Whole genome sequence analysis of *Clostridium difficile* ribotype 017 strains in hospital patients in Cape Town, South Africa

Brian R. Kullin[1], Daniel R. Knight[2], Sharon J. Reid[1], Valerie R. Abratt[1], Thomas V. Riley[2,3,4]

[1]Dept of Molecular and Cell Biology, University of Cape Town, Cape Town, South Africa;
[2]Medical, Molecular and Forensic Sciences, Murdoch University, Murdoch, Western Australia (WA);
[3]Edith Cowan University, Joondalup, WA; [4]PathWest Laboratory Medicine, Nedlands, WA

Introduction

- ***Clostridium difficile*** **PCR ribotype (RT) 017 was the predominant RT in patients attending hospitals in the Cape Town metropole in 2014-2015[1]**
- **Preliminary multilocus variable-number tandem-repeat analysis showed a high level of inter-strain relatedness, indicative of possible patient-to-patient transmission**
- ***In vitro*** **antimicrobial susceptibility testing revealed multidrug resistance for the majority of the isolates**

Aims

To carry out WGS analysis of *C. difficile* RT017 strains from Cape Town hospitals to examine strain transmission and the genetic basis of antimicrobial resistance (AMR)

Methods

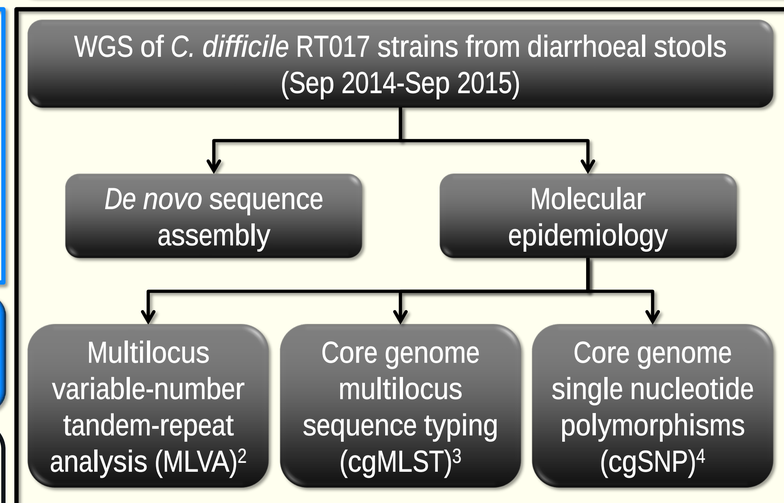

Figure 1: Flowchart of methods

Sample sites

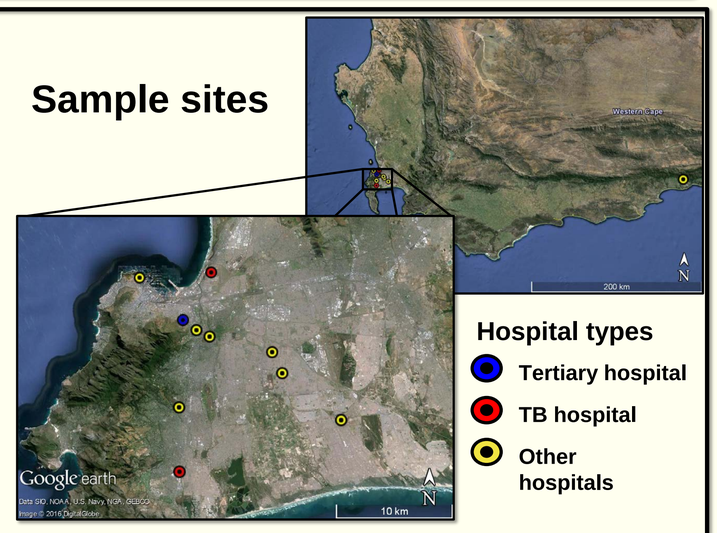

Figure 2: Locations of participating hospitals in Cape Town

Results

SA RT017 *C. difficile* strains harboured a number of AMR determinants

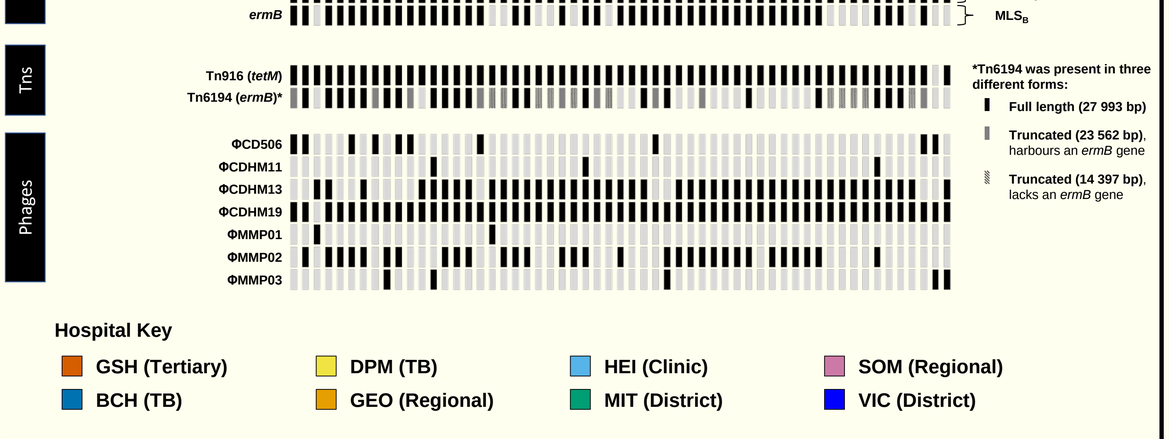

Figure 3: Antimicrobial resistance (AMR) loci and predicted intact prophages present in South African RT017 isolates. The presence and absence of AMR loci, transposons and phages is indicated in each case by black and grey bars respectively. **Note:** Three distinct versions of the Tn6194 transposon were present as indicated by the different bars.

High resolution profiling identified 182 high-quality biallelic SNPs

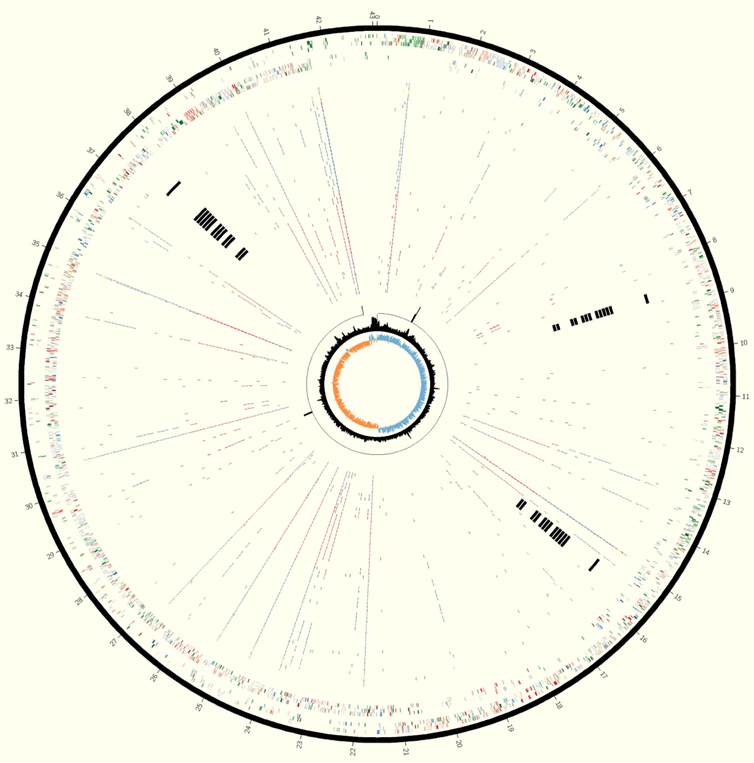

Figure 4: SNPs between *C. difficile* M68 and 57 South African RT017 isolates. The outermost ring represents position with respect to the M68 C. difficile reference genome. The next six rings represent annotated genes on the forward strand (first 3 rings) and the reverse strand (next 3 rings), coloured according to the respective COG function. The two innermost rings represent GC skew (<0% orange, >0% blue) and GC content (black histogram). The third ring represents SNP density in 10 kbp windows. The remaining 57 rings represent the location of SNPs in each strain, coloured according to hospital. Finally, regions of recombination identified by the Gubbins algorithm are represented by black blocks.

COG Key

Cellular processes and signaling
- [D] Cell cycle control, cell division, chromosome partitioning
- [M] Cell wall/membrane/envelope
- [N] Cell motility
- [O] Post-translational modification, protein turnover, and chaperones
- [T] Signal transduction mechanisms
- [U] Intracellular trafficking, secretion, and vesicular transport
- [V] Defense mechanisms
- [W] Extracellular structures
- [Y] Nuclear structure
- [Z] Cytoskeleton

Information storage and processing
- [A] RNA processing and modification
- [B] Chromatin structure and dynamics
- [J] Translation, ribosomal structure and biogenesis
- [K] Transcription
- [L] Replication, recombination and repair

Metabolism
- [C] Energy production and conversion
- [E] Amino acid transport and metabolism
- [F] Nucleotide transport and metabolism
- [G] Carbohydrate transport and metabolism
- [H] Coenzyme transport and metabolism
- [I] Lipid transport and metabolism
- [P] Inorganic ion transport and metabolism
- [Q] Secondary metabolites biosynthesis, transport, and catabolism

Poorly characterised
- [R] General function prediction only
- [S] Function unknown

High-resolution typing suggested intra- as well as inter-hospital strain transmission

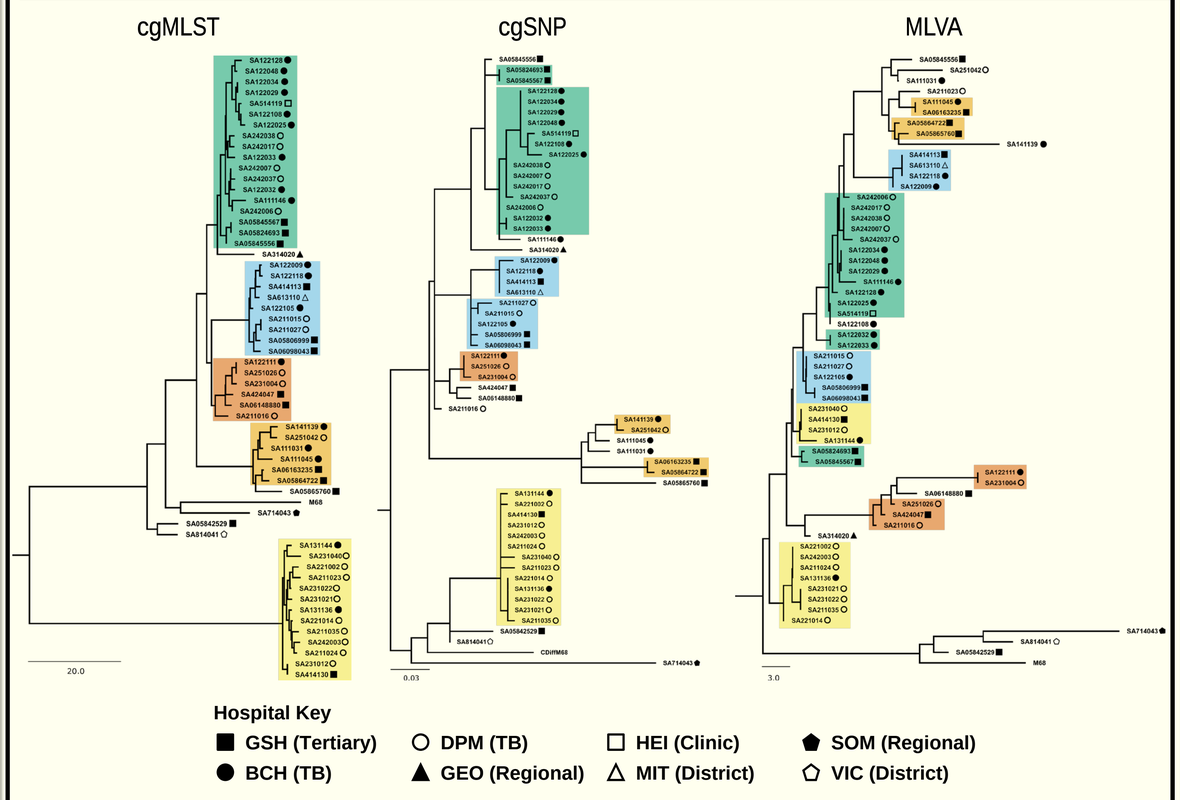

Figure 5: Clustering of isolates by cgMLST, cgSNP and MLVA. Shaded areas represent clusters of clonally-related strains identified by each method. Similar colours are used to facilitate visual comparison between methods. The following cutoffs were used to classify clonally-related strains: cgMLST ≤ 6 loci different, cgSNP ≤ 2 single nucleotide variants, MLVA ≤ 2 summed tandem repeat distance.

cgMLST and cgSNP typing methods were highly congruent

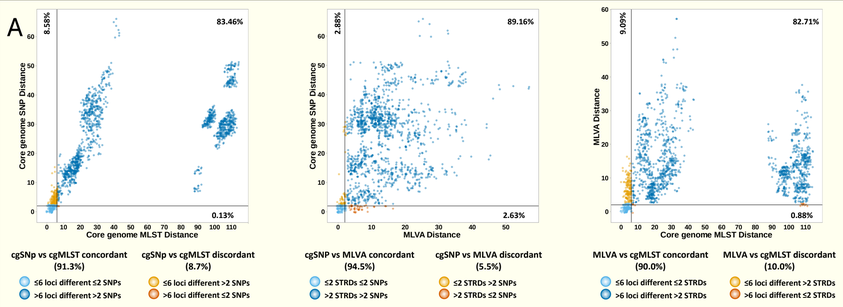

Figure 6: Concordance of high-resolution typing methods. (A) Points in each plot represent pairwise comparisons for every pair of isolates using each of the three typing methods. **(B)** Diversity [Simpson's index] and congruence measurements (adjusted non-directional [Rand] and directional [Wallace] coefficients).

B

Simpson's Index of Diversity

	Simpson's ID [95% CI]
cgMLST	0.814 [0.760-0.868]
cgSNV	0.88 [0.826-0.935]
MLVA	0.902 [0.858-0.947]

Adjusted Rand Coefficient

	cgMLST	cgSNV	MLVA
cgMLST		0.746 [0.587-0.916]	0.537 [0.367-0.719]
cgSNV	0.746 [0.587-0.916]		0.648 [0.465-0.841]
MLVA	0.537 [0.367-0.719]	0.648 [0.465-0.841]	

Adjusted Wallace Coefficient

	cgMLST	cgSNV	MLVA
cgMLST		0.595 [0.429-0.760]	0.395 [0.226-0.565]
cgSNV	1.000 [1.000-1.000]		0.582 [0.406-0.758]
MLVA	0.835 [0.783-0.886]	0.731 [0.579-0.882]	

Strains isolated from the same hospital were not more closely related than strains isolated from different hospitals

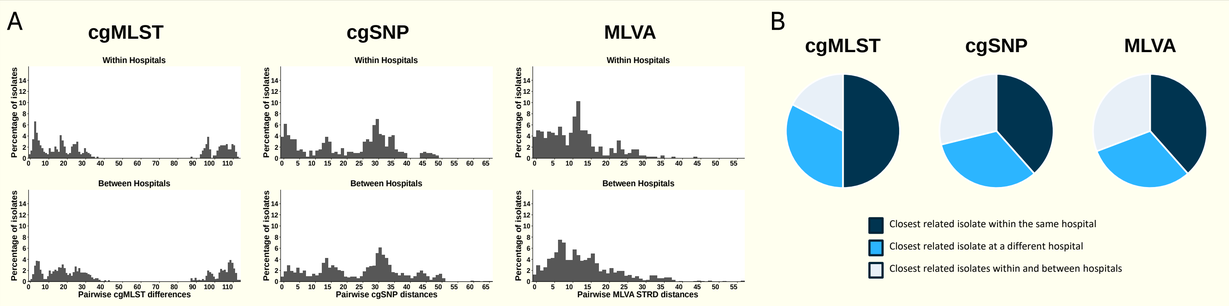

Figure 7: Pairwise comparisons between isolates. (A) Distances according to cgMLST, cgSNP and MLVA between every pair of isolates stratified according to whether the pair was isolated from the same ("Within Hospitals") or different ("Between Hospitals"). **(B)** Proportions of isolates for which the closest related strain was isolated from the same hospital, a different hospital or both.

Conclusions

This is the first study to carry out a detailed genomic analysis of South African *C. difficile* strains. The results suggest that:

1. **SA RT017 isolates harboured a diverse array of antimicrobial resistance determinants, which was correlated with phenotypic resistance to moxifloxacin, erythromycin, rifampicin and tetracycline**
2. **Clonally-related RT017 strains were circulating within patients attending public hospitals in the Cape Town area during the study period**
3. **WGS-based methods (cgSNP and cgMLST) provided similar levels of resolution when used to analyse RT017 strains, whereas MLVA had higher discriminatory power but was a poor predictor of clustering by the WGS-based methods**

References

1. **Kullin, B., J. Wojno, V. Abratt, S.J. Reid,** *Eur. J. Clin. Microbiol. Infect. Dis.* 36, 163-175 (2017)
2. **van den Berg, R.J., I. Schaap, K. E. Templeton, C. H. W. Klaassen, E. J. Kuijper,** *J. Clin. Microbiol.* 45, 1024-1028 (2007).
3. **Bletz, S., S. Janezic, D. Harmsen, M. Rupnik, A. Mellmann,** *J. Clin. Microbiol.* 56, e01987-17 (2018).
4. **Knight, D.R., M.M. Squire, D.A. Collins, T.V. Riley,** *Front. Microbiol.* 7, 2138 (2017)

Funding sources

National Research Foundation
Australia Awards
Carnegie Corporation of New York