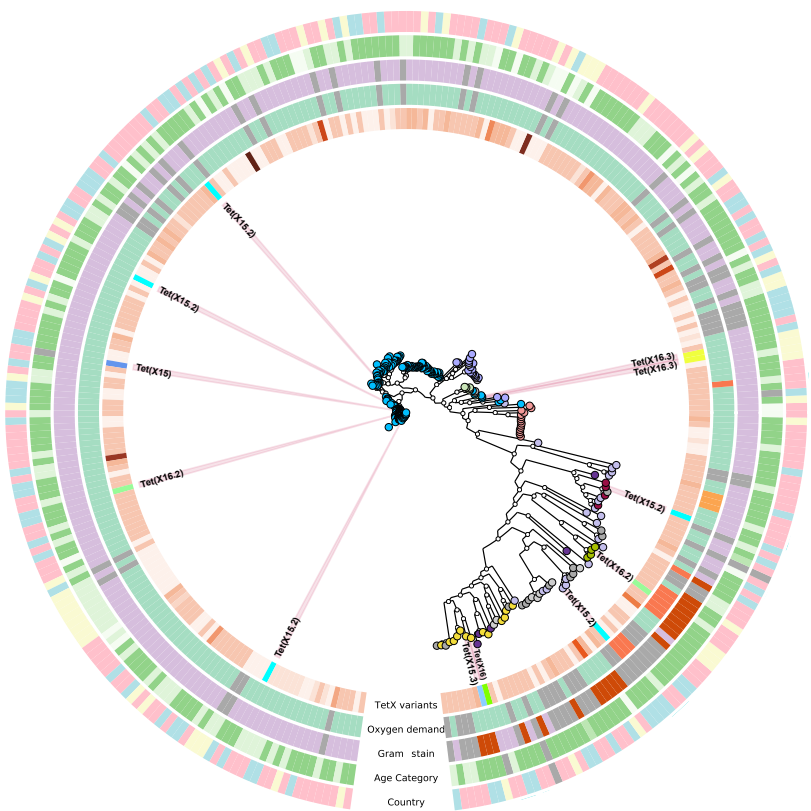
659 Table S9. Detail information of the 896 *tet*(X) carrying bacterial isolates.

660 Table S10. Clusters of the 1218 *tet*(X) positive contigs from the MGBs and bacterial isolates.

661 Table S11. Detail information of the *tet*(X) genomic context types I to XXII.

Figure 1


(a)
1~5
5~10
10~20
20~35
45~60
322
0%
9.40%
USA
CAN
SLV
PER
DNK
SWE
FIN
EST
NOR
ISL
GBR
NLD
DEU
FRA
ESP
ITA
SVK
HUN
AUT
ISR
TZN
RUS
KAZ
MNG
CHN
BGD
MYS
BRN
SGP
IDN
FJI
tet(X2.4)
Non tet(X2) like variants
Other tet(X2) like variants
tet(X2.2)
tet(X2)
tet(X2.5)
Total tet(X) variants
Tet(X) variants
Tet(X15)
Tet(X15.2)
Tet(X15.3)
Tet(X16)
Tet(X16.2)
Tet(X16.3)
Tet(X2)
Tet(X2.2)
Tet(X2.3)
Tet(X2.4)
Tet(X2.5)
Tet(X2.6)
Tet(X2.7)
Tet(X2.8)
Tet(X2.9)
Tet(X2.10)
Tet(X2.11)
Tet(X2.12)
Tet(X2.13)
Tet(X2.14)
Tet(X2.15)
Tet(X2.16)
Age Category
Newborn
Child
School age
Adult
Adult
NA
Phylogenetic tree: taxonomy of MGBs at genus level
Bacteroidaceae
Prevotellaceae
Rikenellaceae
Lachnospiraceae
Ruminococcaceae
Tannerellaceae
Clostridiaceae
Enterococcaceae
Candidatus melainabacteria
NA
Other anaerobe
Oxygen demand
Anaerobic
Aerobic
Facultatively anaerobic
NA
Gram stain
Negative
Positive
NA
Country
America
Europe
Asia
NA


(b)
Tet(X15.2)
Tet(X15.2)
Tet(X15)
Tet(X16.2)
Tet(X16.3)
Tet(X16.3)
Tet(X15.2)
Tet(X16.2)
Tet(X15.2)
Tet(X15.2)
Tet(X16)
Tet(X15.3)
TetX variants
Oxygen demand
Gram stain
Age Category
Country

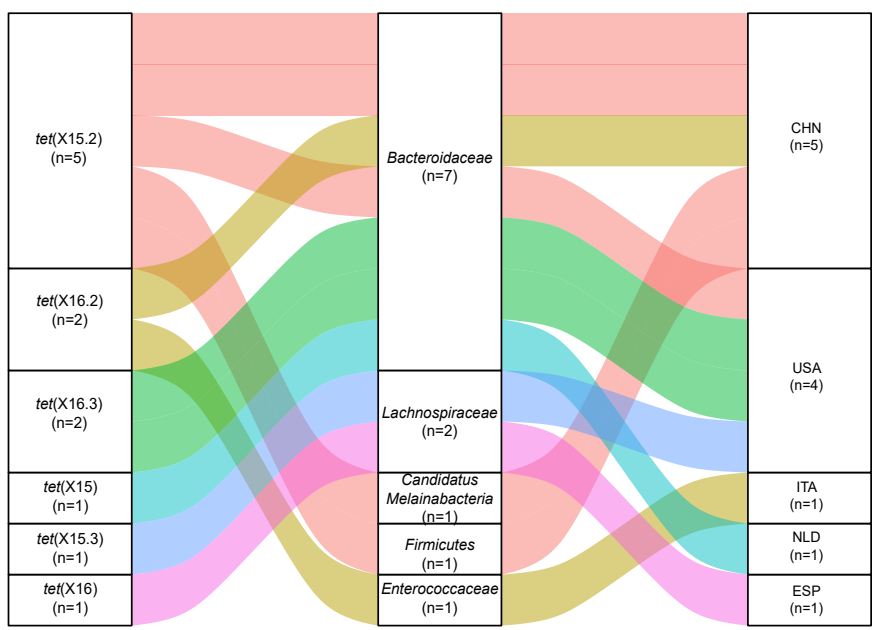
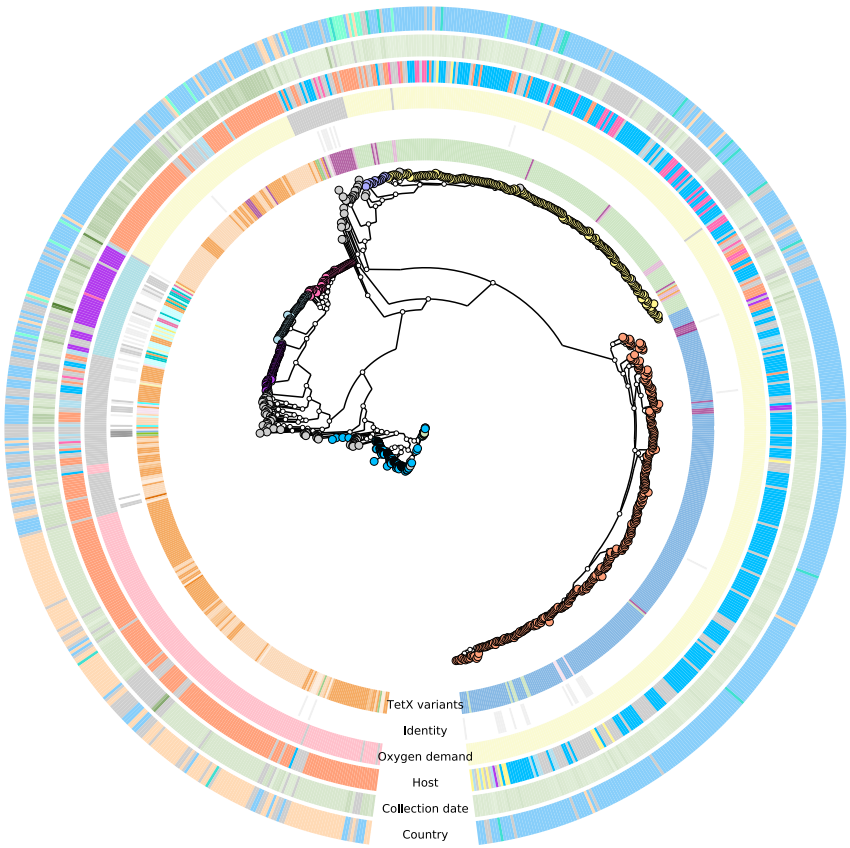

(c)
tet(X15.2) (n=5)
tet(X16.2) (n=2)
tet(X16.3) (n=2)
tet(X15) (n=1)
tet(X15.3) (n=1)
tet(X16) (n=1)
Bacteroidaceae (n=7)
Lachnospiraceae (n=2)
Candidatus Melainabacteria (n=1)
Firmicutes (n=1)
Enterococcaceae (n=1)
CHN (n=5)
USA (n=4)
ITA (n=1)
NLD (n=1)
ESP (n=1)

Figure 2

(a)



Tet(X) variants
- Tet(X2.2)
- Tet(X2.4)
- Tet(X2.5)
- Tet(X2)
- Tet(X2.13)
- Tet(X2.16)
- Tet(X4)
- Tet(X16)
- Tet(X16.2)
- Tet(X3)
- Tet(X15.2)
- Tet(X15.3)
- Tet(X14)
- Tet(X11)
- Tet(X12)
- Tet(X16.3)
- Tet(X5)
- Tet(X6)
- Tet(X7)
- Tet(X13)

Collection year
2020
1966
Unknow

Taxonamy of isolates at species level
- Acinetobacter spp.
- Escherichia coli
- Bacteroides spp.
- Chryseobacterium spp.
- Riemerella anatipestifer
- Elizabethkingia spp.
- Phocaeicola vulgatus
- Salmonella spp.
- Others

Similarity with the *tet*(X)s
- 70%-90%
- 90%-97%
- 97%-99.9%
- 100%

Oxygen demand
- Aerobic
- Anaerobic
- Facultatively anaerobic

Host
- Human
- Pig
- Duck
- Chicken
- Cattle
- Environment
- Unknow

Country
- China
- The USA
- The UK
- Thailand
- Other

(b)

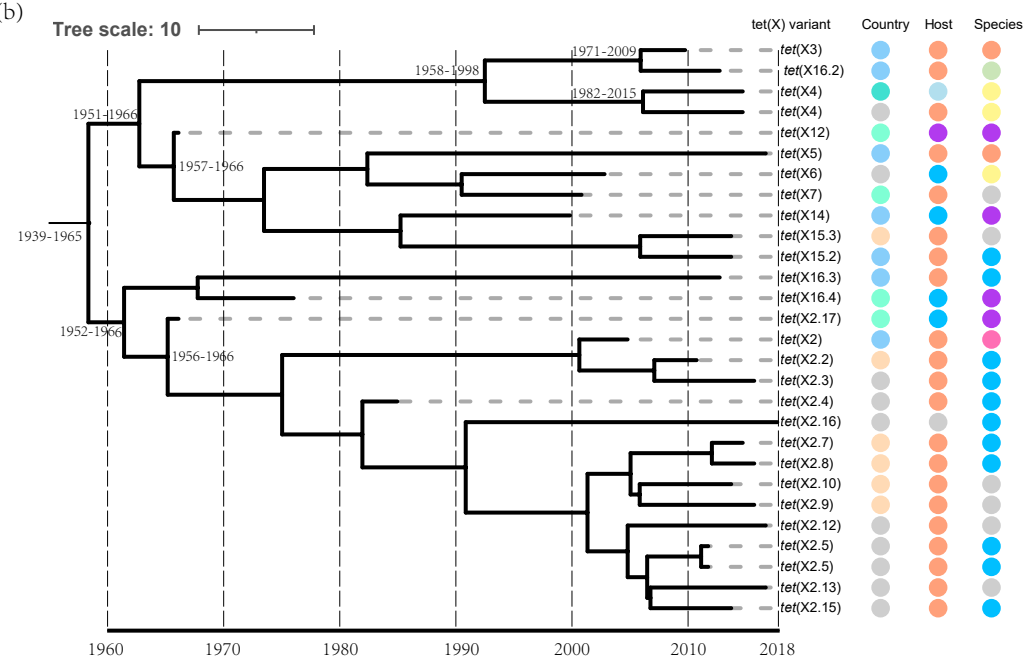


Figure 3

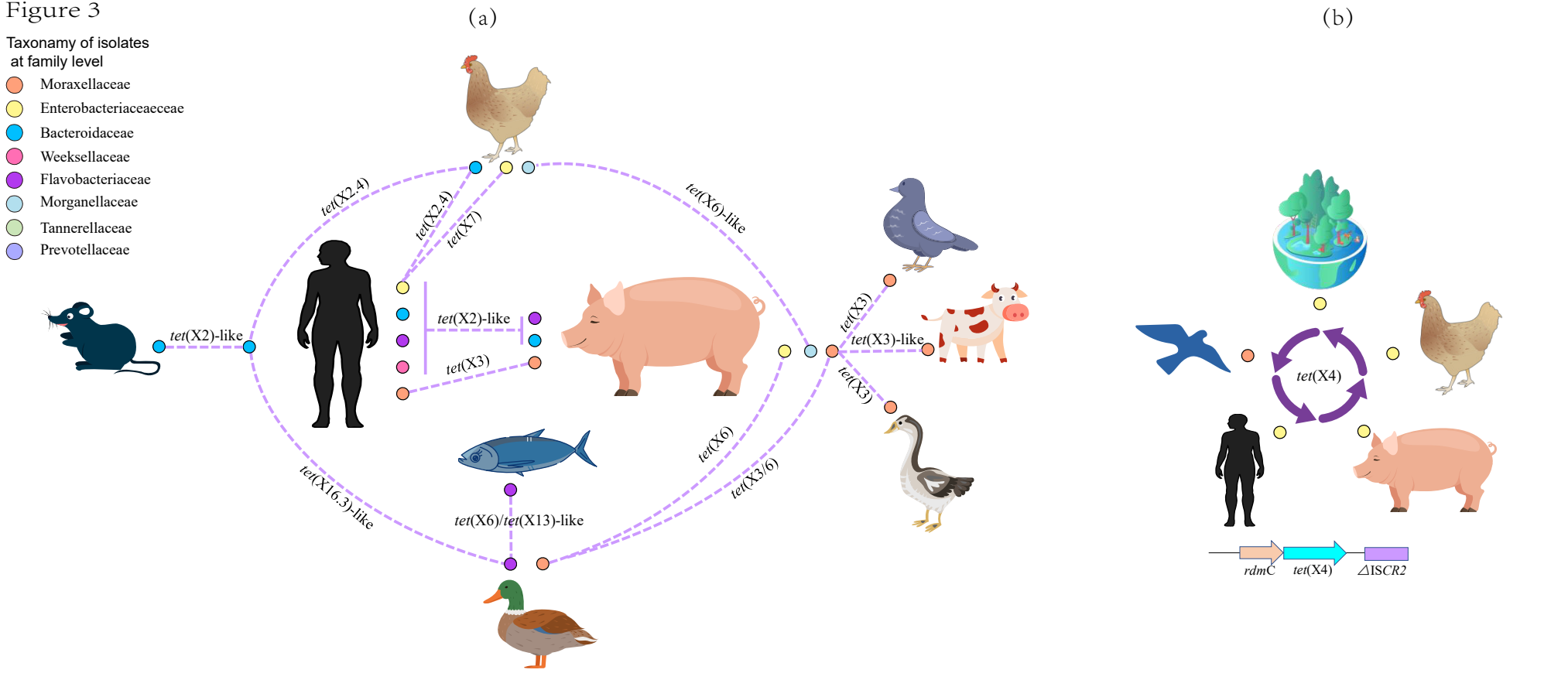

(a)
Taxonamy of isolates at family level
Moraxellaceae
Enterobacteriaceaeceae
Bacteroidaceae
Weeksellaceae
Flavobacteriaceae
Morganellaceae
Tannerellaceae
Prevotellaceae
tet(X2)-like
tet(X2.4)
tet(X2.4)
tet(X7)
tet(X2)-like
tet(X3)
tet(X6)-like
tet(X3)
tet(X3)-like
tet(X3)
tet(X6)
tet(X3/6)
tet(X16.3)-like
tet(X6)/tet(X13)-like
(b)
tet(X4)
rdmC
tet(X4)
ΔISCR2

Figure 4

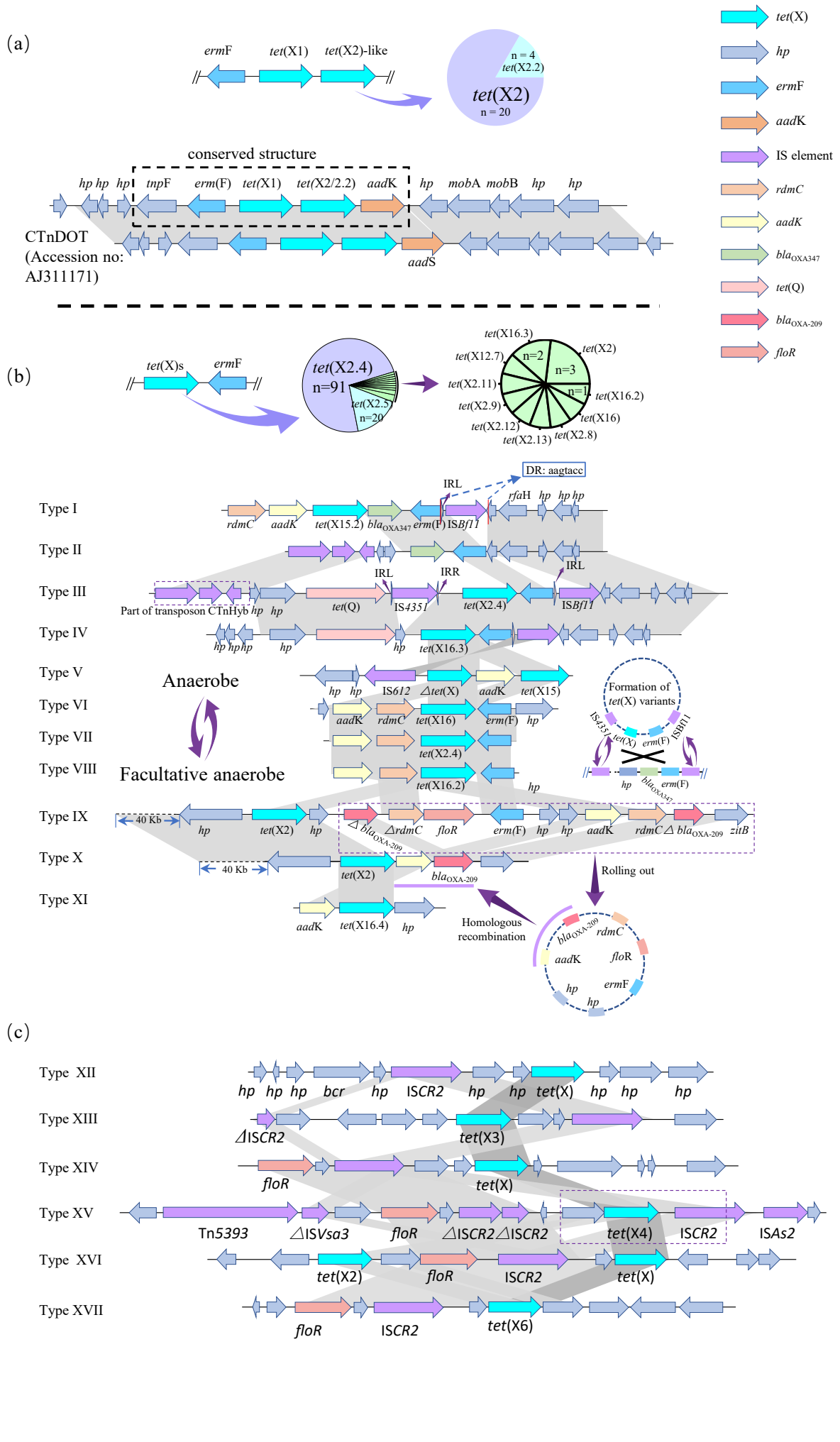
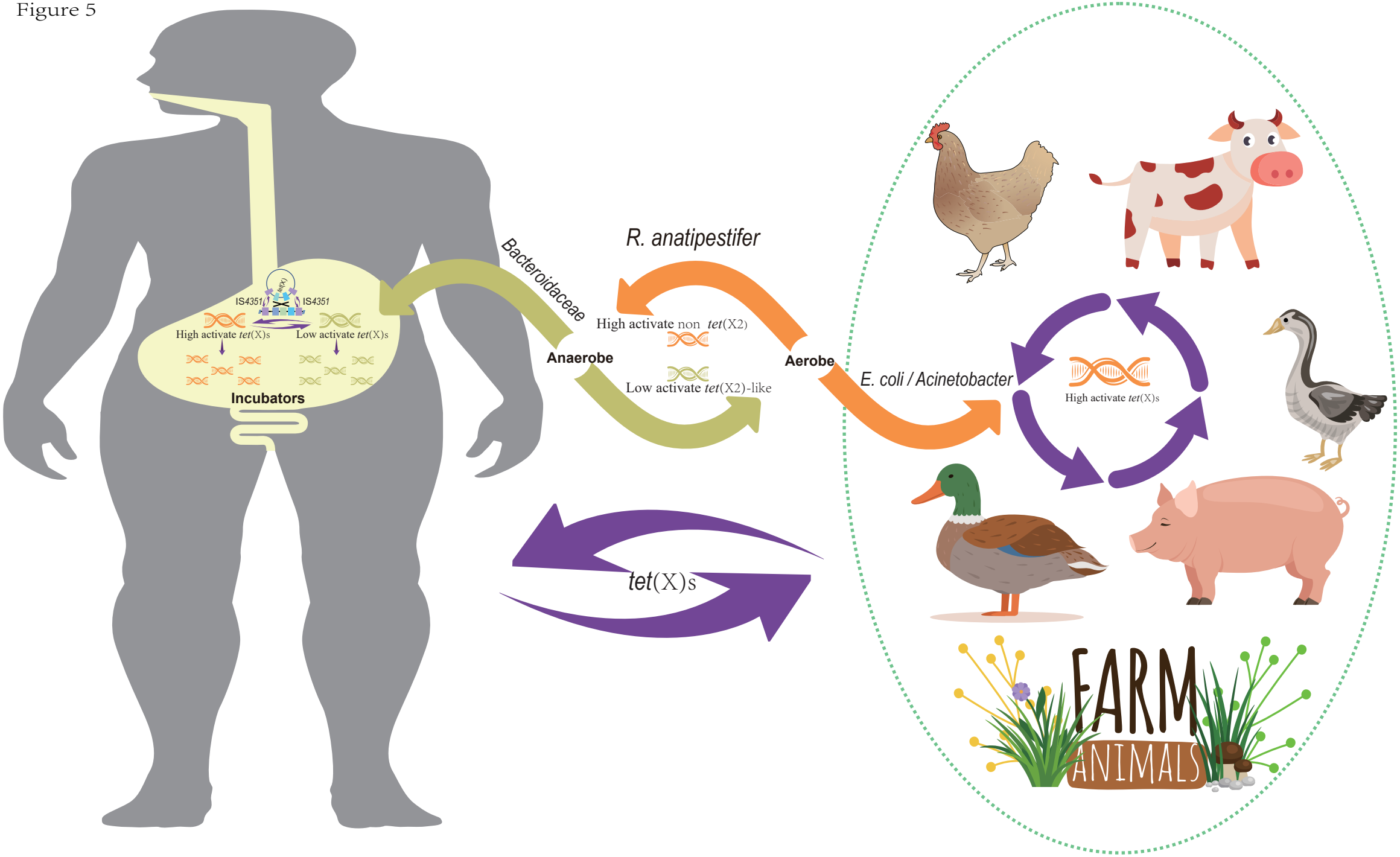

(a)
ermF
tet(X1)
tet(X2)-like
n = 4
tet(X2.2)
tet(X2)
n = 20
conserved structure
hp hp hp
tnpF
erm(F)
tet(X1)
tet(X2/2.2)
aadK
hp
mobA
mobB
hp
hp
CTnDOT
(Accession no:
AJ311171)
aadS
tet(X)
hp
ermF
aadK
IS element
rdmC
aadK
bla OXA347
tet(Q)
bla OXA-209
floR
(b)
tet(X)s
ermF
tet(X2.4)
n=91
tet(X2.5)
n=20
tet(X16.3)
tet(X12.7)
tet(X2.11)
tet(X2.9)
tet(X2.12)
tet(X2.13)
tet(X2)
tet(X16.2)
tet(X16)
tet(X2.8)
n=2
n=3
n=1
Type I
Type II
Type III
Type IV
Type V
Type VI
Type VII
Type VIII
Type IX
Type X
Type XI
DR: aagtacc
IRL
rfaH
rdmC
aadK
tet(X15.2)
bla OXA347
erm(F)
ISBf11
IRR
Part of transposon CTnHyb
tet(Q)
IS4351
tet(X2.4)
tet(X16.3)
Anaerobe
Facultative anaerobe
IS612
△tet(X)
tet(X15)
tet(X16)
tet(X16.2)
Formation of
tet(X) variants
ISBf11
40 Kb
tet(X2)
△ bla OXA-209
△rdmC
floR
erm(F)
aadK
zitB
Rolling out
Homologous
recombination
tet(X16.4)
(c)
Type XII
Type XIII
Type XIV
Type XV
Type XVI
Type XVII
bcr
ISCR2
△ISCR2
tet(X3)
Tn5393
△ISVsa3
△ISCR2△ISCR2
tet(X4)
ISAs2
tet(X6)

Figure 5


IS4351
IS4351
tet(X)
High activate tet(X)s
Low activate tet(X)s
Incubators
Bacteroidaceae
Anaerobe
R. anatipestifer
High activate non tet(X2)
Low activate tet(X2)-like
Aerobe
E. coli / Acinetobacter
High activate tet(X)s
tet(X)s
FARM
ANIMALS