



INTRODUCTION

The National Collection of Type Cultures is a UKHSA-maintained microbial collection established in 1920 that today maintains and supplies approximately 6,000 distinct bacterial strains, many representing significant events in medical history. For example, NCTC 10442 *Staphylococcus aureus* was accessioned in 1966, and has since been recognised as the world's first methicillin-resistant *S. aureus* (MRSA) strain, and NCTC 13846 *Escherichia coli* was accessioned in 2016 following the first worldwide report of transferable colistin resistance. The collection is dynamic [1], continuing to respond to the changing global bacterial landscape, with 125 new strains made available from the NCTC in 2022 alone.

In collaboration with the Wellcome Sanger Institute and Pacific Biosciences, we carried out long-read sequencing of approximately 3,000 NCTC strains, in a project known as NCTC3000. Sequence reads and resulting genome assemblies and annotations were made freely available via ENA/GenBank/DDBJ under [BioProject PRJEB6403](#).

Many NCTC strains were accessioned prior to the widespread usage of antibiotics. Although metadata are available for many strains, anti-microbial resistance (AMR) data are limited. The NCTC3000 dataset provides us with a springboard to obtain deeper annotation of this key collection through a series of focussed projects.

METHODS

Exemplar project 1: an analysis of the plasmid content of 270 *Escherichia coli* strains; plasmids are mobile genetic elements and are implicated in AMR gene mobility in bacteria. Within the MSc. in Bioinformatics project of first author Jake Turnbull, the 270 genome sequences were processed using a bespoke analytical pipeline to determine strain plasmid content. The resulting dataset was then screened for AMR genes, specifically beta-lactamases.

Exemplar project 2: an investigation into an historic strain, NCTC 4136 *Staphylococcus aureus*, isolated in Chicago in 1932 during a case of food poisoning. Genomic analysis indicated this strain, which was isolated prior to the therapeutic use of penicillin, might nonetheless be resistant to this drug. The living strain was challenged with both penicillin and the later-developed antibiotic methicillin, to determine its resistance properties.

RESULTS

NCTC3000 statistics:

- 2,915 strains
- 876 species
- 810 Type strains
- 219 Genera
- 96 Families
- 8 Phyla

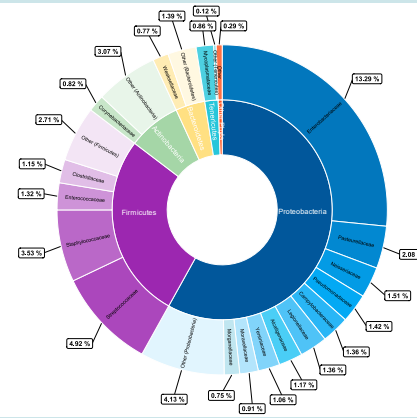


Figure 1: taxonomic distribution of the NCTC3000 dataset

201 out of the 270 *Escherichia coli* strains sequenced within the NCTC3000 dataset were identified as carrying plasmids [3]. Together, these 201 strains contain a collection of 374 putative plasmid contigs, as displayed in Figure 2 to the right.

The 374 putative plasmids were then screened for known beta-lactamase genes, with plasmids found to carry beta-lactamase genes indicated as dark blue bars in Figure 2.

The 270 strains were collected within a time period from before 1920 to 2015. We see that while plasmids have always been prevalent within *Escherichia coli*, plasmids harbouring identifiable beta-lactamase genes can only be seen after 1950, with a recent increase in frequency.

The NCTC3000 project [2] enabled the whole genome sequencing of approximately half of the 6,000 strains accessioned to the NCTC between 1920 and the present day (see Figure 1 to the left).

The project resulted in long-read PacBio sequences for 2,915 strains, which were then further analysed in order to derive genome assemblies and annotations.

The dataset is possibly the largest and highest quality dataset for a microbial collection in the world, and its immediate publication means it is being used by numerous researchers around the globe.

Most importantly, it enables us to add value to the NCTC strains by uncovering information about their functional properties. Their wide temporal spread makes them a highly valuable resource for research into antimicrobial resistance over the last century.

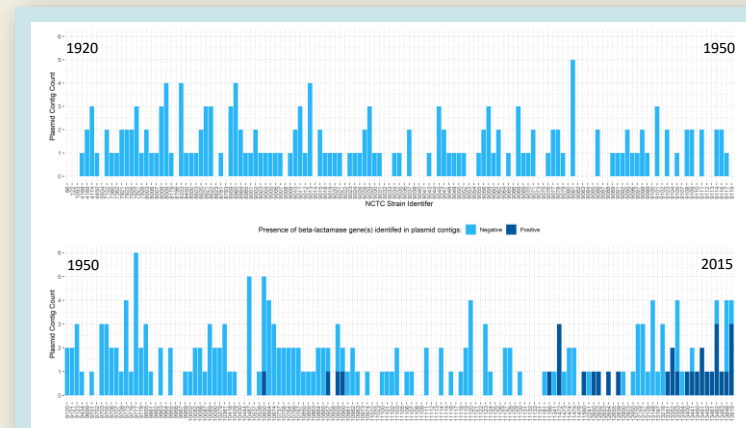


Figure 2: Plasmid-associated AMR genes over time

DISCUSSION

Project 1 found that the 270 *Escherichia coli* strains possessed between 0 and 6 putative plasmid contigs per strain and that the majority carried at least one plasmid. Further analysis of plasmid sequences indicated a growing frequency of AMR genes carried by plasmids over time.

Project 2 shows that genomic analysis can be used to pinpoint strains that may have features of interest. These strains can then be tested in the laboratory to confirm this attribute. This approach has confirmed that NCTC 4136, isolated in 1932, is resistant to penicillin. Is NCTC 4136 in fact the world's earliest living penicillin-resistant *Staphylococcus aureus* strain?

CONCLUSIONS

The NCTC3000 dataset enables us and others to delve deeper into irreplaceable fragments of bacterial history, including strains from both the pre- and post-antibiotic eras.

The dataset adds value to the NCTC collection, helping us to secure the future of a valuable UKHSA and global asset, and provides key genetic and genomic information to researchers in their current and future endeavours.

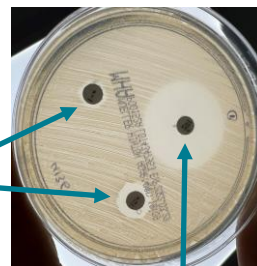
ACKNOWLEDGEMENTS

The generation of the NCTC3000 dataset was funded by Wellcome Trust Grant no: 101503/Z/13/Z 'Creation of an e-resource centre to underpin the provision and use of Type and reference strains of human pathogens'. Pacific Biosciences gave considerable support to the project in the form of advice, equipment, consumables and staff time.

REFERENCES

1. Turnbull JD *et al.* (2022) Notification List of Bacterial Strains Made Available by the United Kingdom National Collection of Type Cultures in 2021. *Microbiol. Resour. Announc.* 11(8):e0035722.
2. Dicks J *et al.* (2023) NCTC3000: a century of bacterial strain collecting leads to a rich genomic data resource. *Microbial Genomics* 9(5):000976.
3. Turnbull JD (2021) Investigating the detection, prevalence, and composition of *Enterobacteriaceae* plasmids using whole genome sequence data derived from selected NCTC strains. MSc. in Bioinformatics thesis, Queen Mary University of London.
4. Parker MT and Lapage SP (1957) Penicillinase production by *Staphylococcus aureus* strains from outbreaks of food poisoning. *J. clin. Path.* 10:313-317.

Bacteria are not able to grow around antibiotics to which they are sensitive.



Small 'rings' indicate largely resistant to penicillin (the two discs have different concentrations of the drug)

Large 'ring' indicates susceptible to methicillin

Figure 3: Penicillin resistance in NCTC 4136

NCTC 4136 is a *Staphylococcus aureus* strain that was isolated from an outbreak of food poisoning in 1932 in Chicago, USA.

Genomic analysis identified a beta-lactamase gene carried by NCTC 4136, potentially underpinning penicillin resistance in this pre-antibiotic era strain.

Antimicrobial susceptibility testing (AST) confirmed that NCTC 4136 is **resistant to penicillin** but is still susceptible to methicillin (see Figure 3 to the left).

Literature analysis [4] shows that NCTC 4136 was once known to be resistant to penicillin, but this lost information has now been recovered by a combination of genomic and laboratory analysis.