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], Frederic 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. *Corresponding author

Clostridium 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. Increasing evidence of a shared ancestry and genetic overlap of RT078, the most common ST11 sublineage, between these populations suggests that CDI may have a zoonotic or foodborne aetiology. Whole genome sequencing and core genome phylogenetics were performed on a diverse 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 sublineages 078, 126, 127, 033 and 288). Phylogenies based on core orthologous genes (1276 loci) showed clustering of clinical and veterinary isolates indicative of very recent shared ancestry. Core genome single nucleotide variant (SNV) analysis provided ultra-fine scale resolution of this lineage, identifying multiple intra-species and inter-species clonal groups (isolates separated by ≤ 2 SNVs in their core genome) in all the major RT sublineages. Many clonal groups comprised isolates spread over a vast geographic area (different states, countries, and continents), indicative of reciprocal long-range dissemination and possible zoonotic/foodborne transmission. Antimicrobial resistance genotypes and phenotypes varied across host species, geographic regions and RTs, and included macrolide/lincosamide resistance (Tn*6194*; *ermB*), tetracycline resistance (Tn*6190*; *tetM* and Tn*6164*; *tet44*), fluoroquinolone resistance (*gyrA/B* mutations) as well as several aminoglycoside resistance cassettes. *C. difficile* ST11 is defined by a large 'open' pan- genome (10378 genes) comprising a core genome of 2058 genes (remarkably, accounting for only 19.8% of the total gene repertoire) and an accessory genome of 8320 genes containing a large and diverse collection of clinically important prophages of the Siphoviridae and Myoviridae. This study provides novel and critical insights on strain relatedness and genetic variability of *C. difficile* ST11, a lineage of significant One Health importance.