Article:
Vernon, J. J. orcid.org/0000-0002-0072-9294, Eyre, D. W., Wilcox, M. H. et al. (1 more author) (Accepted: 2025) Molecular clock complexities of Clostridioides difficile. Anaerobe. ISSN 1075-9964 (In Press)

Molecular clock complexities of *Clostridioides difficile*

Jon J. Vernon [1, 2]*, David W. Eyre [3,4,5], Mark H. Wilcox [1,6], Jane Freeman [1,6]

1. Healthcare-Associated Infections Research Group, Leeds Institute for Medical Research, University of Leeds, Wellcome Trust Brenner Building, St James University Hospital, Leeds, West Yorkshire, LS9 7TF

2. Division of Oral Biology, School of Dentistry, University of Leeds, Wellcome Trust Brenner Building, St James University Hospital, Leeds, West Yorkshire, LS9 7TF.

3. Big Data Institute, Nuffield Department of Population Health, University of Oxford, Oxford, OX3 7LF.

4. NIHR Oxford Biomedical Research Centre, University of Oxford, OX3 9DU

5. NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, University of Oxford, OX3 9DU

6. Microbiology, Leeds Teaching Hospitals Trust, Leeds, UK.

*corresponding author

Email: j.j.vernon@leeds.ac.uk

Tel: 0113 3437121

Abstract

Objectives

Reconstruct the phylogenetic status of a collection of historical *Clostridioides difficile* isolates and evaluate the congruence of their evolutionary trajectories with established molecular clock models.

Methods

Phylogenetic analysis was performed on Illumina sequence reads from previously analysed historic *C. difficile* isolates (1980–86; *n*=75) demonstrating multiple antimicrobial resistances. Data was grouped by ribotype (RT), including comparators from European surveillance (2012–13) and phylogenetic studies (1985–2010). Reads were mapped to CD630/CD196 reference genomes and compared using recombination-adjusted maximum likelihood trees. Prediction intervals for expected SNP differences by age were calculated using a Poisson distribution and molecular clock estimates (0.74 SNPs per genome/per year). Root-to-tip analysis was performed to determine the date of most common recent ancestor of genomes sharing a ribotype.

Results

Moxifloxacin-resistant (>16 mg/L) RT027 isolate JV67 (1986) was two SNPs distinct from a 2006 genome, fewer than the expected lower estimate (4.4 SNPs) under current molecular clock calculations; (p=3.93x10^{-5}). For isolate JV02 (1981), the 13 SNP divergence from a 2008 isolate was consistent with expectations (5.9 SNPs; p=0.07). JV73 (1983) demonstrated an 8 SNP difference, which although above the expected lower limit (5.5 SNPs), was outside the 95% prediction interval; (p=

4.51x10^{-3}). Only sixty-nine percent of historical genomes fit within the prediction interval for the number of SNPs expected compared to recent isolates, with fewer SNPs observed more frequently than expected. Root-to-tip analysis demonstrated only a weak linear correlation.

Conclusions

C. difficile molecular clock estimations may be more complex than previously considered, with periods of spore quiescence potentially complicating analyses.

Keywords

Clostridioides difficile, phylogenetics, antimicrobial resistance, fluoroquinolone, ribotype 027, bacterial evolution.

Introduction

Clostridioides difficile PCR ribotype (RT) 027 remains one of the most prevalent disease-causing strains across Europe (1, 2). This hyper-virulent type has been associated with multiple international outbreaks (3-5) and is synonymous with fluoroquinolone resistance (6, 7). The evolution of RT027 is of major interest to enable a greater understanding of the important genetic alterations that have led to the emergence of hyper-virulence. Phylogenetic analysis has suggested that this RT experienced a population expansion period around the turn of the century, with evidence of multiple horizontal gene transfer events (8). This was demonstrated through evidence identifying complementary single nucleotide polymorphisms (SNPs) between isolates with large evolutionary distances. Comparisons of whole genome sequences between modern, epidemic and "historical", non-epidemic 027 strains have revealed five large genomic regions of difference, indicative of recent acquisitions in evolutionary terms, due to their absence in older lineages (9). However, no genetic differences were identified in the pathogenicity locus between RT027 isolates from pre or post the 2003/4 North American outbreak (10). This lends greater weight to the argument that excess fluoroquinolone use and subsequent resistance in this RT was the major influential factor driving its emergence (11, 12). He *et al* reported the presence of two main fluoroquinolone resistant lineages (FQR1 and FQR2) for RT027, both acquiring Thr-82→Ile mutations in *gyrA* through separate evolutionary events. FQR1 originated in North-East USA, whereas FQR2 was more widespread across Canada and North America and was identified as the source of international dissemination of the original outbreak (10).

We previously reported the genotypic and phenotypic analysis of a historical catalogue of 75 *C. difficile* isolates from 1980-86, including the discovery of three moxifloxacin (MXF) resistant RT027 isolates (13), pre-dating reports of this important phenotype (from around the turn of the century (14, 15)) by approximately twenty years. Here we sought to further investigate this collection through phylogenetic analysis, with an aim to validate its age authenticity. Whilst investigating this catalogue we identified a series of findings which indicated a greater depth of complexities to the use of molecular clock predictions in *C. difficile*.

Material and Methods

Strain and sequence data

HiSeq 3000 sequencing (Illumina, USA) read files from our previously reported historical *C. difficile* collection originating from 1980-86 (13) were used for further phylogenetic analyses in this study. Forty-seven UK isolates, representing common RTs; 001 (*n*=6, 1981-1983), 002 (*n*=3, 1981-1983), 014 (*n*=4, 1981-1983), 015 (*n*=16, 1981-1984), 020 (*n*=13, 1980-1983), 027 (*n*=3, 1981-1986) and 078 (*n*=2, 1981-1982) were selected for analysis due to the availability of comparator sequences. Sequence data can be accessed under BioProject PRJNA704635. These isolates exhibited a range of antimicrobial resistances, including to moxifloxacin with all RT027 isolates deemed resistant (16-32 mg L^{-1}; Table 1) as previously described (13).

Molecular Typing

C. difficile isolates were genotyped by the UKHSA *C. difficile* Ribotyping Network (CDRN), Leeds, UK; based on the methods of Stubbs *et al*. (16), with the addition of capillary electrophoresis. Multi-locus variable number tandem repeat analysis (MLVA) was performed using the enhanced fingerprinting service of the CDRN (17).

Phylogenetic analysis

Next generation sequencing data of seven common PCR RT groups; 001, 002, 014, 015, 020, 027 and 078, was processed through an established bioinformatics pipeline created for the analysis of bacterial genomic data, briefly described below (18, 19). Additional comparator sequences from these RTs were included from the European, multicentre, prospective, biannual, point-prevalence study of *Clostridium difficile* infection in hospitalised patients with diarrhoea (EUCLID), (552 sequences [2012-2013], BioProject PRJNA398458) (20) and He *et al.* (10) (149 RT027 sequences [1985-2010], BioProject PRJEB2318 and accession numbers: FN668375, FN665652, FN665653, FN668944, FN668941-FN668943, FN545816, ERA000207, ERA000208, FN538970 and FN545816).

A SNP calling workflow was used based on a previously published methodology (19). Briefly, Illumina sequencing data was mapped to the *C. difficile* 630 reference genome (21), with the exception of RT027 (clade 2) isolates, which were mapped to CD196 (9). Mapping was performed with Stampy (v.1.0.11) (22), variants were identified using SAMtools mpileup (v.0.1.12-10) (23) and filtered requiring a read consensus of >75%, a heterozygous call under a diploid model, and a minimum coverage of five reads, including one read in each direction. Maximum likelihood phylogenetic trees were generated with PhyML (v3.1) (24), and adjusted for the impact of recombination using ClonalFrameML (v1.125) with default settings (25).

ClonalFrameML uses the sequence data and a statistical model to identify clusters of genetic variants likely to have been introduced by recombination. Each of the branch lengths of the phylogeny are then adjusted to only reflect variants likely to have been introduced by mutation, and not those arising through recombination. The Interactive Tree of Life (v.4.2) (26) was used to analyse subsequent data.

Prediction intervals for the number of SNPs expected given the time between isolates being obtained were calculated based on evolutionary rates (0.74 SNPs [95% Confidence Interval 0.22–1.4] per genome, per year) (19), assuming a Poisson distribution for SNPs. These were compared to observed SNP and age differences between nearest neighbouring taxa on maximum likelihood trees. Root-to-tip analysis was performed using RStudio 2024.04.02 (Posit, PBC). Briefly, the RT027 phylogenetic tree, rooted with the 1985 isolate from He *et al*. (10), was read into R using the 'ape' package. Root-to-tip distances were calculated with the 'adephylo' package and plotted against isolate ages. Linear regression analysis was performed, and the x-intercept was determined.

Results

Phylogenetic analyses

Phylogenetic analyses were performed by grouping RT sequences, mapping them to relevant reference genomes and creating maximum likelihood trees with comparator genomes. The purpose of this was to determine where the historical isolates fit within the existing knowledge of the evolutionary picture and confirm that the previously determined historic isolates with fluoroquinolone resistance matched their expected molecular clock predictions. Isolates recovered from the

historical catalogue (1980-86) generally formed clusters with at least one other strain from the same collection (Figure 1 & Figure 2), providing evidence to support the expected age of the original samples. The estimated number of SNPs from the closest "modern" isolate (EUCLID study, 2012-13) (1) was commonly identified as >13. Figure 1 visualises the tree for RT027 isolates with a colour gradient representing their ages and labelling reflecting place of origin. The three historical RT027 isolates, JV02 (1981), JV67 (1986) and JV73 (1983) did not cluster with each other. Instead, only eight and two SNPs distinguished JV73 (1983) and JV67 (1986) from genomes recovered in 2008 and 2006; (Figure 1, see expanded panels). This finding was the same with or without correction for recombination. The lower estimate of SNPs expected between JV67 (1986) and its nearest neighbour on the phylogenetic tree (from 2006) was 4.4 SNPs, therefore a finding of 2 SNPs was outside of the expected range based on current molecular clock estimations. For JV73 (1983), a minimum of 5.5 SNPs were expected from isolates 25 years apart, and 8 SNPs were observed. Using a Poisson model, we calculated the probability of observing these SNP counts (or less) over the time elapsed between samples as p=4.51 x 10^{-3} and p=3.93 x 10^{-5}, for JV73 (1983) and JV67 (1986) respectively. This means that if the model is true, the probability of observing the number of SNPs occurring within the time difference is significantly outside the predicted range. Isolate JV02 (1981) demonstrated closest relatedness to a UK isolate from 2008 with 13 SNP differences, which remained above the predicted lower estimate of 5.9 SNPs; (p=0.07) for the 27-year period between isolate collection. Figure 2 represents the maximum likelihood tree for RT001 strains from the historic collection (JV04 [1982], JV23 [1981], JV28 [1982], JV34

[1981], JV37 [1983] and JV63 [1983]) and how they fit with comparator strains from 2012-2013. This figure demonstrates how isolates JV37 (1983) and JV63 (1983) are phylogenetically distant from neighbouring genomes, exhibiting 6,696 and 6,229 SNP differences from isolates collected 30 years later in 2013 (Figure 2). These differences far exceed the upper limit of 42.0 SNPs predicted by current molecular clock estimations. In contrast, the 1981 (JV23 and JV34), and the 1982 isolates (JV04 and JV28), exhibited much lower SNP divergences, 12.7 and 25.3 SNPs, respectively. These values fall within the expected ranges for comparisons with neighbouring 2013 isolates (32-33 years apart), which are 7.0-44.8 SNPs for the 1981 isolates and 6.8-43.4 for the 1982 isolates; (Figure 2). Only partial concordance of the data was seen with previously reported rates of evolution. With differences between collection periods of ~30 years and an estimated rate of *C. difficile* evolution (~0.74 SNPs per genome, per year) (19), only 69% of neighbouring sequences on a phylogenetic tree fit within the 95% prediction interval of a Poisson distribution for the expected number of SNPs (Figure 3). There was an excess of pairs of sequences with SNP differences that were smaller than expected based on 0.74 SNPs per year (E.g. JV67 is 20 years from the nearest neighbour and therefore around 14.8 SNPs [4.4-28.0 95% Confidence Interval] would be expected, as opposed to the two observed).

Historical RT078 isolates JV14 (1982) and JV22 (1981) demonstrated identical genomes, representing six SNP differences from an Irish strain (2013) collected 32-33 years later; (Supplementary Figure 1). The lower estimates of expected SNPs for JV14 (1982) and JV22 (1981) were 6.8 and 7.0, respectively. Therefore,

prediction intervals for these differences were outside of the 95% probability projection, based on current molecular clock estimations; p=0.0003.

Root-to-tip analysis for RT027 revealed a weak linear correlation (R^2=0.08) between root-to-tip distance and isolate ages, with an inferred date for the most common recent ancestor for RT027 of 1946 (Figure 5).

Discussion

Phylogenetic analyses were conducted to assess the evolutionary status of *C. difficile* isolates recovered from a historical collection originated from the early 1980s. Harnessing the knowledge of the *C. difficile* molecular clock (19, 27), these genomes were evaluated for age authenticity through their relatedness to modern genomes. Since this type of analysis presents only a probability of relatedness, it revealed hidden complexities. In cases where distant relationships were identified, marked by thousands of SNPs separating the genomes (Figure 2), evolutionary distinctions could be inferred. However, most instances were not so definitive.

The unexpected discovery of MXF-resistant RT027 isolates from our previous study (13), pre-dating the agent's introduction raised questions regarding the accuracy and reliability of their reported ages. This resistance may have emerged due to exposure to earlier quinolones such as nalidixic acid or may have occurred spontaneously and been selected once fluoroquinolones were introduced. This could have been the result of resistance-determining mutations lying dormant only to be selected when faced with future antimicrobial challenges. This discovery is compelling as it demonstrates the powerful potential for antimicrobial resistance to exist as 'silent' mutations long before the widespread use of an antibiotic. The

implications are significant, with 'dormant' resistance determinants potentially undermining the efficacy of even newly developed drugs and posing significant challenges to future antimicrobial stewardship. To clarify the ancestral relationships of these historical isolates to modern *C. difficile* strains a phylogenetic analysis was performed. The RT027 comparator database was extensive, including isolates spanning over 30 years. However, comparable historical data for other RTs was unavailable at the time of analysis. Nonetheless, in some instances, considerable genomic distinction from modern genomes provided convincing conclusions (Figure 2).

Whilst spatial clustering and root proximity on phylogenetic trees offer valuable insights into genomic ancestry, the degree of SNP variation from the closest modern relative can offer additional confidence in determining the evolutionary age of an isolate. Prediction intervals were calculated to represent a probabilistic lower limit of evolution, based on current *C. difficile* molecular clock estimations (0.74 SNPs per genome, per year) (19, 27). The intervals reflect the likelihood of observing a certain number of SNPs or fewer over the years separating a genome from its nearest neighbour on the phylogenetic tree. It was observed that only 69% of genomes from the historical collection had SNP differences within a 95% prediction interval (Figure 3, Supplementary Figures 1-5), whereas 95% would have been expected if the evolutionary rate and model assumptions were accurate. This suggests that some genomes may have evolved more slowly than anticipated, potentially due to periods of dormancy in a quiescent spore state. This indicates that, although probabilistic lower limits of evolution are based on predictions which can be adjusted iteratively as new data become available, spore dormancy

periods make accurate predictions challenging. Whilst this study is not sufficiently powered to re-calculate the evolutionary rate of *C. difficile*, the increasing availability of genomic data for this organism offers the opportunity to re-examine these predictions through future study. Nonetheless, the influence of spore dormancy remains poorly understood, making it difficult to account for.

Unexpectedly, the three historical RT027 isolates clustered with modern genomes of EUCLID study strains (Figure 1). Combined with fluoroquinolone resistance pre-dating previously reported time scales (10), these isolates did not align with the expected characteristics of early RT027 strains. Whilst the closest relative of isolate JV02 (isolated in the UK in 1981) was a 2008 Glasgow isolate, the prediction interval probability for 13 or fewer SNPs occurring over 27 years was p=0.07. This observed SNP difference would fall into the expected range (95% prediction interval) of the Poisson distribution of the molecular clock, fitting current *C. difficile* evolutionary understanding. Nonetheless, isolates JV67 (1986) and JV73 (1983), separated by only eight and two SNPs respectively, conflicted with the notion that they were greater than 25 years older than comparator genomes. Further analysis of branches comprising older isolates (dating back to 1985) indicated the assimilation of some modern strains, denoting close ancestries. Near the root of the phylogram (Figure 4), isolates >20 years apart revealed differences of ~8 SNPs, while others spanning even more years showed zero SNP differences where a minimum of 3.5 or 2.4 SNPs would be expected (e.g. 1991-2007 and 1993-2004), falling outside of the 95% prediction interval. This suggests greater complexity in temporal phylogenetic analyses. Root-to-tip analysis revealed weak correlations between root-to-tip distances and isolate ages of publicly available

RT027 genomes, whether or not strains from this study were included (Figure 5). The addition of these historical isolates shifted the most common recent ancestor date from 1962 to 1946. The weak correlation suggested that prolonged dormancy in a quiescent spore form may introduce significant uncertainty into *C. difficile* evolutionary estimates. Although the data suggests these isolates were similar to contemporary strains, other outliers like the 1990 and 1992 isolates (Figure 5) further illustrate this complexity.

The complex nature of individual *C. difficile* RT evolutionary trees may be partially attributable to the sporulation of this bacterium (28). Since genetic replication halts when *C. difficile* exists in the dormant spore form, evolution is significantly slower in these spore-forming organisms (29, 30). Establishing how long a specific strain has remained in this state is challenging, complicating the determination of precise temporal evolutionary links. Whilst it is assumed that the historical isolates existed in spore form for over 30 years, the duration that modern strains have spent in this state is essentially unknown. *C. difficile* can be harboured as spores asymptomatically in humans and animals (31) and can persist in the environment for extended periods (32). This indeterminant period of dormancy contributes to the difficulty of accurately dating isolates through phylogenetic analysis. Superdormant spores have been described in *Bacillus* spp., which could be associated with thermotolerance (33, 34). A similar concept has also been proposed in *C. difficile* (35). Such a possibility may be expected to be associated with a particularly slow rate of DNA mutations. Longitudinal studies could provide further insights into how spore dormancy impacts molecular clock estimations.

The notable observation of two clusters of identical genomes, spanning 11 and 16 years, appeared to contradict the current understanding of *C. difficile* evolution. Closer analysis revealed a possible explanation: all isolates clustered with zero SNP differences (Figure 4) originated from Arizona, with one isolate reportedly from 1991 clustering amongst strains from 2006/2007 (8). These genomes were sequenced in a study by Songer *et al.*, which identified RT027 in retail meat (36). A follow-up study from the same group (37) used MLVA to show that many of these isolates were indistinguishable, indicating probable specimen contamination involving the 1991 isolate. These findings underscore the need for caution when interpretating external data sets.

The previous discovery of three historical RT027 isolates exhibiting MXF resistance before the drug's market introduction in 1999 (13) added new layers to the understanding of fluoroquinolone resistant lineages. While the acquisition of the well-studied pathogenicity locus is known to occur through mobile genetic elements (38), the MXF resistance observed in these isolates is conferred by a single point mutation (Thr-82→Ile) which can spontaneously arise. Although minimal quinolone pressure would have been present historically to drive such mutations, this SNP has been demonstrated to provide a fitness advantage (39). An associated increase in growth rate would provide competitive benefit and indicate a possible reason for evolutionary retention of a spontaneous mutation. Ultimately, these isolates reveal additional complexity to the current understanding of FQR1 and FQR2 lineages (8), with potential alternative lineages originating from Europe.

To minimise the risks that the RT027 isolates in this study originated from modern sources, repeat attempts were made to re-isolate the same strains from the

312 original specimen tubes and unopened samples associated with the same patient.
313 RT027 strains were consistently isolated on three occasions and compared to the
314 original finding using MLVA, with all subsequent strains proving indistinguishable
315 from the original (data not shown). To further validate the results, MLVA was
316 applied to confirm that these historical strains were distinct from 633 RT027
317 isolates analysed by the CDRN enhanced fingerprinting service. No matches were
318 observed, indicating no direct connection to any of the 633 strains processed over
319 the previous eight years. Furthermore, isolates JV67 (1986) and JV73 (1983)
320 demonstrated distinctly different profiles from any other (>10 SNPs).

321 Although it is unlikely, based on these thorough investigations, that the historical
322 collection's isolates were modern, this possibility cannot be excluded entirely.
323 Nonetheless, this does not account for the overall picture that only 69% of pairs fit
324 within 95% prediction intervals, where 95% would be expected.

325 This study describes the use of phylogenetic analyses to assess the age
326 authenticity of a historic collection of *C. difficile* isolates, including some of the
327 earliest known RT027 MXF-resistant strains. It demonstrates that whilst most
328 phylogenetic analyses conform to current molecular clock estimations, some
329 isolates fell outside these expectations. Therefore, these findings suggest that *C.*
330 *difficile* spore dormancy might have a greater influence on evolutionary rates than
331 previously recognized.

Acknowledgements

Elements of this manuscript have been presented at the European Congress of Clinical Microbiology and Infectious Diseases, Madrid, Spain, 2018 (#P0386).

We acknowledge the contribution of Dr Chris Randall, University of Leeds, for his support with genomic assemblies using CLC Genomics Workbench.

The authors are also thankful to Professor Peter Borriello, Public Health England laboratories, Colindale, UK, for the initial isolate collection.

Source of Funding

This work was supported by a Leeds Institute of Biological and Clinical Sciences, University of Leeds, PhD studentship.

Conflicts of Interest

JF and MHW are supported in part by the National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre (BRC) (NIHR213331). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.
All other authors: none to declare.

Contributions of Authors

352 **Jon Vernon**: Conceptualization, methodology, formal analysis, investigation, data
353 curation, writing – original draft, writing – review & editing, visualization. **David**
354 **Eyre**: methodology, formal analysis, investigation, writing – review & editing,
355 visualization. **Mark Wilcox**: Conceptualization, supervision, writing – review &
356 editing, funding acquisition. **Jane Freeman**: Conceptualization, methodology, formal
357 analysis, writing – review & editing, supervision, funding acquisition.

References

1. Davies KA, Ashwin H, Longshaw CM, Burns DA, Davis GL, Wilcox MH. Diversity of Clostridium difficile PCR ribotypes in Europe: results from the European, multicentre, prospective, biannual, point-prevalence study of *Clostridium difficile* infection in hospitalised patients with diarrhoea (EUCLID), 2012 and 2013. Euro Surveill. 2016;21(29).

2. Freeman J, Vernon J, Morris K, Nicholson S, Todhunter S, Longshaw C, et al. Pan-European longitudinal surveillance of antibiotic resistance among prevalent *Clostridium difficile* ribotypes. Clin Microbiol Infect. 2015;21(3):248 e9- e16.

3. Loo VG, Poirier L, Miller MA, Oughton M, Libman MD, Michaud S, et al. A predominantly clonal multi-institutional outbreak of *Clostridium difficile*–associated diarrhea with high morbidity and mortality. N Engl J Med. 2005;353(23):2442-9.

4. McDonald LC, Killgore GE, Thompson A, Owens RC, Kazakova SV, Sambol SP, et al. An epidemic, toxin gene-variant strain of *Clostridium difficile*. N Engl J Med. 2005;353(23):2433-41.

5. Kuenzli AB, Burri S, Casanova C, Sommerstein R, Buetti N, Seth-Smith HMB, et al. Successful management of a *Clostridioides difficile* ribotype 027 outbreak with a lean intervention bundle. J Hosp Infect. 2020;106(2):240-5.

6. Freeman J, Vernon J, Pilling S, Morris K, Nicholson S, Shearman S, et al. The ClosER study: results from a three-year pan-European longitudinal surveillance of antibiotic resistance among prevalent *Clostridium difficile* ribotypes, 2011–2014. Clin Microbiol Infect. 2017.

7. Drudy D, Kyne L, O'Mahony R, Fanning S. *gyrA* mutations in fluoroquinolone-resistant *Clostridium difficile* PCR-027. Emerg Infect Dis. 2007;13(3):504-5.

8. He M, Sebaihia M, Lawley TD, Stabler RA, Dawson LF, Martin MJ, et al. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. Proc Natl Acad Sci U.S.A. 2010;107(16):7527-32.

9. Stabler RA, He M, Dawson L, Martin M, Valiente E, Corton C, et al. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol. 2009;10(9):R102.

10. He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, Martin MJ, et al. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. Nat genet. 2013;45(1):109-13.

11. Freeman J, Bauer MP, Baines SD, Corver J, Fawley WN, Goorhuis B, et al. The changing epidemiology of *Clostridium difficile* infections. Clin Microbiol Rev. 2010;23(3):529-49.

12. Pepin J, Saheb N, Coulombe MA, Alary ME, Corriveau MP, Authier S, et al. Emergence of fluoroquinolones as the predominant risk factor for *Clostridium difficile*-

associated diarrhea: A cohort study during an epidemic in Quebec. Clin Infect Dis. 2005;41(9):1254-60.

13. Vernon J, Wilcox MH, Freeman J. Antimicrobial resistance progression in the United Kingdom: A temporal comparison of *Clostridioides difficile* antimicrobial susceptibilities. Anaerobe. 2021;70:102385.

14. Hoogkamp-Korstanje JAA, Roelofs-Willemse J. Comparative *in vitro* activity of moxifloxacin against Gram-positive clinical isolates. J Antimicrob Chemother. 2000;45(1):31-9.

15. Wilcox MH, Fawley W, Freeman J, Brayson J. In vitro activity of new generation fluoroquinolones against genotypically distinct and indistinguishable *Clostridium difficile* isolates. J Antimicrob Chemother. 2000;46(4):551-6.

16. Stubbs SLJ, Brazier JS, O'Neill GL, Duerden BI. PCR targeted to the 16S-23S rRNA gene intergenic spacer region of *Clostridium difficile* and construction of a library consisting of 116 different PCR ribotypes. J Clin Microbiol. 1999;37(2):461-3.

17. Fawley WN, Freeman J, Smith C, Harmanus C, van den Berg RJ, Kuijper EJ, Wilcox MH. Use of highly discriminatory fingerprinting to analyze clusters of *Clostridium difficile* infection cases due to epidemic ribotype 027 strains. J Clin Microbiol. 2008;46(3):954-60.

18. De Silva D, Peters J, Cole K, Cole MJ, Cresswell F, Dean G, et al. Whole-genome sequencing to determine transmission of *Neisseria gonorrhoeae*: an observational study. Lancet Infect Dis. 2016;16(11):1295-303.

19. Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, et al. Diverse sources of *C. difficile infection* identified on whole-genome sequencing. N Engl J Med. 2013;369(13):1195-205.

20. Eyre DW, Davies KA, Davis G, Fawley WN, Dingle KE, De Maio N, et al. Two Distinct Patterns of *Clostridium difficile* Diversity Across Europe Indicating Contrasting Routes of Spread. Clin Infect Dis. 2018;67(7):1035-44.

21. Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, et al. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat Genet. 2006;38(7):779-86.

22. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 2011;21(6):936-9.

23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

24. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003;52(5):696-704.

25. Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput Biol. 2015;11(2):e1004041.

26. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44(Web Server issue):W242-W5.

27. Didelot X, Eyre DW, Cule M, Ip CL, Ansari MA, Griffiths D, et al. Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. Genome Biol. 2012;13(12):R118.

28. Elliott B, Androga GO, Knight DR, Riley TV. *Clostridium difficile* infection: Evolution, phylogeny and molecular epidemiology. Infect, Genet Evol. 2017;49:1-11.

29. Weller C, Wu M. A generation‐time effect on the rate of molecular evolution in bacteria. Evolution. 2015;69(3):643-52.

30. Pearson T, Busch JD, Ravel J, Read TD, Rhoton SD, U'ren JM, et al. Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. Proc Natl Acad Sci U.S.A. 2004;101(37):13536-41.

31. Crobach MJT, Vernon JJ, Loo VG, Kong LY, Péchiné S, Wilcox MH, Kuijper EJ. Understanding *Clostridium difficile* Colonization. Clin Microbiol Rev. 2018;31(2).

32. Fekety R, Kim K-H, Brown D, Batts DH, Cudmore M, Silva J. Epidemiology of antibiotic-associated colitis: Isolation of *Clostridium difficile* from the hospital environment. Am J Med. 1981;70(4):906-8.

33. Ghosh S, Setlow P. Isolation and characterization of superdormant spores of Bacillus species. J Bacteriol. 2009;191(6):1787-97.

34. Ghosh S, Setlow P. The preparation, germination properties and stability of superdormant spores of *Bacillus cereus*. J Appl Microbiol. 2010;108(2):582-90.

35. Rodriguez-Palacios A, Lejeune JT. Moist-heat resistance, spore aging, and superdormancy in *Clostridium difficile*. Appl Environ Microbiol. 2011;77(9):3085-91.

36. Songer JG, Trinh HT, Killgore GE, Thompson AD, McDonald LC, Limbago BM. *Clostridium difficile* in retail meat products, USA, 2007. Emerg Infect Dis. 2009;15(5):819.

37. Marsh JW, Tulenko MM, Shutt KA, Thompson AD, Weese JS, Songer JG, et al. Multi-locus variable number tandem repeat analysis for investigation of the genetic association of *Clostridium difficile* isolates from food, food animals and humans. Anaerobe. 2011;17(4):156-60.

38. Dingle KE, Elliott B, Robinson E, Griffiths D, Eyre DW, Stoesser N, et al. Evolutionary history of the *Clostridium difficile* pathogenicity locus. Genome Biol Evol. 2014;6(1):36-52.

39. Vernon JJ, Wilcox MH, Freeman J. Effect of fluoroquinolone resistance mutation Thr-82→Ile on *Clostridioides difficile* fitness. J Antimicrob Chemother. 2018;74(4):877-84.

40. Davies KA, Longshaw CM, Davis GL, Bouza E, Barbut F, Barna Z, et al. Underdiagnosis of *Clostridium difficile* across Europe: the European, multicentre, prospective, biannual, point-prevalence study of *Clostridium difficile* infection in hospitalised patients with diarrhoea (EUCLID). Lancet Infect Dis. 2014;14(12):1208-19.

			Minimum Inhibitory Concentration (mg L^{-1})						
Strain ID	**Specimen Date**	**PCR Ribotype**	**VAN**	**MTZ**	**FDX**	**MXF**	**CLI**	**CIP**	**ERY**
JV04	1982	001	1	0.125	0.015	2	16	16	<4
JV23	1981	001	4	1	0.008	32	64	>64	>256
JV28	1982	001	1	0.25	0.03	1	16	32	<4
JV34	1981	001	4	0.5	0.03	16	4	>64	>256
JV37	1983	001	1	0.25	0.06	2	8	16	<4
JV63	1983	001	0.5	0.125	0.015	1	2	16	<4
JV29	1981	002	1	0.125	0.125	1	4	16	<4
JV31	1981	002	1	0.125	0.03	2	8	16	<4
JV41	1983	002	1	0.125	0.03	2	8	16	<4
JV25	1981	014	1	0.25	0.06	1	32	16	16
JV33	1982	014	0.5	0.25	0.06	1	2	16	<4
JV40	1983	014	0.5	0.125	0.06	0.125	2	16	<4
JV69	1981	014	0.5	0.125	0.015	1	2	16	<4
JV03	1982	015	0.5	0.06	0.03	1	8	32	<4
JV07	1982	015	0.5	0.125	0.06	1	4	32	<4
JV09	1981	015	1	0.125	0.015	1	4	32	<4
JV10	1982	015	1	0.125	0.03	1	4	32	<4
JV11	1982	015	1	0.25	0.03	2	16	32	<4
JV17	1982	015	1	0.125	0.03	2	>64	32	<4
JV26	1981	015	1	0.125	0.06	1	8	32	<4
JV36	1983	015	1	0.25	0.06	2	8	16	<4
JV42	1983	015	0.5	0.125	0.03	2	4	16	<4
JV43	1984	015	1	0.25	0.125	2	2	16	<4
JV45	1983	015	1	0.125	0.03	2	1	16	<4
JV54	1982	015	0.5	0.125	0.015	1	2	16	<4
JV55	1982	015	0.5	0.125	0.015	1	0.25	16	<4

JV61	1983	015	0.5	0.125	0.03	1	2	16	<4
JV62	1983	015	0.5	0.125	0.015	1	2	16	<4
JV66	1982	015	0.5	0.125	0.015	1	1	16	<4
JV05	1982	020	0.25	0.06	0.06	1	4	32	<4
JV06	1982	020	0.25	0.06	0.015	1	2	32	<4
JV12	1982	020	0.5	0.25	0.06	2	16	16	<4
JV13	1982	020	0.5	0.25	0.06	2	4	32	<4
JV15	1983	020	0.5	0.25	0.06	2	16	32	<4
JV27	1981	020	0.5	0.25	0.06	1	8	32	<4
JV35	1983	020	1	0.25	0.125	2	8	32	<4
JV44	1983	020	0.5	0.25	0.125	2	2	16	<4
JV48	1980	020	0.5	0.125	0.008	1	2	16	<4
JV51	1981	020	0.5	0.125	0.015	1	2	16	<4
JV52	1981	020	0.5	0.125	0.015	1	4	16	<4
JV53	1981	020	0.5	0.125	0.015	1	4	16	<4
JV64	1983	020	0.5	0.125	0.015	1	2	16	<4
JV02	1981	027	0.5	1	0.06	32	8	>64	>256
JV67	1986	027	0.5	1	0.06	16	2	>64	>256
JV73	1983	027	0.5	1	0.06	16	2	>64	>256
JV14	1982	078	0.5	0.125	0.004	1	>64	16	<4
JV22	1981	078	1	0.125	0.03	2	>64	16	<4

Table 1: Demographics of the historical isolates used in phylogenetic analyses, including age, PCR ribotype and antimicrobial susceptibilities against seven comparator antimicrobials taken from Vernon et al (2021). (13) *VAN-vancomycin, MTZ-metronidazole, FDX-fidaxomicin, MXF-moxifloxacin, CLI-clindamycin, CIP-ciprofloxacin, ERY-erythromycin.*

Figure 1: Phylogenetic tree representation of "historical" ribotype 027 isolates amongst >350 comparator 027 strains from 1985–2013, acquired from the published He et al. (10) and the EUCLID studies. (20, 40) *Maximum likelihood trees were estimated as previously (19), and generated with the Interactive Tree of Life. (26) Colour gradient represents isolate age. The expanded panels show the clusters of isolates adjacent to JV02 (1981) and JV67 (1986; red box) and JV73 (1983; purple box). The number of SNPs distance and the year and location of neighbouring isolates is also shown. Supplementary Table 1 provides links between the tip labels to the original metadata.*

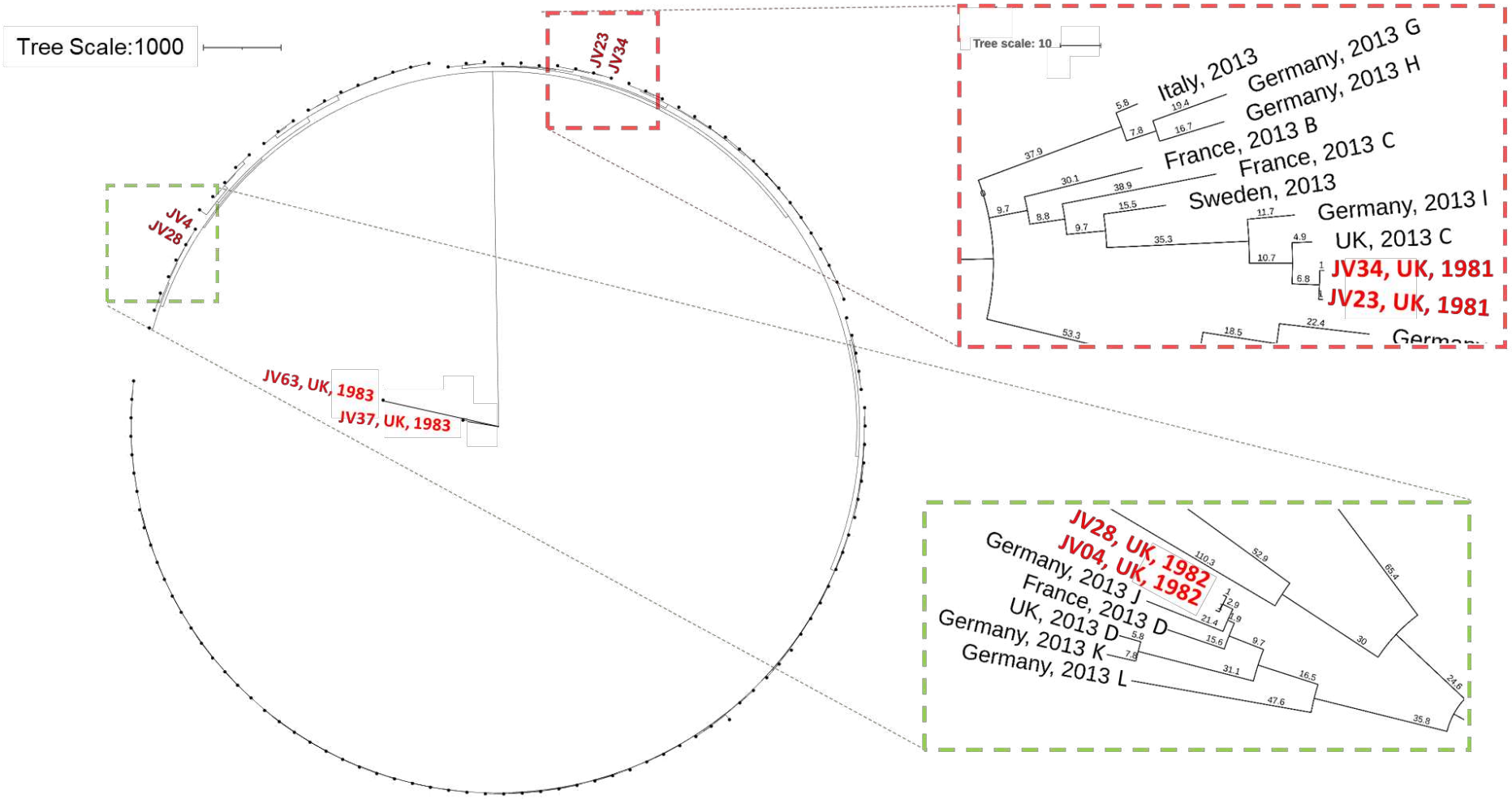


Figure 2: Phylogenetic tree representation of "historical" ribotype 001 isolates amongst comparator 001 strains from 2013, acquired from the EUCLID study. (20, 40)*Maximum likelihood trees were estimated as previously, (19) and generated using the Interactive Tree of Life. (26) The expanded panels show the clusters of isolates adjacent to JV23 (1981) and JV34 (1981; red box) and JV04 (1982) and JV28 (1982; green box). The number of SNPs distance and the year and location from neighbouring isolates is also shown. Supplementary Table 1 provides links between the tip labels to the original metadata.*

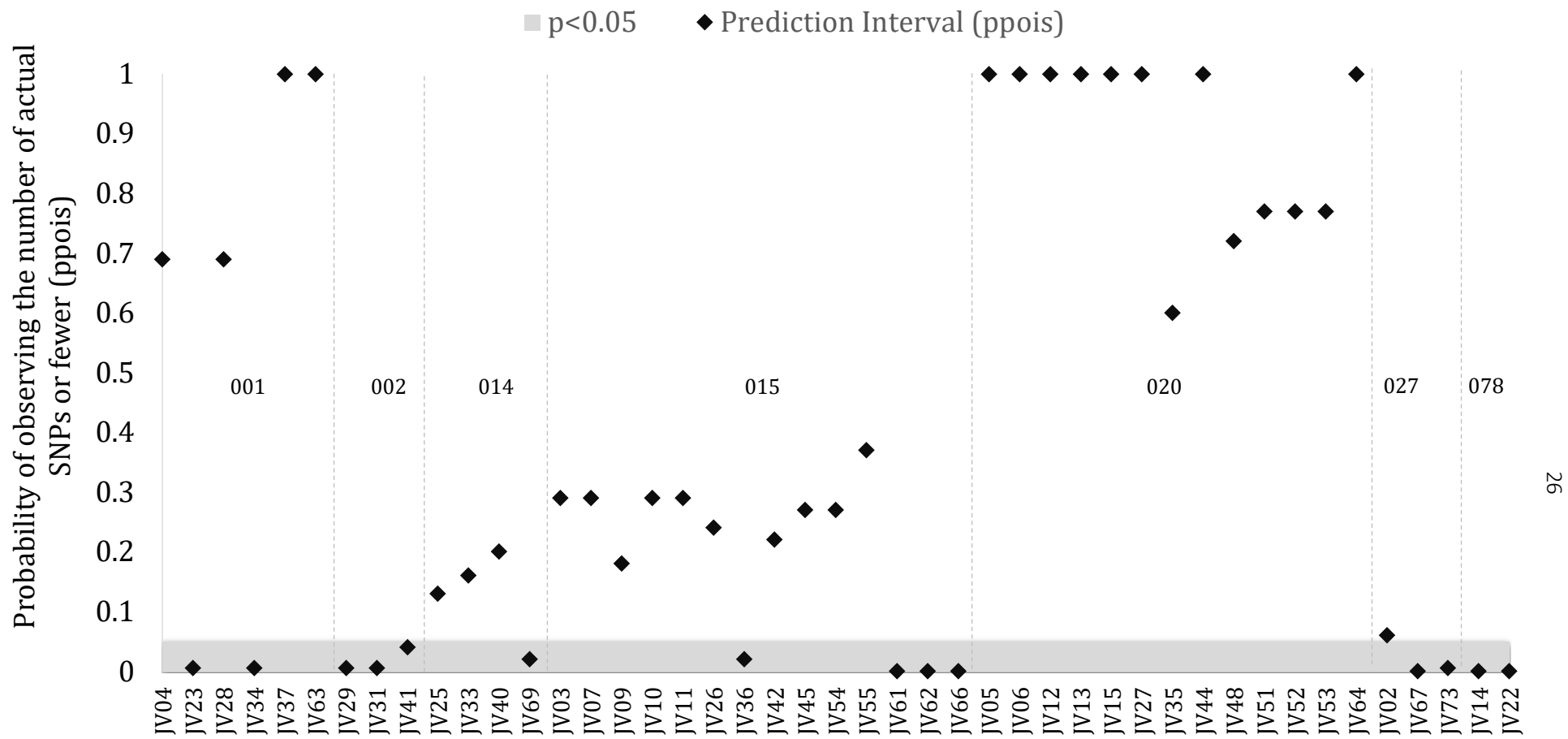


Figure 3: Prediction intervals representing the probabilities of the observed number of single nucleotide polymorphisms (SNPs) or fewer occurring in the time period separating neighbouring taxa, based on Poisson distribution of Clostridioides difficile molecular clock estimations (0.74 SNPs per genome, per year). (19) *The numeric labels on the x-axis refer to PCR ribotype groups, the y-axis is the probability of the expected number of SNPs occurring in the given time, from 0 (impossible) to 1 (certain).*

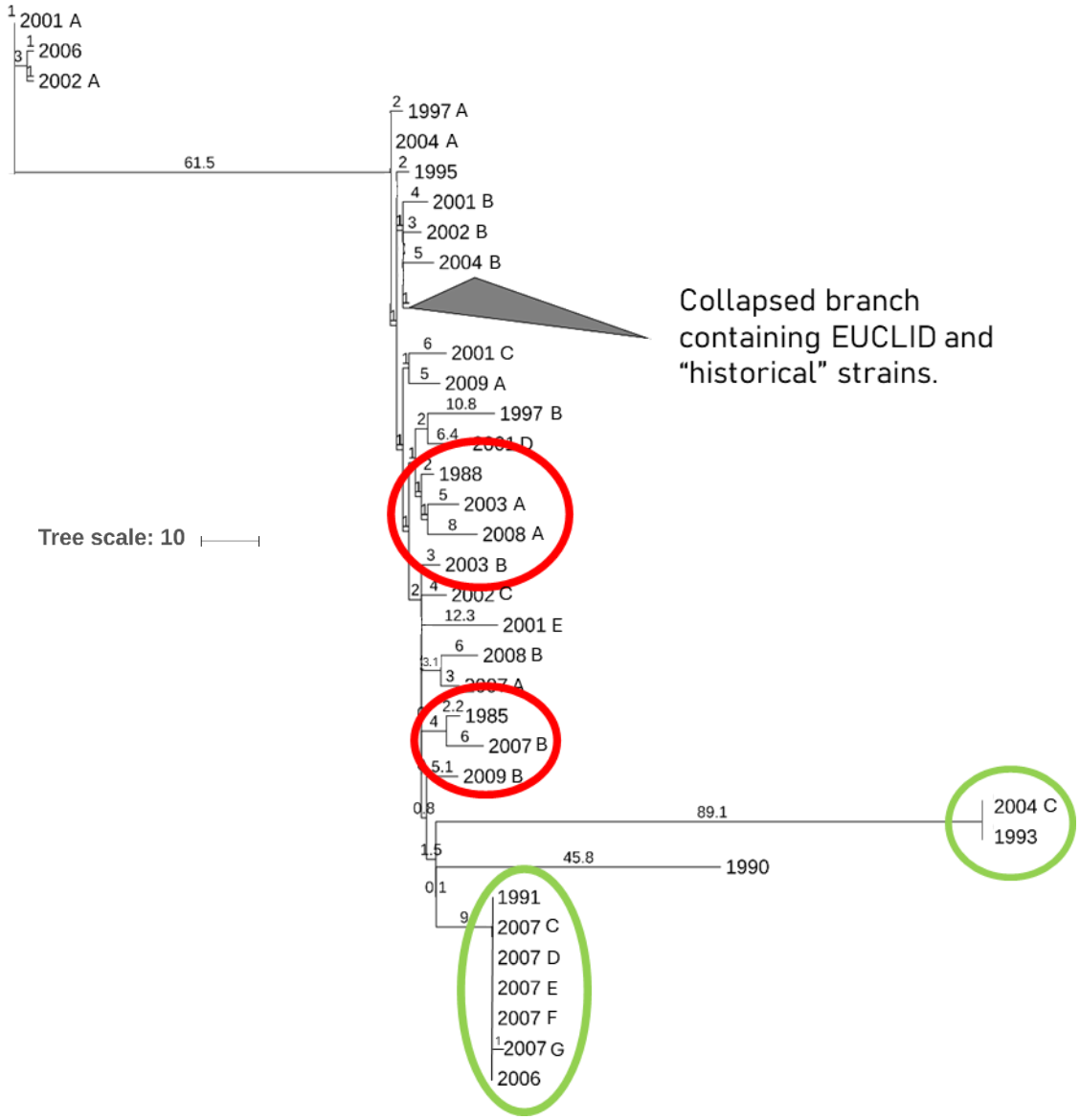
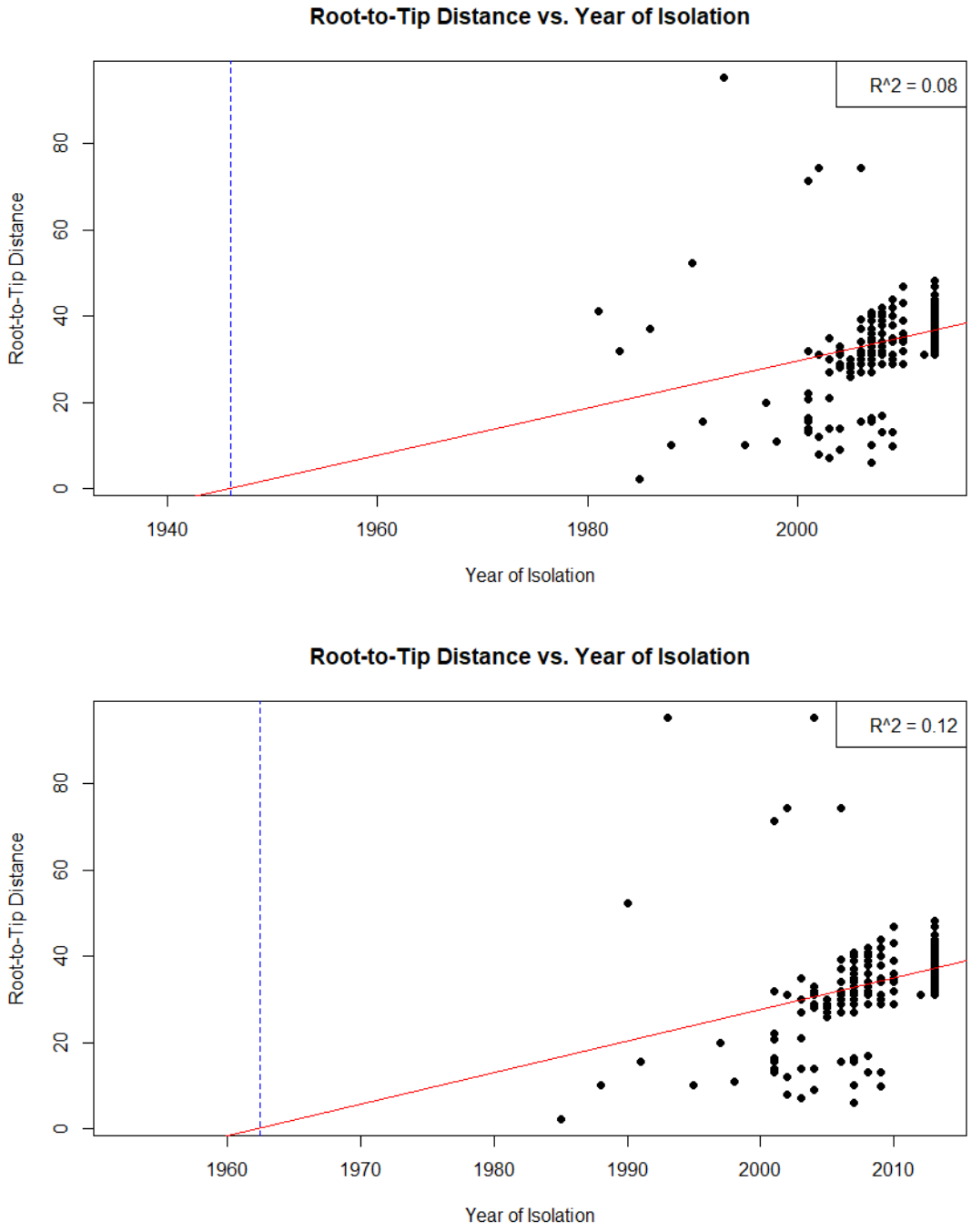


Figure 4: Expanded phylogenetic tree of ribotype 027 genomes from 1985–2009, acquired from the published He et al. study (10) and EUCLID (European, multicentre, prospective, biannual, point-prevalence study of Clostridium difficile infection in hospitalised patients with diarrhoea) studies (20, 40). *Red circles highlight branches of closely related strains originating between up to 20 years apart. Green circles highlight taxa clusters consisting of temporally distant genomes, identified as zero SNPs different. Maximum likelihood trees were estimated as previously, (19) and generated with the Interactive Tree of Life. (26) The numbers on the branches refer to the number of SNPs between isolates. Supplementary Table 1 provides links between the tip labels to the original metadata.*



511

512 ***Figure 5: Root-to-tip distance versus year of isolation analysis for ribotype (RT)***
513 ***027 isolates from the He et al (10) and EUCLID (20, 40) studies.*** *Top: Plot including*
514 *the three RT027 isolates from this study, JV02 (1981), JV67 (1986) and JV73 (1983).*
515 *The x-intercept is 1946. Bottom: Plot excluding the isolates from this study. The x-*
516 *intercept is 1962.*

517