Genomic and evolutionary analysis of *Clostridium difficile* ST11: a genetically diverse lineage of significant One Health importance

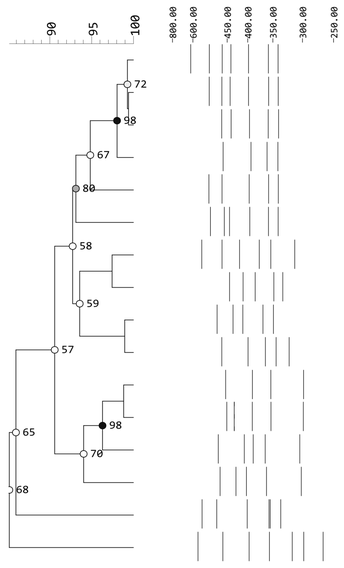
Daniel R. Knight[1], Grace O. Androga[1,2], Frédéric Barbut[3], Catherine Eckert[3], Stuart Johnson[4], Patrizia Spigaglia[5], Kazuhiro Tateda[6], Pei-Jane Tsai[7] and Thomas V. Riley[2,8,9]*

[1]School of Biomedical Sciences, Faculty of Health and Medical Sciences, The University of Western Australia, Nedlands, Western Australia, Australia; [2]PathWest Laboratory Medicine, Department of Microbiology, Queen Elizabeth II Medical Centre, Nedlands, Western Australia, Australia; [3]National Reference Laboratory for *C. difficile*, Hospital Saint-Antoine, Assistance Publique Hôpitaux de Paris, Paris, France; [4]Loyola University Medical Center and Hines Veteran Affairs Hospital, Illinois, USA; [5]Infectious, Parasitic and Immunomediated Diseases, Istituto Superiore di Sanità, Rome, Italy; [6]Department of Microbiology and Infectious Diseases, Toho University School of Medicine, Tokyo, Japan; [7]Department of Microbiology and Immunology, National Cheng Kung University, Medical College, Tainan, Taiwan; [8]Department of Population Health, School of Medical and Health Sciences, Edith Cowan University, Joondalup, Western Australia, Australia; [9]Department of Medical and Molecular Sciences, School of Veterinary and Life Sciences, Murdoch University, Murdoch, Western Australia, Australia.

BACKGROUND, STUDY AIMS AND STRAIN COLLECTION

C. difficile sequence type (ST) 11 is a diverse evolutionary lineage comprising at least 5 PCR ribotypes (RTs) that contribute considerably to the global burden of *C. difficile* infection (CDI) in humans and production animals[1]. Increasing evidence of genetic overlap of RT 078, the most common ST11 sub-lineage, between humans and animals suggests that CDI may have a zoonotic or foodborne aetiology[1,2,3]. However, almost nothing is currently known about the phylogenomics of non-078 ST11 RTs 126, 127, 033 and 288. Specifically, the exact evolutionary and epidemiological relationship between these RTs and RT078, and the antimicrobial resistance (AMR) repertoire and zoonotic potential of these ST11 sub-lineages, remain open questions.

Here, whole genome sequencing (WGS) and high-resolution core genome phylogenetics were used to examine relatedness and clonal transmission in a collection of 207 ST11 isolates of clinical (*n*=150) and veterinary/environmental (*n*=57) origin from Australia and 12 countries across Asia, Europe and North America (16 RTs including major ST11 sub-lineages 078, 126, 127, 033 and 288, **Fig. 1**). We also characterised the ST11 resistome, pan-genome, and prophage content in this population.



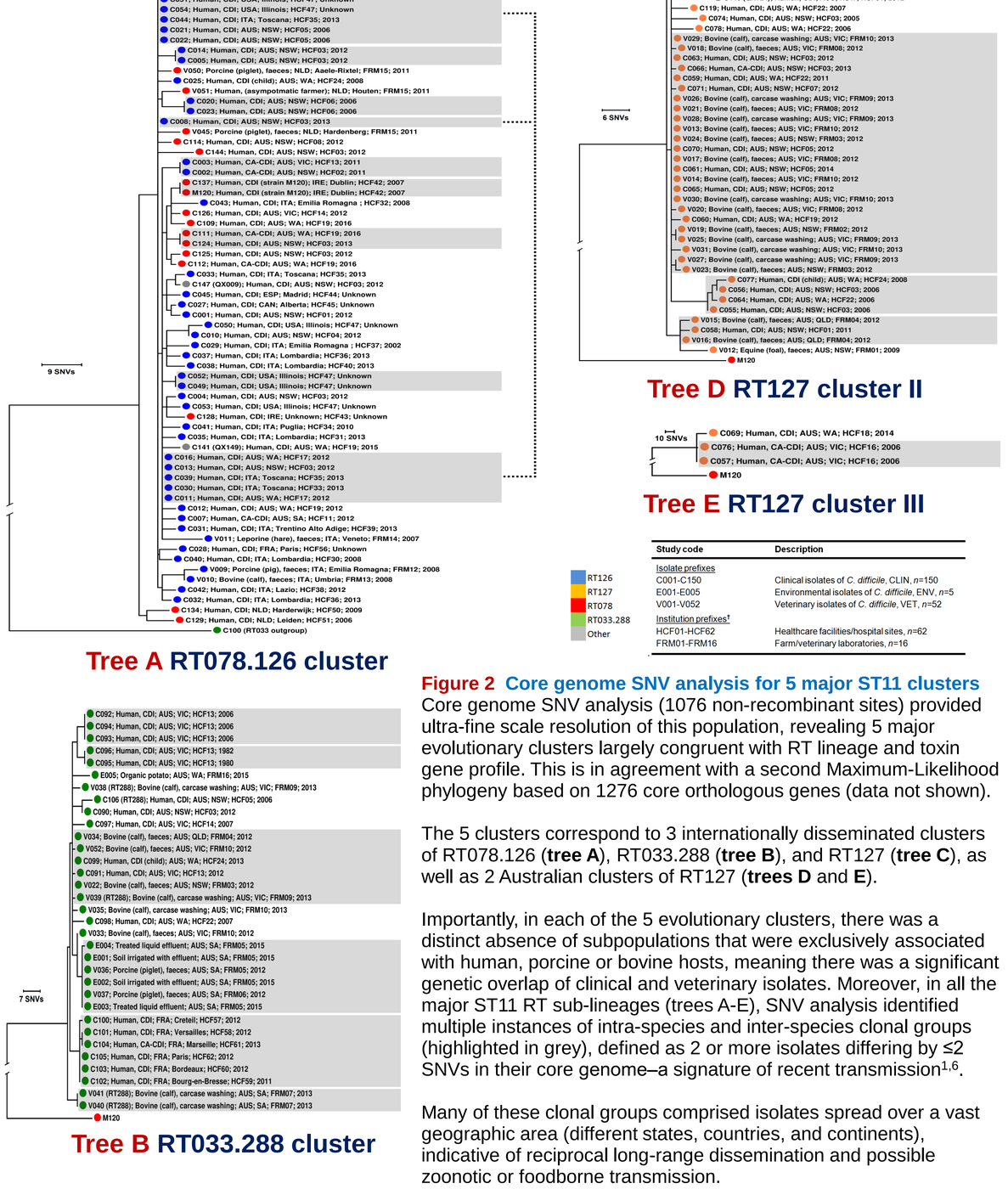
PCR ribotype (toxin gene profile)	N CLI	N VET	N ENV	Grand Total N (%)
QX528 ($A^+B^+CDT^+$)	1	0	0	**1 (0.5)**
RT078 ($A^+B^+CDT^+$)†	30	10	0	**40 (19.3)**
RT126 ($A^+B^+CDT^+$)‡	58	11	0	**69 (33.3)**
RT127 ($A^+B^+CDT^+$)	34	20	0	**54 (26.1)**
QX360 ($A^+B^+CDT^+$)	1	0	0	**1 (0.5)**
QX399 ($A^+B^+CDT^+$)	1	0	0	**1 (0.5)**
QX009 ($A^-B^+CDT^+$)	1	0	0	**1 (0.5)**
QX220 ($A^+B^+CDT^+$)	1	0	0	**1 (0.5)**
QX404 ($A^+B^+CDT^+$)	1	0	0	**1 (0.5)**
QX124 ($A^+B^+CDT^+$)	1	0	0	**1 (0.5)**
RT033 ($A^-B^-CDT^+$)	16	7	5	**28 (13.5)**
RT288 ($A^-B^-CDT^+$)	1	4	0	**5 (2.4)**
QX060 ($A^-B^+CDT^+$)	1	0	0	**1 (0.5)**
QX538 ($A^+B^+CDT^+$)	1	0	0	**1 (0.5)**
QX149 ($A^+B^+CDT^+$)	1	0	0	**1 (0.5)**
QX216 ($A^+B^+CDT^+$)	1	0	0	**1 (0.5)**
Total	**150**	**52**	**5**	**207**

Figure 1 Molecular epidemiology of *C. difficile* strain population
Dendrogram cluster analysis of 16s-23s rRNA intergenic spacer region (ISR) banding patterns for 16 unique *C. difficile* ST11 RTs analysed in this study (*n*=207). Origin: CLIN, clinical; VET, veterinary; ENV, environmental.

METHODS

C. difficile culture, PCR ribotyping, toxin gene PCR, antimicrobial susceptibility testing (MIC) and WGS (illumina) were performed as previously described[4,5]. Detailed methods for the following bioinformatic analyses are available in our recent publication[5]: *in silico* multi-locus sequence tying (MLST) and AMR profiling (ResFinder, BIGSdb, SRST2); core orthologous gene analysis (MAFFT, RAxML, GET_HOMOLOGUES, ClonalframeML,); core genome single nucleotide variant (SNV) analysis (SMALT, SAMtools, vcftools, SnpEff, Gubbins and RAxML); pan-genome analysis (Roary and PanGP); prophage analysis (PHASTER); and transposon characterisation (MUMmer, Artemis, EasyFig).

Intra- and interspecies transmission of globally disseminated ST11 clones provides compelling evidence that CDI may have a zoonotic or foodborne aetiology



Study code	Description
Isolate prefixes	
C001-C150	Clinical isolates of *C. difficile*, CLIN, *n*=150
E001-E005	Environmental isolates of *C. difficile*, ENV, *n*=5
V001-V052	Veterinary isolates of *C. difficile*, VET, *n*=52
Institution prefixes†	
HCF01-HCF62	Healthcare facilities/hospital sites, *n*=62
FRM01-FRM16	Farm/veterinary laboratories, *n*=16

Figure 2 Core genome SNV analysis for 5 major ST11 clusters
Core genome SNV analysis (1076 non-recombinant sites) provided ultra-fine scale resolution of this population, revealing 5 major evolutionary clusters largely congruent with RT lineage and toxin gene profile. This is in agreement with a second Maximum-Likelihood phylogeny based on 1276 core orthologous genes (data not shown).

The 5 clusters correspond to 3 internationally disseminated clusters of RT078.126 (**tree A**), RT033.288 (**tree B**), and RT127 (**tree C**), as well as 2 Australian clusters of RT127 (**trees D** and **E**).

Importantly, in each of the 5 evolutionary clusters, there was a distinct absence of subpopulations that were exclusively associated with human, porcine or bovine hosts, meaning there was a significant genetic overlap of clinical and veterinary isolates. Moreover, in all the major ST11 RT sub-lineages (trees A-E), SNV analysis identified multiple instances of intra-species and inter-species clonal groups (highlighted in grey), defined as 2 or more isolates differing by ≤2 SNVs in their core genome–a signature of recent transmission[1,6].

Many of these clonal groups comprised isolates spread over a vast geographic area (different states, countries, and continents), indicative of reciprocal long-range dissemination and possible zoonotic or foodborne transmission.

ST11 shows high levels of AMR mediated by DNA gyrase mutations and a diverse reservoir of clinically important transposable elements

Phenotype	% NS	Genotype: AMR gene(s)	N	Genomic context	Phenotype-genotype concordance
Intrinsic					
Fluoroquinolone resistance	27.6	*gyrA*	47	QRDR mutation: Thr82Ile	100%
		gyrB	6	QRDR mutations: Ser366Val, Ser416Ala, Asp426Asn, Glu466Val	
Acquired					
Tetracycline resistance	42.7	*tetM*	92	Tn*6190*	100%
		tet-44	6	Tn*6164*	
		tet-40	75	†*Megasphaera elsdenii* /*Streptococcus suis*	
		tetO	1	†*Streptococcus suis* /*Campylobacter jejuni*	
MLS_B resistance	38.9	*ermB*	27	Tn*6194*	36.1%
Aminoglycoside resistance	100‡	*aph3'-III-sat4A-ant6'-Ia*	82	†*Erysipelothrix rhusiopathiae*	N.A‡
		aac6-aph2'	4	†*Enterococcus faecalis*	
		aac6-Im-aph2'-Ib	2	†*Enterococcus faecalis* /*Escherichia coli*	
		ant6-Ib-ant9-Ia	6	Tn*6164*	
Glycopeptide resistance	0	*vanB2* operon	1	Tn*1549*-like (see **Fig. 3**)	0%

NS, non-susceptible; MLSB, macrolide-lincosamide-streptogramin-B phenotype; ‡*C. difficile* is inherently resistant to aminoglycosides; QRDR, quinolone resistance determining region; †no discernible transposons were identified, gene(s) shared 100% homology with synonymous loci in other species

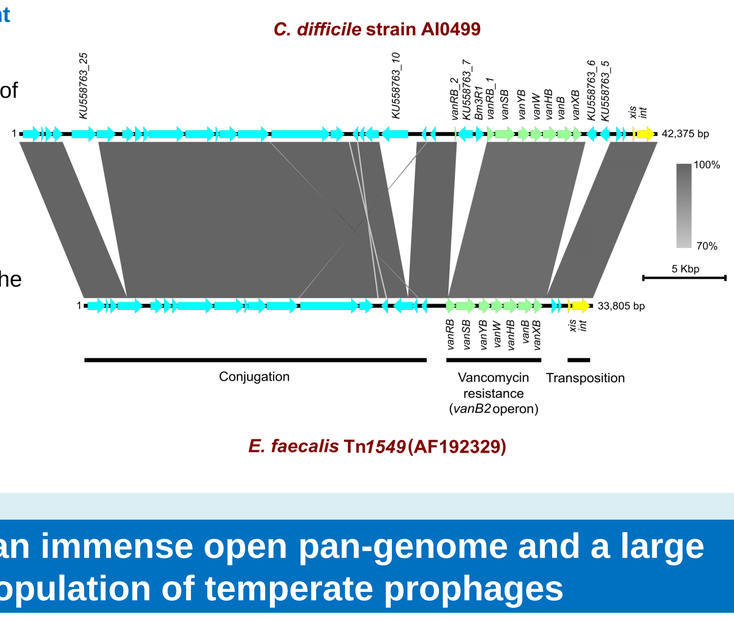
Table 1 AMR loci, underlying genomic context and associated phenotype
Multiple AMR loci were identified and found to be carried on numerous conjugative transposons, many of which are capable of transfer within *C. difficile* and between *C. difficile* and other species[1,7]. Non-synonymous mutations within *gyrA* and *gyrB* and the presence of tetR genes were both good predictors of antimicrobial phenotype. Conversely, the presence of *ermB* was a poor predictor of phenotype, indicating alternative mechanisms of MLS_B resistance are in play.

C. difficile ST11 strain AI0499 harbors a cryptic *vanB2* resistance operon, the first such report for this species

Figure 3 Genetic architecture and comparative analysis of Tn*1549*-like element
C. difficile AI0499 (RT033, $A^-B^-CDT^+$) was isolated from a neonatal calf at an abattoir in Australia in 2013[8]. AI0499 harboured a cluster of 7 syntenic vancomycin resistance genes with >99% seq ID to *vanXB*, *vanB*, *vanHB*, *vanW*, *vanYB*, *vanSB,* and *vanRB*.

Analysis of regions flanking the *vanB* gene cluster revealed a 43kbp element showing significant sequence identity and synteny with Tn*1549*, a conjugative transposon linked with the emergence and global dissemination of vancomycin-resistant enterococci[9].

Notably, AI0499 did not show any reduced susceptibility to vancomycin *in vitro* (MIC 1 mg/L), possibly due to an aberrant and non-functional *vanRB* gene (fragmented by *Bm3R1* and orf7). See Knight *et al.* 2016[10].



ST11 is characterised by an immense open pan-genome and a large heterogeneous population of temperate prophages

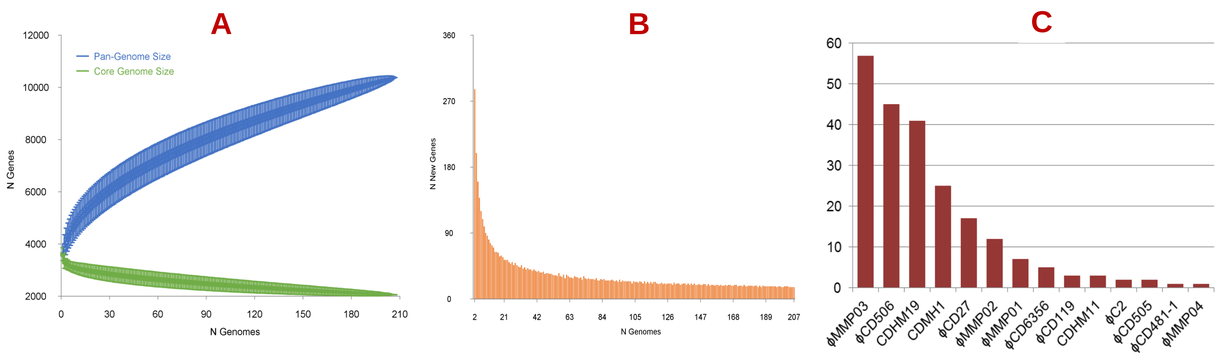


Figure 4 Pan-genome, core genome and prophage analysis for 207 *C. difficile* ST11 genomes
(A) The ST11 lineage is defined by a large 'open' pan-genome (10378 genes) comprising a core genome of 2058 genes (accounting for only 19.8% of the total gene repertoire) and an accessory genome of 8320 genes. **(B)** At *n*=207, an average of 16 new 'strain-specific' genes are contributed to the gene pool. **(C)** A total of 221 intact prophages were identified, comprising 14 different types, many of which are clinically important and have putative roles in the fitness and virulence of the host bacterium[1].

REFERENCES

1. Knight et al. 2015. Clin. Microbiol. Rev. 28: 721
2. Rupnik et al. 2008. J. Clin. Microbiol. 46: 2146
3. Hensgens et al. 2012. Clin. Microbiol. Infect. 18: 635
4. Knight et al. 2017. Front. Microbiol. 7:2138
5. Knight and Riley. 2016. J. Antimicrob. Chemother. 71: 2213
6. Eyre et al. 2013. N. Engl. J. Med. 369: 1195
7. Mullany et al. 2015. Res. Microbiol. 166: 361
8. Knight et al. 2016. Clin. Microbiol. Infect. 22: 266.e1-7
9. Howden et al. 2014. mBio. 4: e00412-13
10. Knight et al. 2016. mSphere. 1: e00177-16