Novel insights into *C. difficile* biology and evolution revealed by whole genome sequencing

Kate Dingle, Oxford University

Oxford Biomedical Research Centre
Enabling translational research through partnership

NHS
National Institute for Health Research

Introduction

• Large numbers of newly available microbial genomes are enhancing understanding of pathogen biology and evolution.

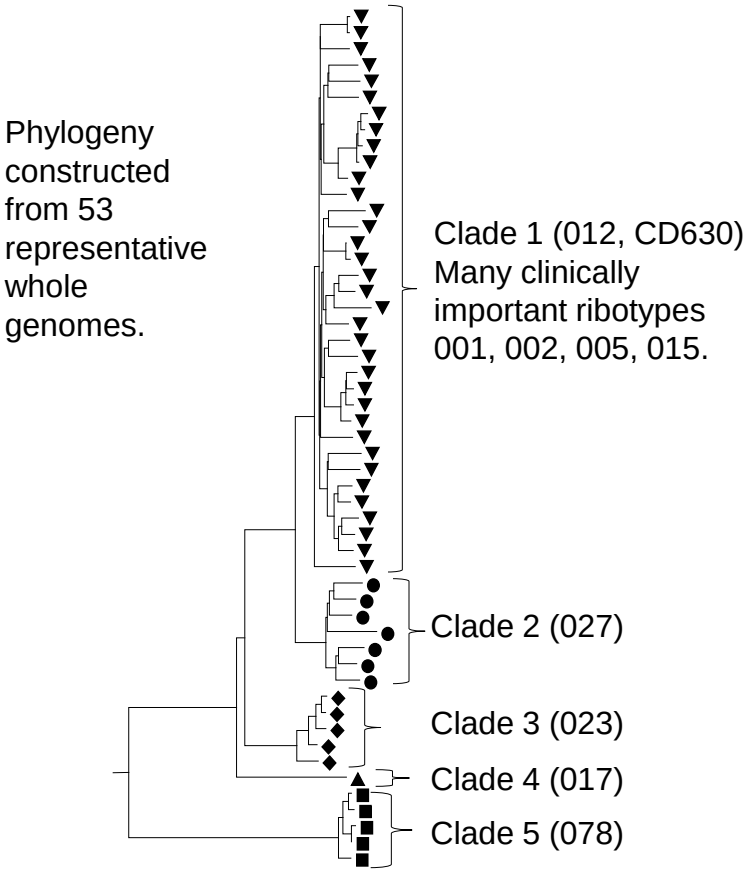
• Loci of interest can now be studied using long contiguous sequences determined without prior knowledge of gene content or sequence.

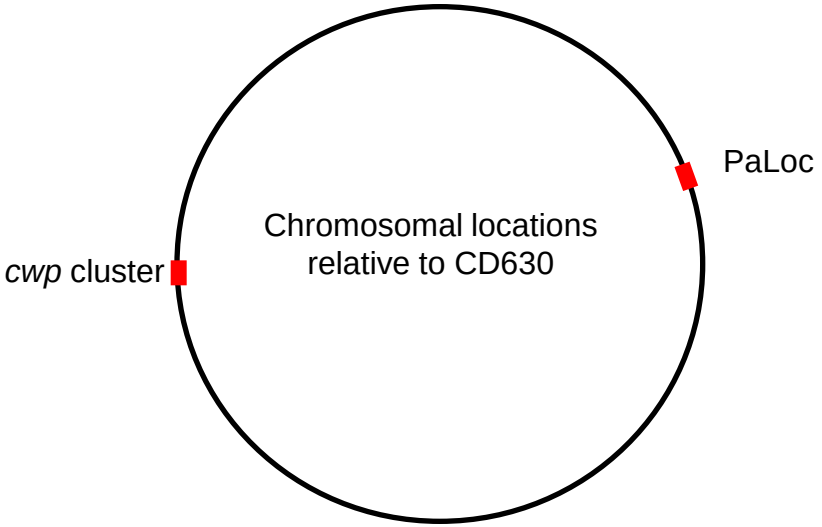
• Careful choice of isolates representing the population structure and different clinical phenotypes enhances this approach.

• WGS derived from a large collection of clinical isolates from Oxford and Leeds, UK, and Australia (n>2000) were used.

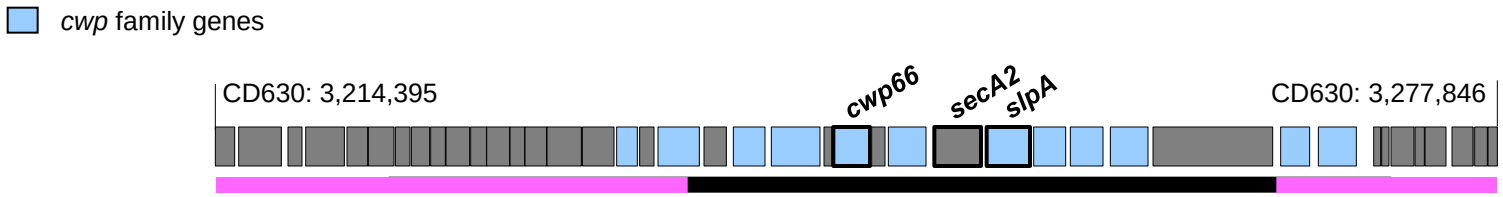
Two clinically important regions of the *C. difficile* genome were investigated with respect to the population structure

- **36.6kb cell wall protein (*cwp*) cluster** containing the *slpA* gene encoding the S-layer; immunodominant antigen, determines serotype.
- **19kb Pathogenicity locus (PaLoc)** encoding toxins responsible for the symptoms of CDI.



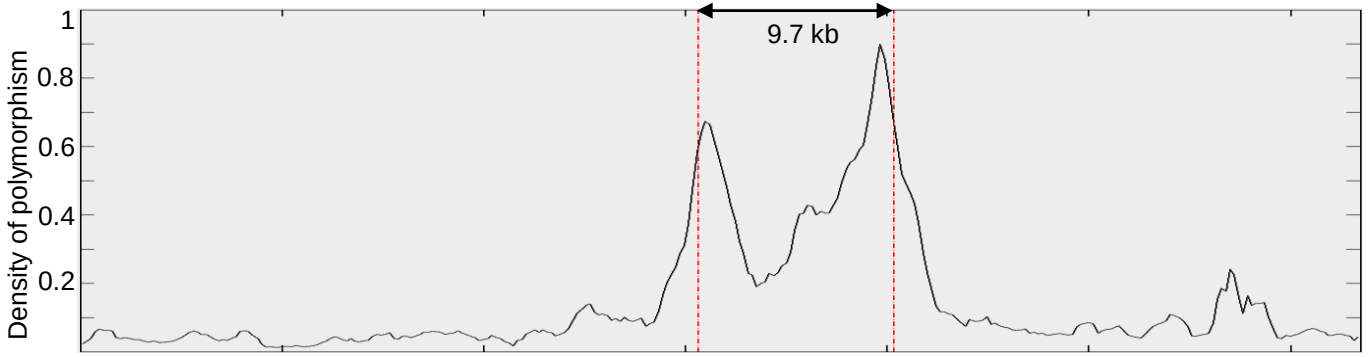


Genetic organisation in and around the *cwp* cluster.

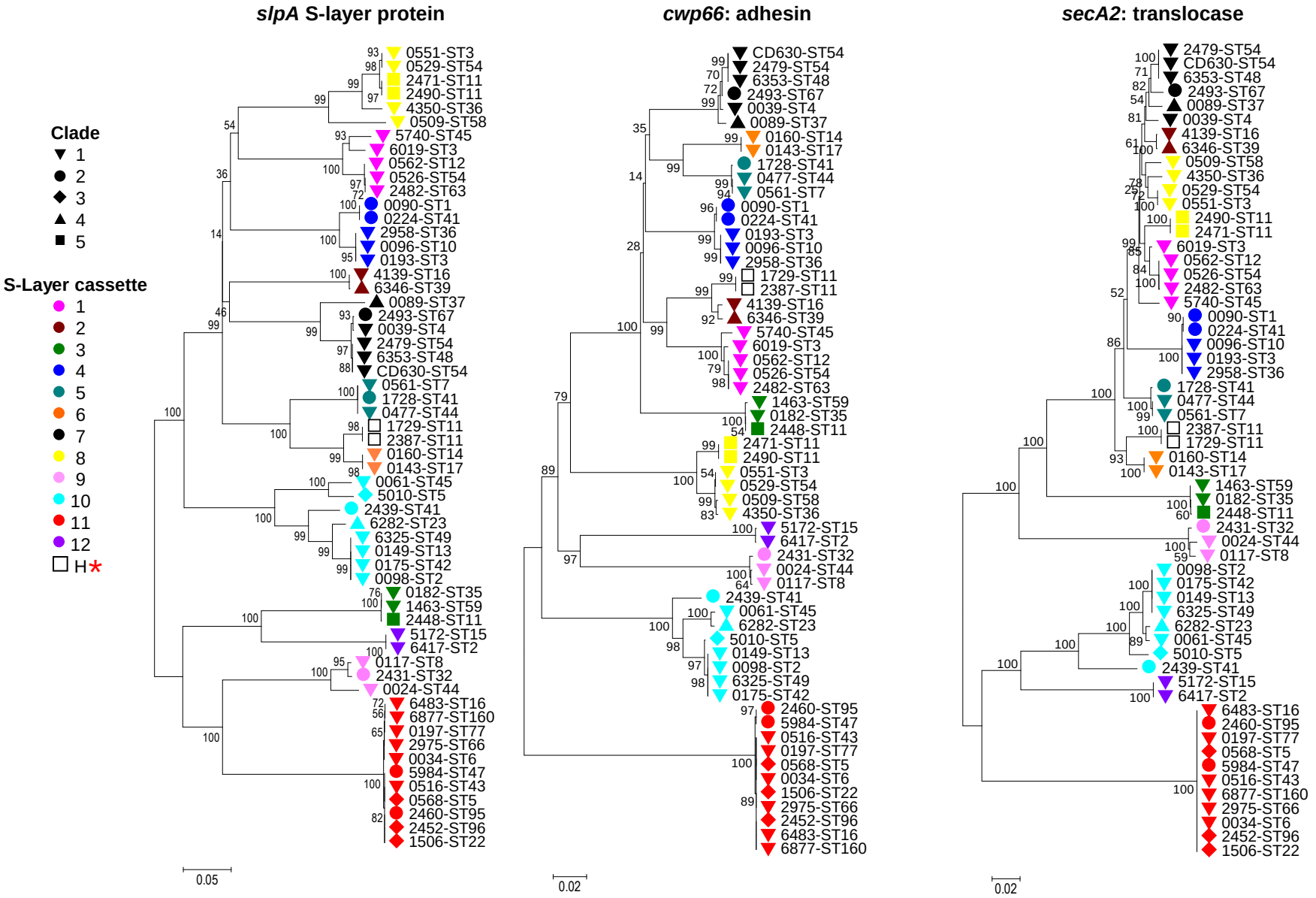


cwp cluster flanking sequences (pink) follow the five clades phylogeny.

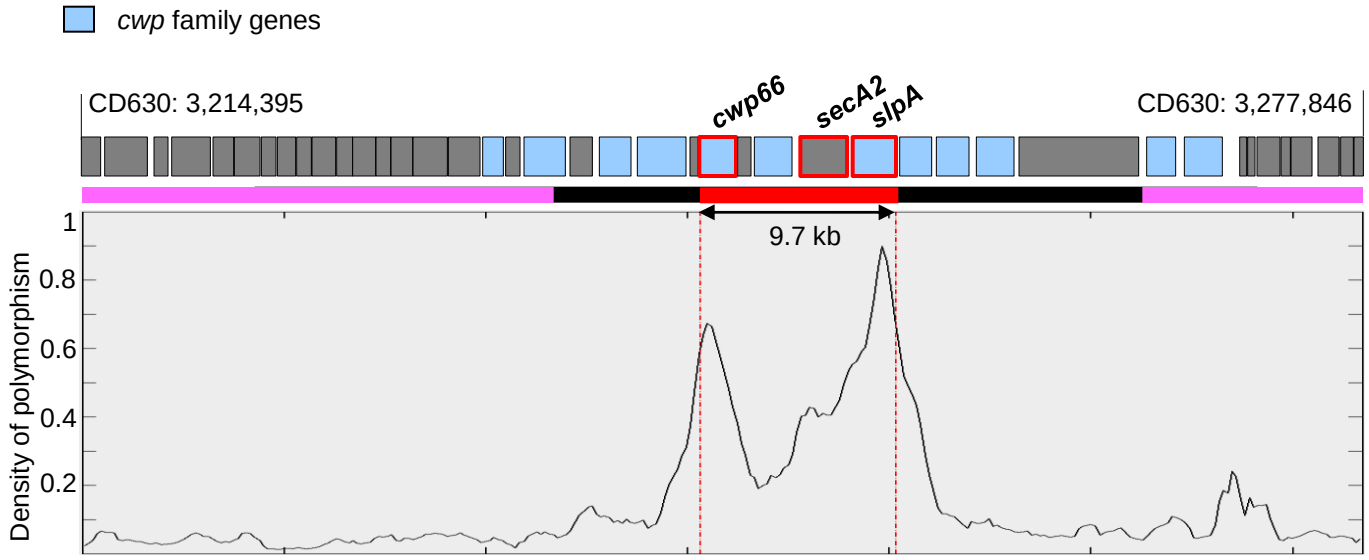
Density of polymorphism across the *cwp* cluster.



Phylogeny of *slpA*, *cwp66* and *secA2*



A 10kb S-layer cassette

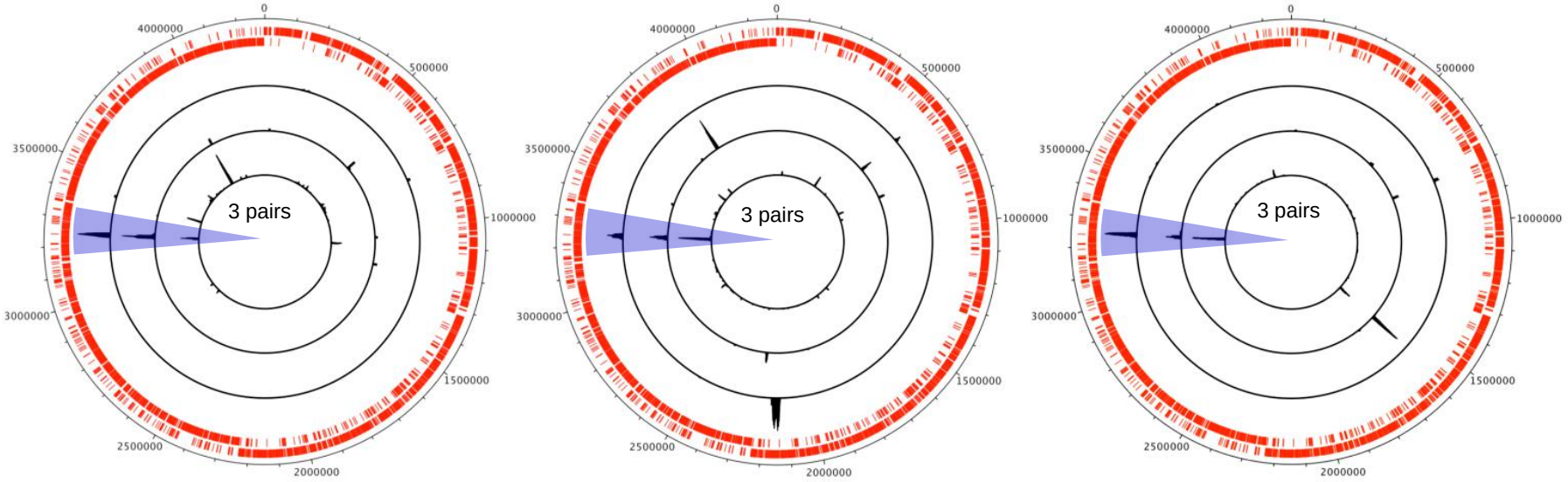


~10kb cassette:

- Linkage of *cwp66*, *secA2*, *slpA*.
- Coincides with peak in genetic diversity.
- Suggests frequent genetic exchange of the cassette involving “switching” by homologous recombination.

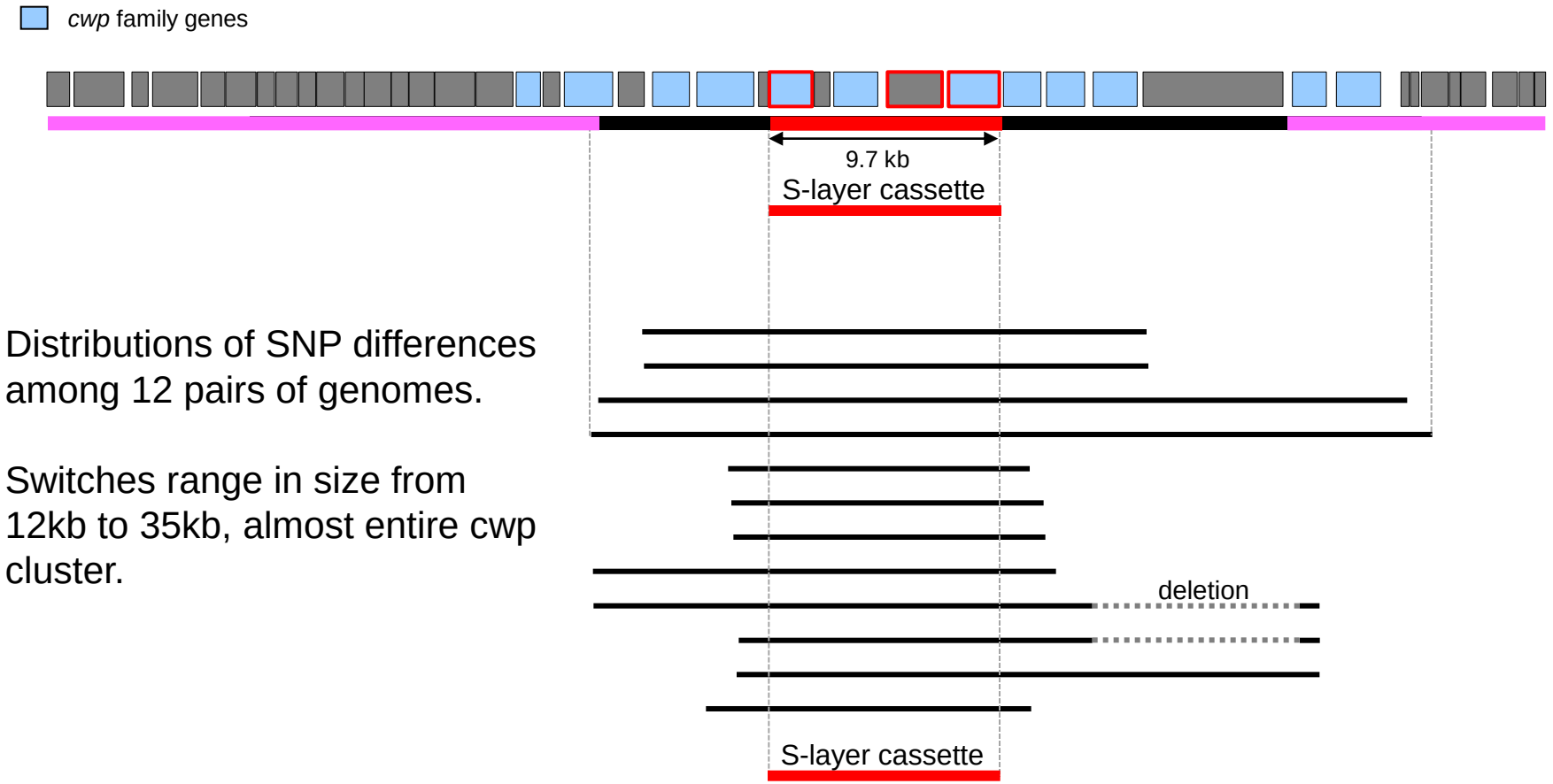
Recombinational switching of the S-layer cassette at the chromosomal scale

Red lines indicate the ORFs on the forward and reverse strands of CD630.



- Each pair of genomes chosen to be very closely related, but with distinct S-layer cassettes.
- Plotted whole-genome distributions of SNPs between the nine pairs
- Blue highlighting indicates the location of the *cwp* cluster within the chromosome.

Size of recombination events involving *cwp* cluster

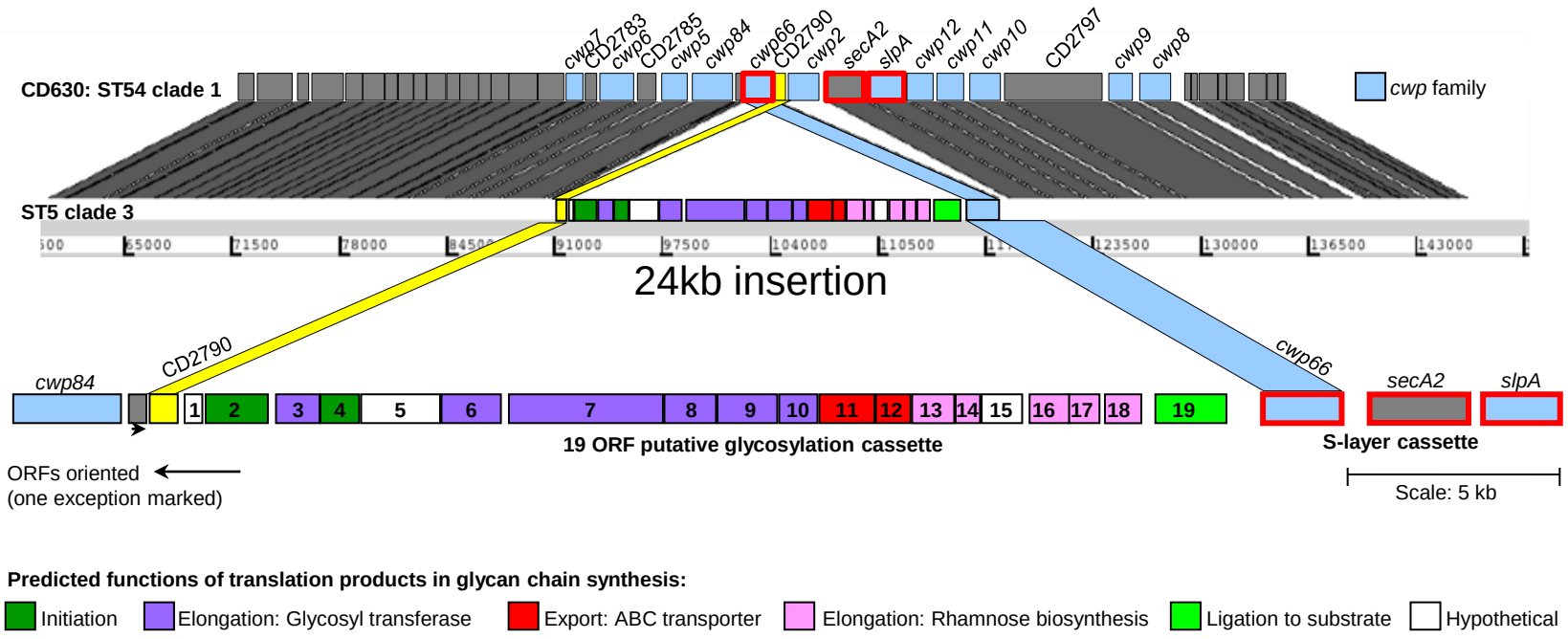


Distributions of SNP differences among 12 pairs of genomes.

Switches range in size from 12kb to 35kb, almost entire cwp cluster.

Switches size constrained to maintain linkage of genes in S-layer cassette.

Putative S-layer glycosylation cluster within one of the S-layer cassettes



Cassette found in clade 3 (023, ST5), and some members of clades 1 and 2. Always occurs with the same *slpA*, *cwp66* and *secA2* variants as part of an expanded S-layer cassette.

Summary

S-layer switching

- Changes S-layer “serotype” while the rest of the genome remains the same.

- Immune avoidance may contribute to temporal changes in clinically important *C. difficile* genotypes.

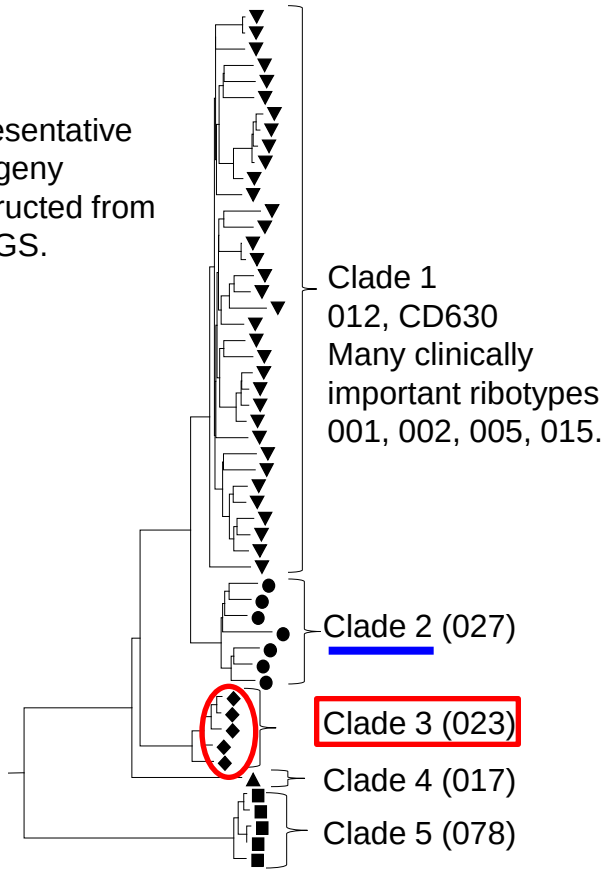
- Demonstrates the need for data on both S-layer cassette and genotype as neither can predict the other.

- Is relevant to vaccine development.

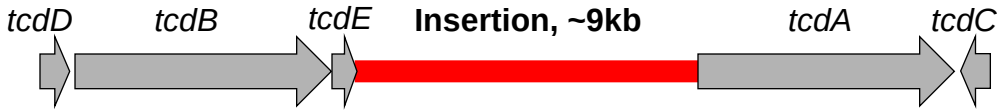
Just accepted by JID, shortly available online.

Cross population comparison of PaLoc sequences

Representative phylogeny constructed from 53 WGS.

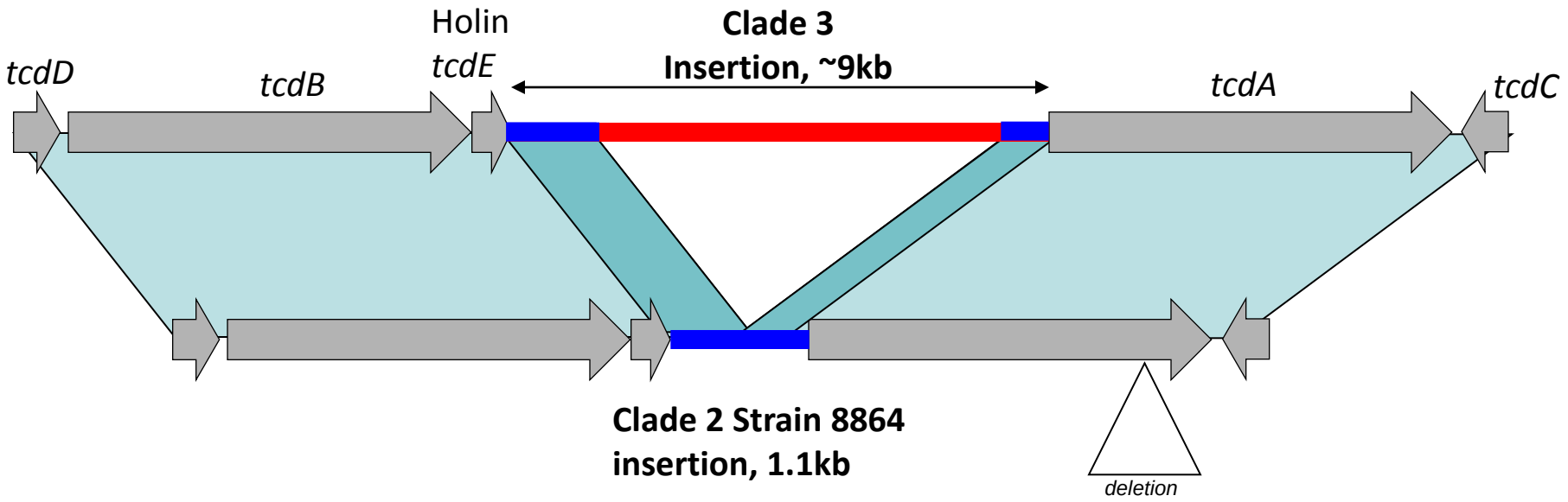


- PaLocs as expected.
- Except novel ~9kb insertion in Clade 3 (023).
- Present in all clade 3 UK and Australian isolates studied (n~100) representing five STs: unlikely a recent insertion.



- One PaLoc containing an insertion in this position described previously.
- Clade 2 (ST62, 036) strain 8864.
- Insertion was smaller, at 1.1kb.

Clade 3 PaLoc vs clade 2 8864

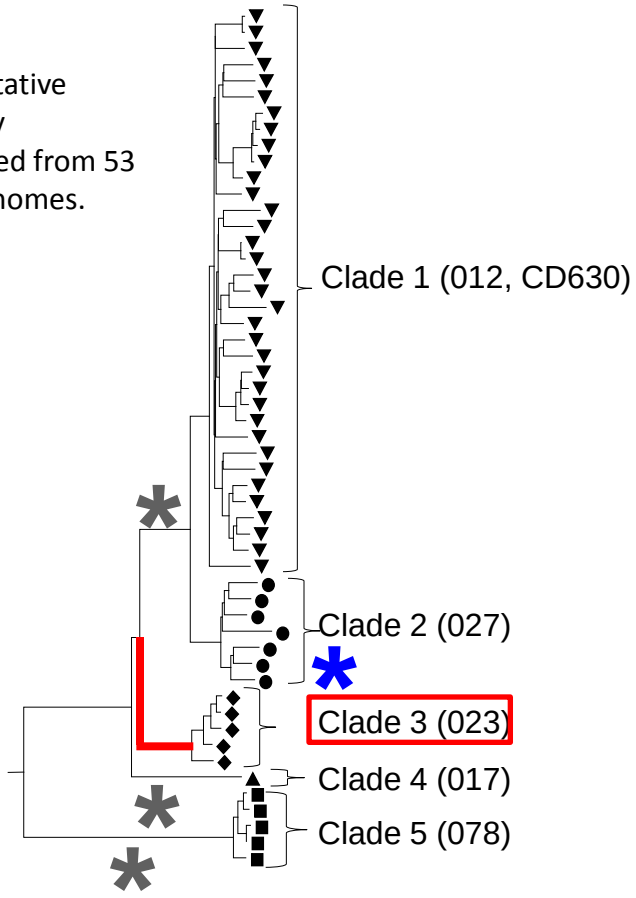


96% nucleotide sequence identity across the region in common.

Was this sequence once common to all PaLocs as part of an ancestral bacteriophage? Investigated phylogenetically and by sequence annotation.

Origin of the PaLoc insertion: phylogenetics

Representative phylogeny constructed from 53 whole genomes.



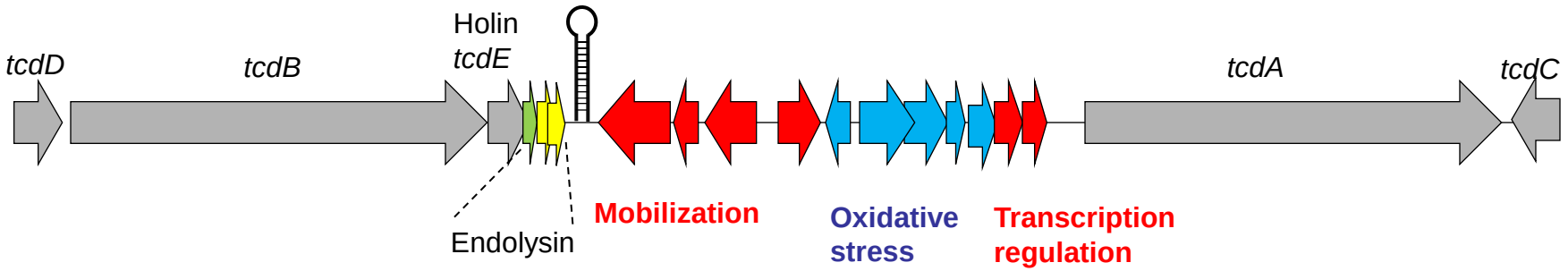
The most likely phylogenetic explanation is two related PaLoc insertions occurred on separate occasions, one insertion to the ancestor of clade 3 and one to 8864.

* Three independent identical deletions would be required to explain the sequences we observe today.

* Absence of the 1.1kb 8864 insertion from its close relatives in clade 2 suggests recent horizontal acquisition.

Origin of the PaLoc insertion: Annotation

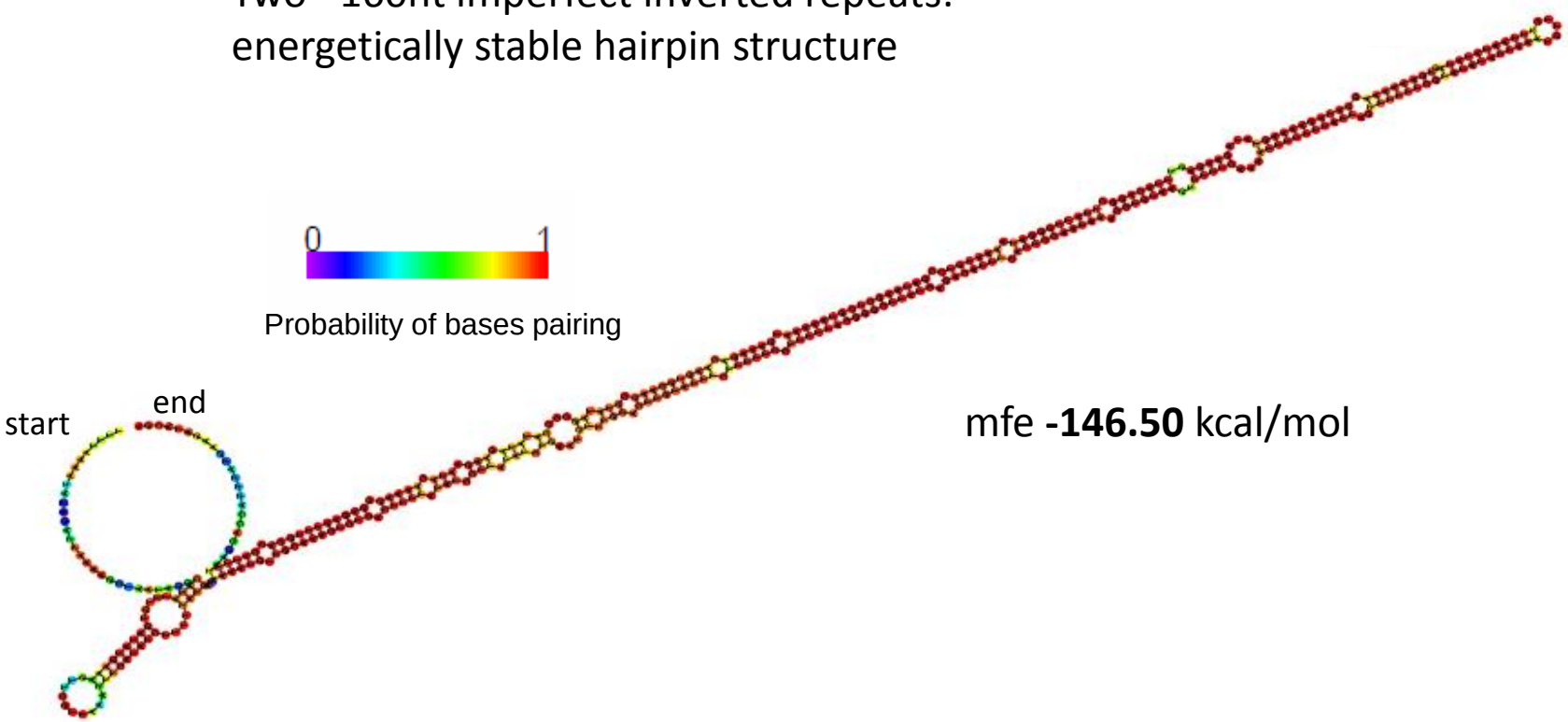
- Coding regions related to integrative bacteriophage and transposable elements.
- "Mobilization module"
    - **Tyrosine recombinase**: integrase, Tn916-like, lamboid bacteriophages.
    - **Xis**: directionality of integration versus excision.
    - **Rep-trans**: N-terminal: HTH cro/cI family. C-terminal possible topoisomerase which cleaves specifically at origins of transfer or replication.
    - **Imperfect palindrome**: 320 nt total length, energetically stable hairpin, possible bacteriophage *ori*.



- Endolysin: Concatenation of 3 adjacent amino acid sequence fragments gives 92% coverage, 75% aa identity, with N-acetyl-L-alanine amidase *C. difficile* ATCC43255 (also present in 8864).

Palindrome: 320 nt total length

Two ~160nt imperfect inverted repeats:
energetically stable hairpin structure

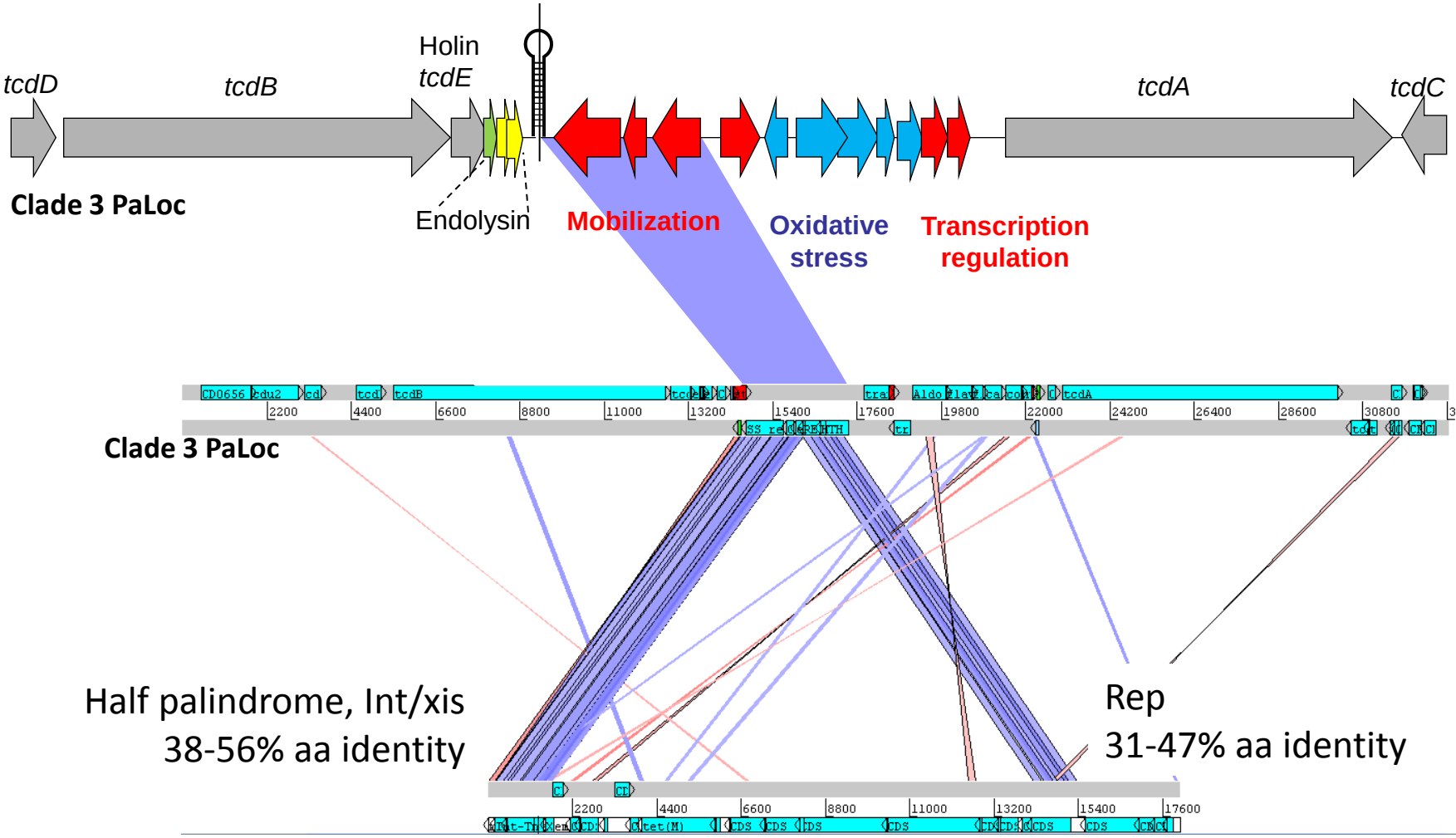


mfe **-146.50** kcal/mol

RNAfold web server: predict secondary structures of single stranded RNA or DNA sequences.

Mobilisation module: Sequence homology with Tn916:

integrase same family as bacteriophage lambda, xerC/D.



Tn916 *E. faecalis* (Flannagan et al., 1994)

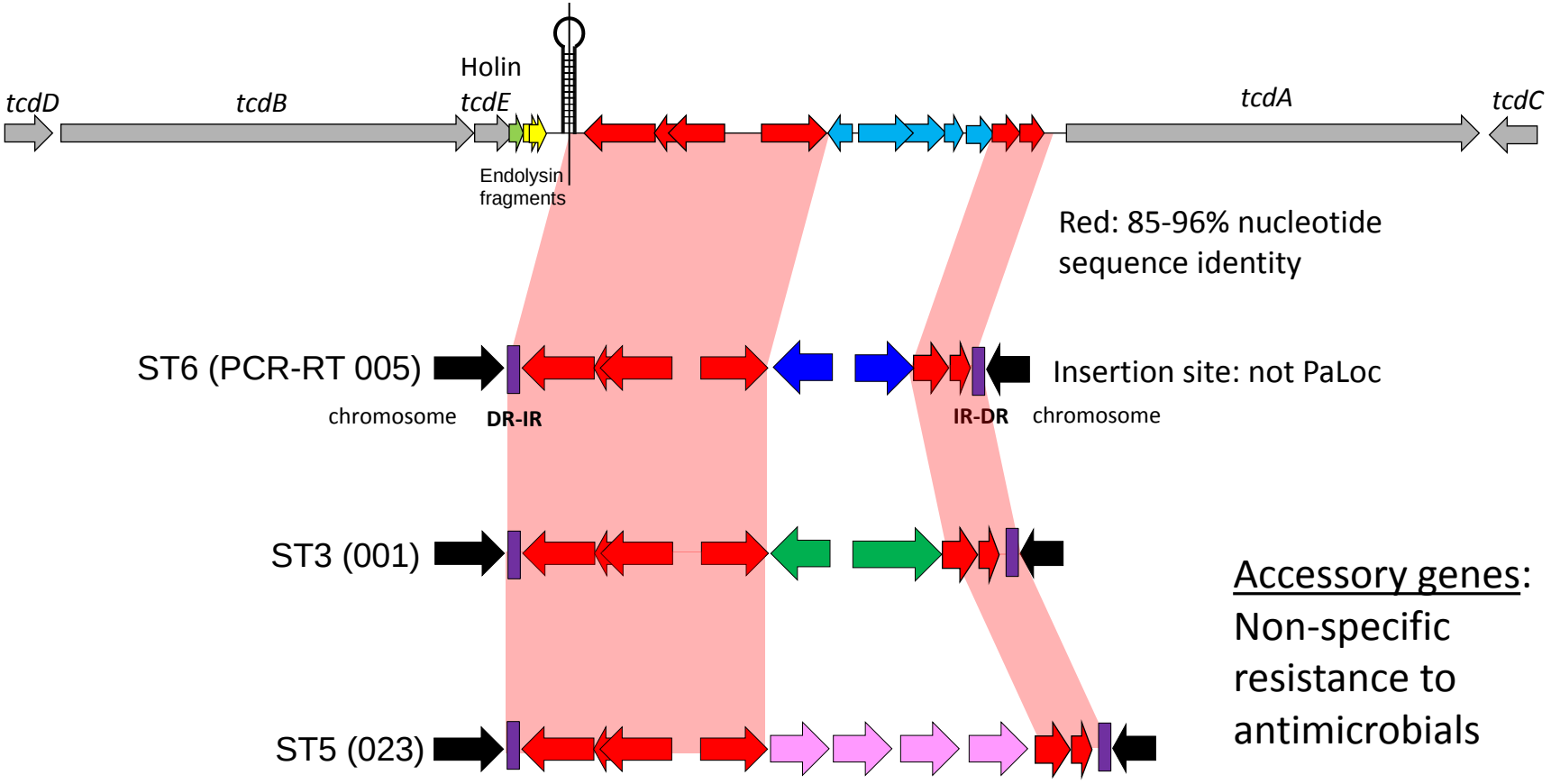
Are any sequences (bacteriophages? transposons?) closely related to the clade 3 PaLoc insertion present elsewhere in our isolate collection?

Excluding Tn916-like conjugative transposons.

Examined available WGS (n>2000) for sequences sharing at least 70% nucleotide sequence identity.

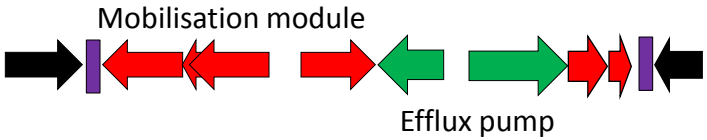
A family of mobile genetic elements

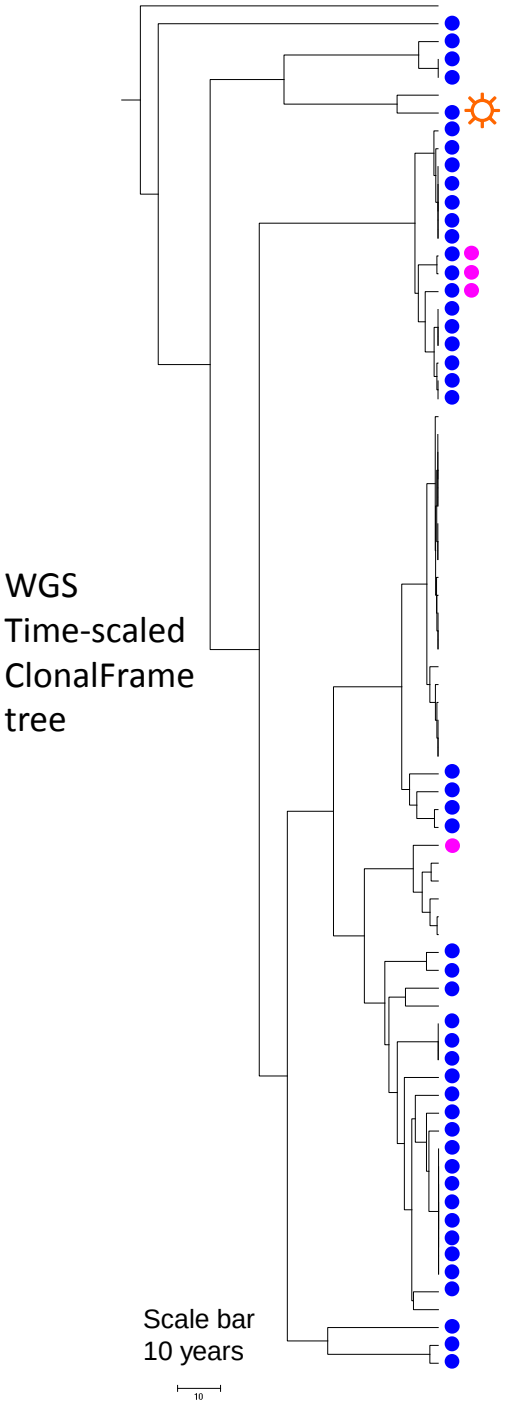
- Isolates of all five clades contained members of a family of mobile elements closely related to central ~8kb of PaLoc insertion.
- Occurred in a wide variety of PaLoc-independent chromosomal locations.
- Alignment of elements with PaLoc highlights accessory genes.



Accessory genes: Non-specific resistance to antimicrobials

Evidence of recent mobility within ST3 (PCR-RT 001)



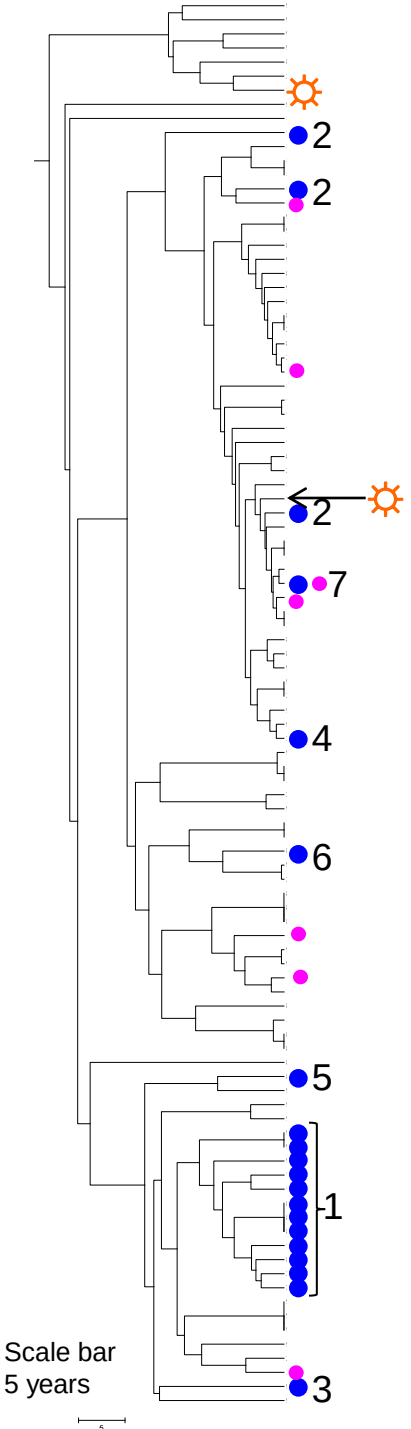


- Mobility demonstrated using a large number (76) genomes.
- One 1982 Australian isolate ☼
- Four Leeds isolates •

- Tn integrated in a <u>single chromosomal location</u>.

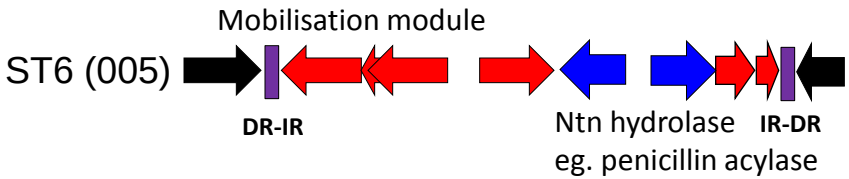
- Integration once and excision <u>6</u> times, 4 in last 10 years.

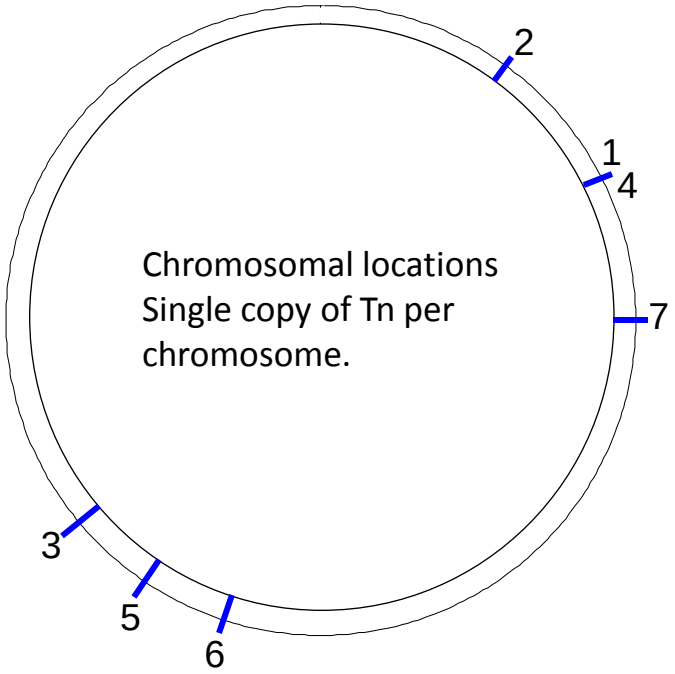
Recent mobility within ST6 (PCR-RT 005)



WGS tree
ST6 (005)
76 genomes
Oxford (67)
Leeds (7)
Australia (2)

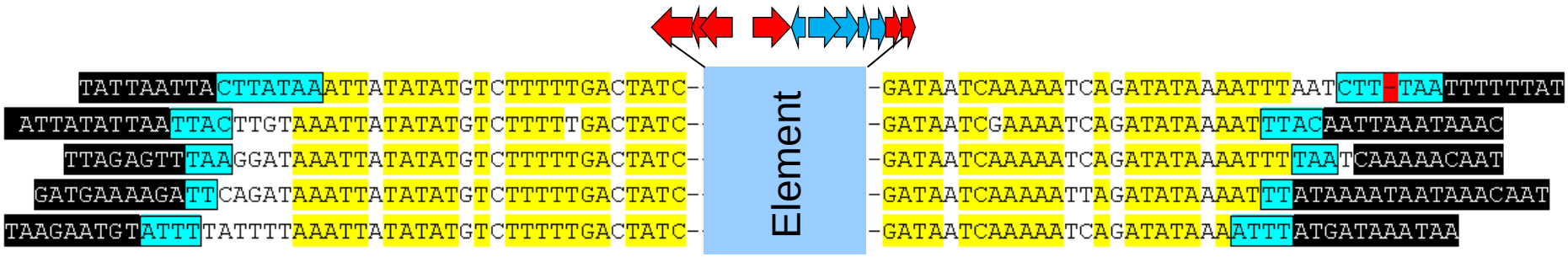
● Transposable element present.





Termini characteristic of mobile genetic elements

5’ and 3’ element/chromosome junctions: shown for <u>five</u> different elements

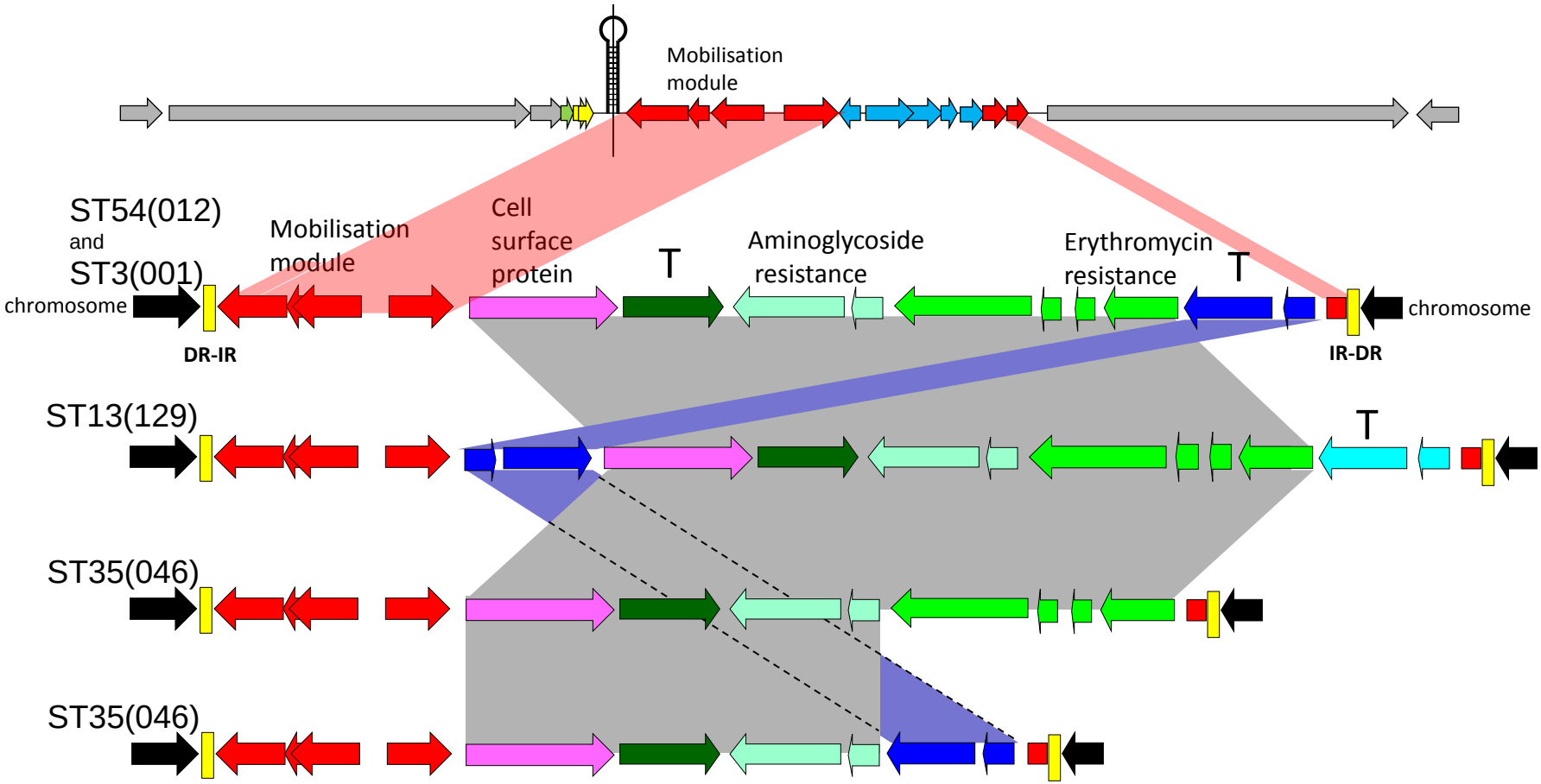


Imperfect inverted repeats: complementary bases highlighted.

Target site duplication: direct repeats

Chromosomal sequences flanking element, AT rich (in common with Tn916), most insertions between ORFs**.**

Acquisition of Additional Accessory Genes



Leucine rich repeats (LRR, protein-protein interactions) and Ig-like folds (bacterial cell surface proteins).

99-100% identity to *Enterococcus faecalis/faecium, Staph aureus, Strep pneumo*.

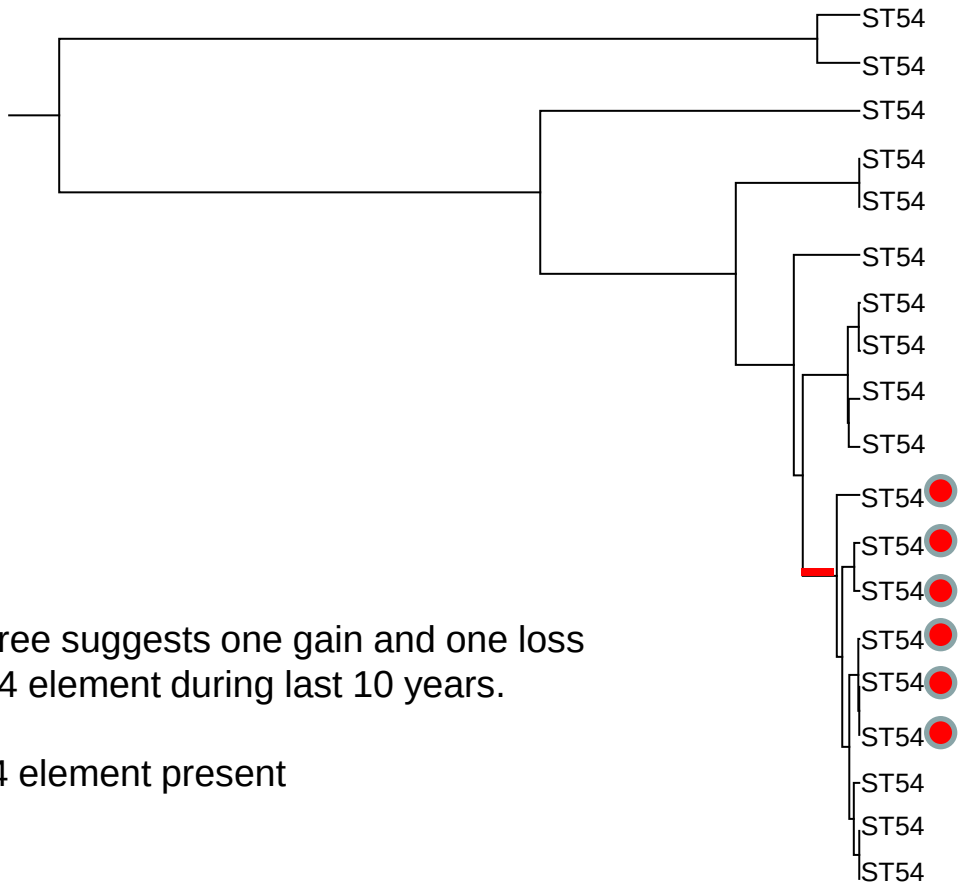
90-99% identity to species among normal human gut flora

Flanking chromosomal sequences

PaLoc insertion-like element

T: transposase

Evidence of mobility in ST54 (012)

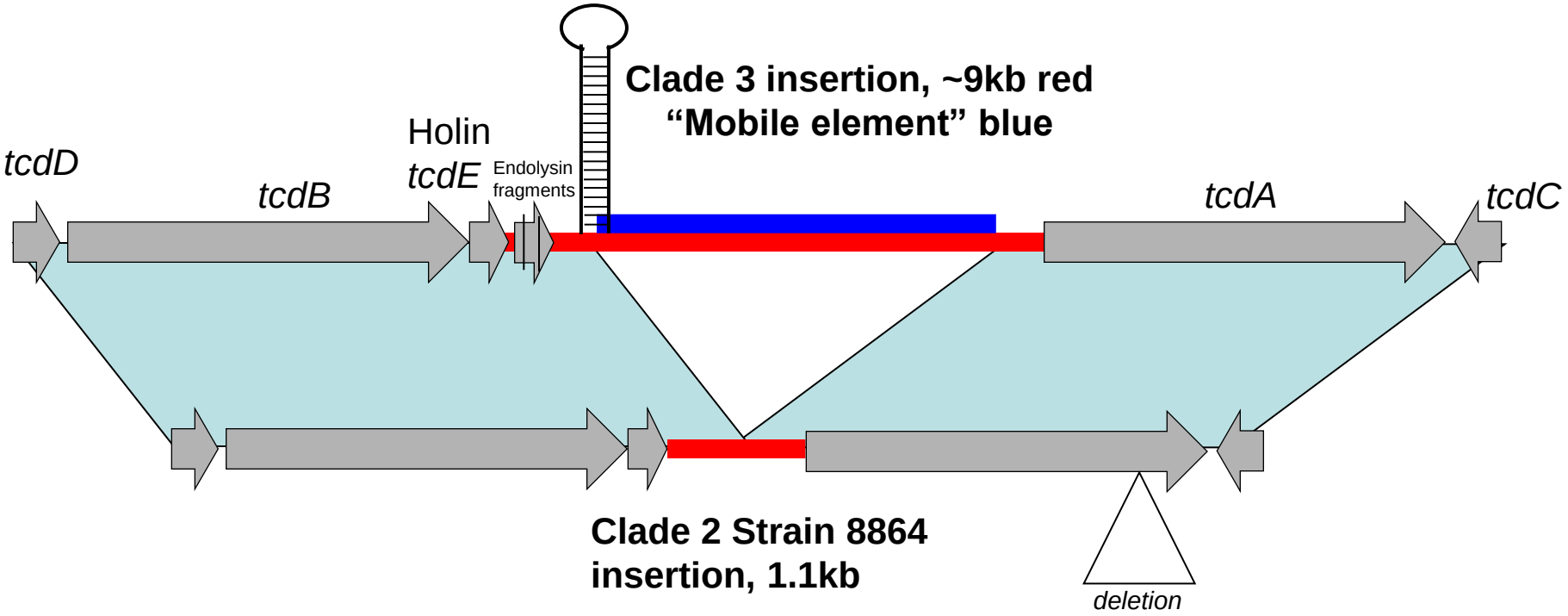


WGS tree suggests one gain and one loss of ST54 element during last 10 years.

ST54 element present

50
years

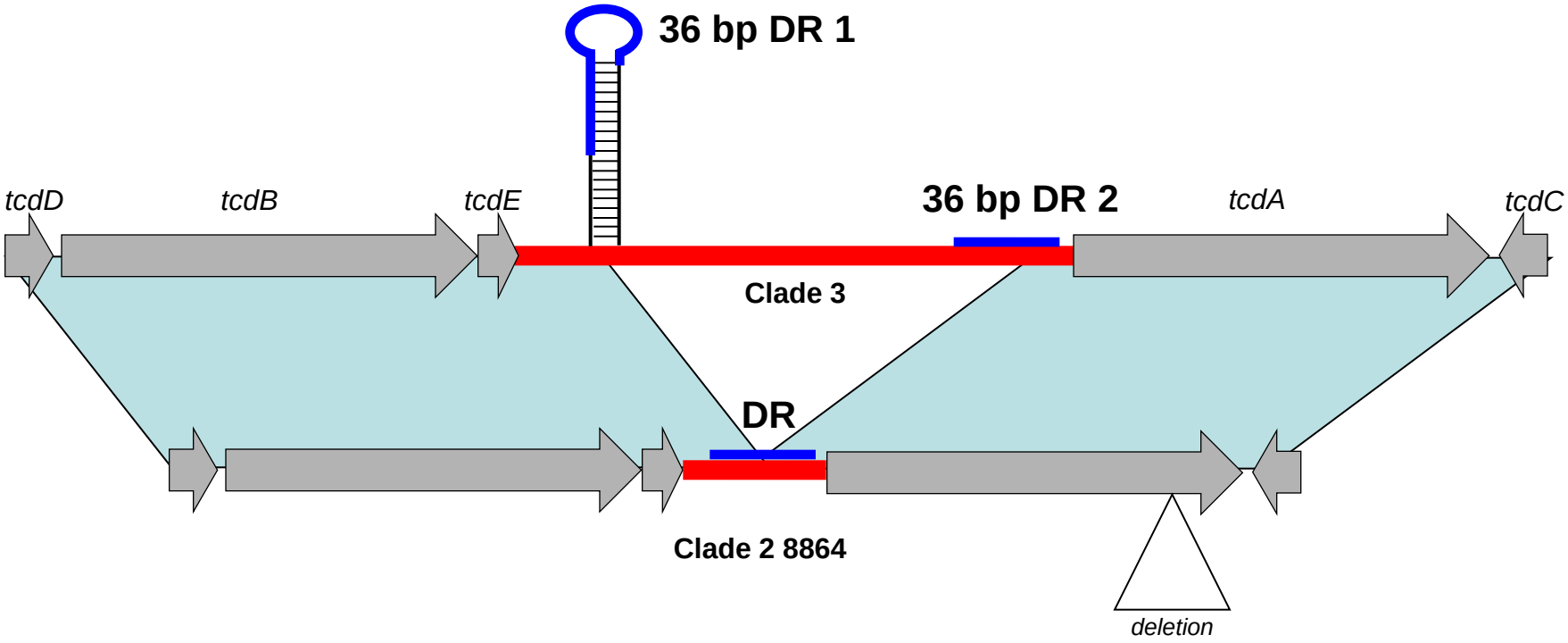
How are mobile elements generated?



Missing the "mobile element" sequence

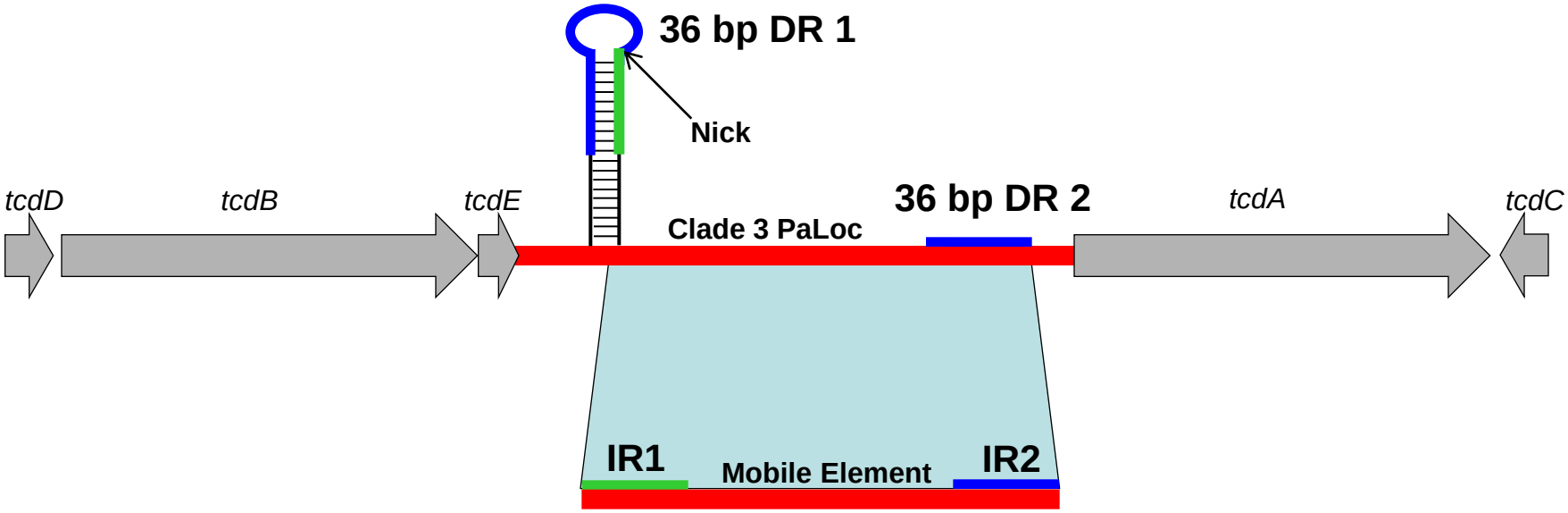
Mechanism may be palindrome mediated excision

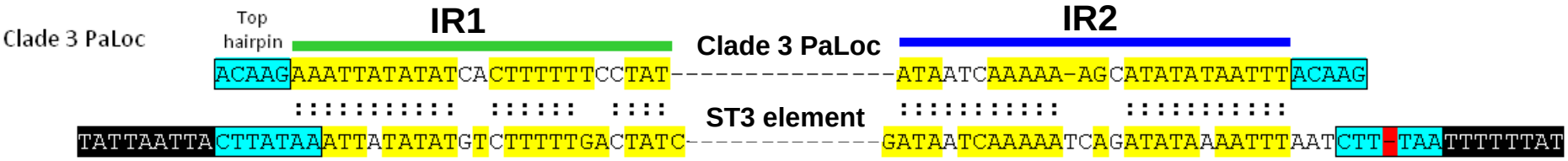
- Chromosomes containing large palindromes are unstable (*E. coli*).
- Eliminated by recombination between two direct repeats (DR), one in the palindrome and one nearby.
- Excises sequences between DR.
- A single DR remains. (Leach 1996).



Sequence missing from 8864 corresponds to mobile elements and it contains a single copy of the DR sequence.

Excision creates terminal inverted repeats of element?





Summary

- The clade 3 Pathogenicity Locus is 28kb, containing a bacteriophage related insertion of ~9kb.
- Phylogenetic data suggest this is unlikely to be part of an ancestral PaLoc bacteriophage, but instead a more recently acquired phage-like insertion.
- A fragment of endolysin-like sequence within the typical PaLoc suggest that 19kb PaLocs may have undergone a deletion.
- Palindrome mediated excision by recombination between two perfect 36bp DR would create a hypothetical mobile DNA element closely related to observed mobile elements.
- Is this a general mechanism whereby mobile elements can be generated?
- Do these mobile elements represent a missing-link between bacteriophages and conjugative transposons?

Acknowledgements

Oxford University/John Radcliffe Hospital, Oxford

Xavier Didelot
David Eyre
David Griffiths
Tanya Golubchik
Keith Jolley
Timothy Peto
Azim Ansari
Alison Vaughan
Esther Robinson
Rosalind Harding
Sarah Walker
Derrick Crook

University of Leeds/The General, Infirmary Leeds

Warren Fawley
Mark Wilcox

University of Western Australia/Pathwest Laboratory Medicine

Briony Elliott
Tom Riley

Oxford Biomedical Research Centre
Enabling translational research through partnership

NHS
National Institute for Health Research

UKCRC Modernising Medical Microbiology

wellcome trust

MRC Medical Research Council

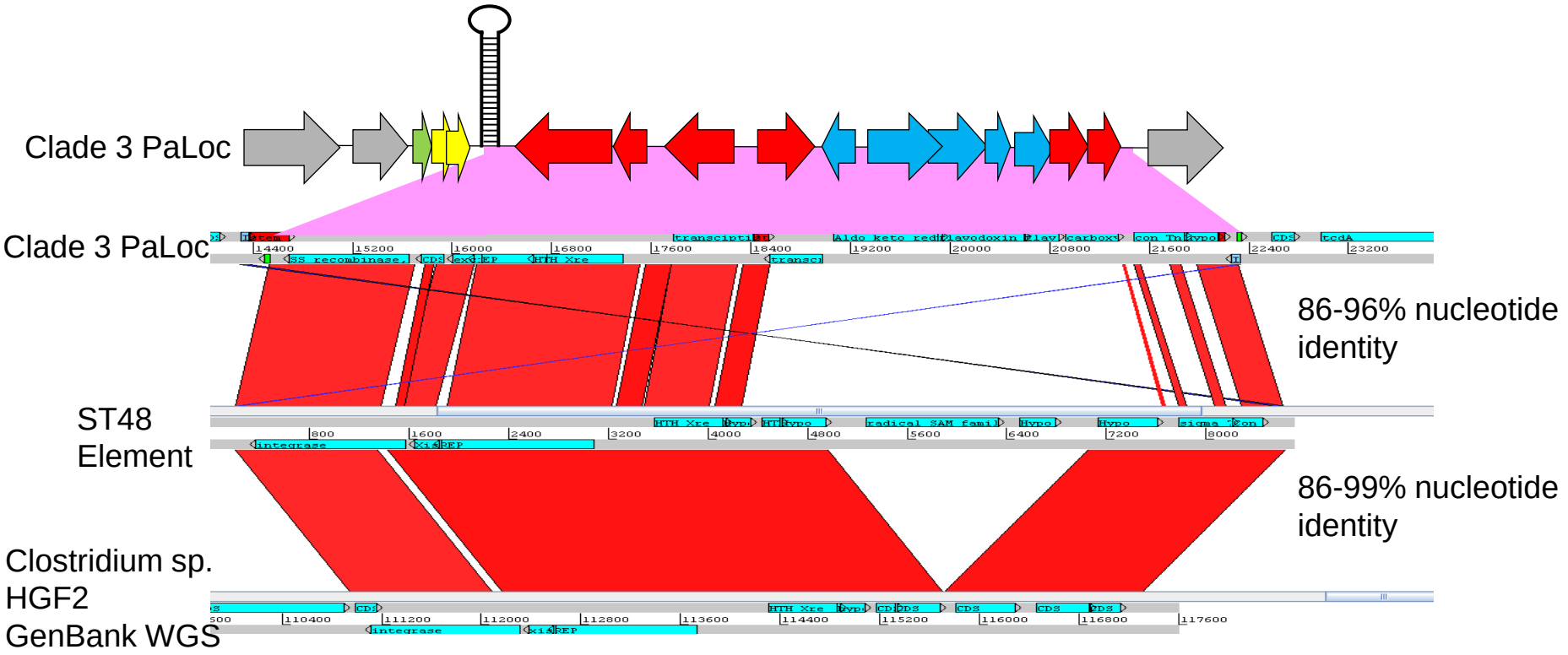
wellcome trust sanger institute

Health Protection Agency

UNIVERSITY OF OXFORD

Elements found in genomes of species of normal gut flora

GenBank WGS: *Clostridium sp.* HGF2, *Ruminococcus*, *Bifidobacterium*, *Lachnospiraceae, Coprobacillus.*



ST48 element has acquired chloramphenicol resistance relative to HGF2 *Clostridium sp*.