# From Escherich to the *Escherichia coli* genome: how a commensal bacterium shaped the history of modern microbiology

Guillaume Méric[1], Matthew D. Hitchings[2], Ben Pascoe[1,3], Samuel K. Sheppard[1,3,4]*

[1]The Milner Centre for Evolution, Department of Biology and Biochemisty, University of Bath, Bath, BA2 7AY, UK; [2]Swansea University Medical School, Institute of Life Science, Singleton Park, Swansea, SA2 8PP, UK; [3]MRC Cloud Infrastructure for Microbial Bioinformatics (CLIMB) Consortium, Institute of Life Science, UK; [4]Department of Zoology, University of Oxford, Oxford, UK. *Corresponding author: Professor Samuel K. Sheppard: s.k.sheppard@bath.ac.uk.

In 1885, the pioneering Bavarian paediatrician Theodor Escherich was battling against neonatal dysentery when he first isolated "*Bacterium coli commune*" from the stool of infants in the laboratory of Otto von Bollinger in Munich [1]. Later known as *E. coli,* this organism has shaped bacterial genetics and its 130 year history is synonymous with the rise of modern microbiology. *E. coli* exists principally as a harmless component of natural gut microflora of animals including humans and, while it can cause disease, it might seem surprizing that this single lineage within the enormous complexity of the prokaryote tree of life has become so influential.

From early discoveries, understanding of some of the world's most common diseases has been driven by the exchange of bacterial isolates between experimental laboratories. In 1920, the archives of the Lister Institute became the National Culture Type Collection (NCTC). This collection, which contains isolates that date back to the very infancy of infectious disease research, underpins a wealth of internationally recognised standard microbiology procedures and has driven understanding of the enormous genotypic and phenotypic diversity within and between bacterial species. In addition to reference strains, the NCTC collection contains a wealth of information including isolate records, lab books and personal communications from some of Europe's most influential microbiologists. But until now, the identity of Escherich's original *E. coli* strain, the most famous bacterium of all, has remained obscure.

In 2015, whilst searching through the NCTC and Lister Institute document archives, we were able to piece together the journey of the first *E. coli* isolate through the laboratories of some of history's most famous microbiologists (**Figure 1**). Around 1900, Escherich's brother-in-law, Meinhard von Pfaundler, sent the *"Bacterium coli commune*" strain from "Professor Escherich's laboratory" to

Herbert Durham at the Pathological Laboratory of the University of Cambridge [2]. Durham gave the original Escherich strain to Albert MacConkey and Harriette Chick, then at the Royal Commission on Sewage Disposal in Liverpool, who acknowledged the gift in their 1900 and 1901 studies published in the Thomson-Yates Laboratories Reports [3-6]. MacConkey then moved to the Lister Institute in November 1901 [7] and later was joined by Harriet Chick, who after graduating from London University, went on to become the first woman to be appointed a research fellow at the Lister Institute, much to the displeasure of some of her male colleagues [8]. Strain collections that accompanied MacConkey and Chick from Liverpool formed the basis of MacConkey's landmark 1905 paper describing *B. coli* and lactose-fermenting bacteria [5]. The *B. coli* strain was subsequently used by Arthur Boycott, then assistant pathologist at the Lister Institute, and referred to as 'the original Escherich strain' in his work on typhoid fever published in 1906 [9]. This 'stock strain' was archived as *B. coli communis* at the Lister Institute in 1911 by Francis Bainbridge [10], and the institute's collections became the NCTC in 1920 [11], under the curation of Ralph St John-Brooks and Mabel Rhodes. The 'Original Escherich strain' was assigned the name of NCTC86 (**Figure 2**) and *Bacterium coli commune* was officially reclassified as *Escherichia coli* in 1919 [12] in honour of Theodor Escherich, who had died in 1911 in Vienna [13].

In the last decade, advances DNA sequencing technologies and bioinformatics analyses have allowed the efficient sequencing of large numbers of bacterial genomes. By sequencing the Original Escherich strain and comparative analysis of 4,856 predicted coding sequences, we were able to get a detailed picture of this important strain and how it relates to other sequenced *E. coli.* Based upon a core and accessory genome analysis it was possible to reconstruct a 2,600 core gene phylogeny relative to 70 publically available genomes (See supplementary material on http://www.sheppardlab.com/nctc86), and characterize accessory genes with known functions such as host colonization and virulence that would potentially imply a role in dysentery, as originally postulated by Escherich.

The original Escherich strain belonged to the sequence type 10 cluster within the common phylogroup A, and was strikingly similar to isolate B41 from a pig sampled in 1980, with 64.8% of alleles being identical. Surprisingly, the genome contained no known pathogenicity islands and well known virulence and colonisation factors, including Shiga toxin and attachment, adhesion and invasion genes (*hek, sfaEFA, vpeRABC*), were absent. Furthermore, reported resistance to penicillin and erythromycin was not reflected in the presence of any known resistance allele, and resistance to synthetic antibiotics, such as sulphonamides or quinolones, was not detected.

Based on these findings it is likely that, in 1885, Escherich isolated a non-pathogenic strain that was part of the natural gut micrbiota of the new-born subject, rather that the cause of neonatal dysentery. The question remains, how did a harmless commensal bacterium go on to be central for understanding fundamental biological processes such as bacterial conjugation [14], phage genetics [15], the topography of gene structure [16], and genomics [17]? In fact, the reason for its early discovery and its adoption as a model organism are probably the same. *E. coli* is ubiquitous in multiple environments and acquiescent to laboratory culture and this hat has led to it becoming arguably the most well understood organism on Earth.


1. Escherich T (1885) Die Darmbakterien des Neugeborenen und Säuglings. Fortschr Med 3: 515-522; 547-554.
2. Durham HE (1901) Some theoretical considerations upon the nature of agglutinins, together with further observations upon *Bacillus typhi abdominalis*, *Bacillus enteritidis*, *Bacillus coli communis*, *Bacillus lactis aerogenes*, and some other bacilli of allied character. J Exp Med 5: 353-388.
3. MacConkey A (1900) The differentiation and isolation from mixtures of the *Bacillus coli communis* and *Bacillus typhosus* by the use of sugars and the salts of bile. In: Boyce R, Sherrington CS, editors. The Thompson Yates Laboratories Report: The University Press of Liverpool.
4. Chick H (1901) The distribution of *B. coli commune,* Part II. In: Boyce R, Sherrington CS, editors. The Thompson Yates Laboratories Report: The University Press of Liverpool.
5. Macconkey A (1905) Lactose-Fermenting Bacteria in Faeces. J Hyg (Lond) 5: 333-379.
6. MacConkey AT (1908) Bile Salt Media and their advantages in some Bacteriological Examinations. J Hyg (Lond) 8: 322-334.
7. G.F.P. (1931) Obituary. Dr. A.T. MacConkey. Nature 127: 980-981.
8. Chick J (2011) War on Disease: a History of the Lister Institute. BMJ 343: d8209.
9. Boycott AE (1906) Observations on the Bacteriology of Paratyphoid Fever and on the Reactions of Typhoid and Paratyphoid Sera. J Hyg (Lond) 6: 33-73.
10. Bainbridge FA (1911) The Action of Certain Bacteria on Proteins. J Hyg (Lond) 11: 341-355.
11. Russell JE (2014) Historical Perspectives. The National Collection of Type Cultures. Microbiologist: Society for Applied Microbiology.
12. Castellani A, Chalmers AJ (1919) Manual of Tropical Medicine, 3rd edition: William Wood and Company, New York.
13. Friedmann HC (2006) Escherich and *Escherichia*. Adv Appl Microbiol 60: 133-196.
14. Lederberg J, Tatum EL (1946) Gene recombination in Escherichia coli. Nature 158: 558.
15. Susman M (1995) The Cold Spring Harbor Phage Course (1945-1970): a 50th anniversary remembrance. Genetics 139: 1101-1106.
16. Benzer S (1961) On the Topography of the Genetic Fine Structure. Proc Natl Acad Sci U S A 47: 403-415.
17. Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, et al. (1997) The complete genome sequence of Escherichia coli K-12. Science 277: 1453-1462.
18. Supplementary material for the genomic characterisation of *E. coli* NCTC86 can be found on http://www.sheppardlab.com/nctc86