



*Citation for published version:*

Méric, G, Hitchings, MD, Pascoe, B & Sheppard, SK 2016, 'From Escherich to the *Escherichia coli* genome', *The Lancet Infectious Diseases*, vol. 16, no. 6, pp. 634-636. [https://doi.org/10.1016/S1473-3099\(16\)30066-4](https://doi.org/10.1016/S1473-3099(16)30066-4)

*DOI:*

[10.1016/S1473-3099\(16\)30066-4](https://doi.org/10.1016/S1473-3099(16)30066-4)

*Publication date:*

2016

*Document Version*

Peer reviewed version

[Link to publication](#)

**University of Bath**

**Alternative formats**

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# 1 From Escherich to the *Escherichia coli* genome: how a commensal bacterium shaped the 2 history of modern microbiology

3  
4 Guillaume Méric<sup>1</sup>, Matthew D. Hitchings<sup>2</sup>, Ben Pascoe<sup>1,3</sup>, Samuel K. Sheppard<sup>1,3,4\*</sup>

5  
6 <sup>1</sup>The Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath,  
7 Bath, BA2 7AY, UK; <sup>2</sup>Swansea University Medical School, Institute of Life Science, Singleton  
8 Park, Swansea, SA2 8PP, UK; <sup>3</sup>MRC Cloud Infrastructure for Microbial Bioinformatics (CLIMB)  
9 Consortium, Institute of Life Science, UK; <sup>4</sup>Department of Zoology, University of Oxford, Oxford,  
10 UK. \*Corresponding author: Professor Samuel K. Sheppard: [s.k.sheppard@bath.ac.uk](mailto:s.k.sheppard@bath.ac.uk).

11  
12 In 1885, the pioneering Bavarian paediatrician Theodor Escherich was battling against neonatal  
13 dysentery when he first isolated “*Bacterium coli commune*” from the stool of infants in the  
14 laboratory of Otto von Bollinger in Munich [1]. Later known as *E. coli*, this organism has shaped  
15 bacterial genetics and its 130 year history is synonymous with the rise of modern microbiology. *E.*  
16 *coli* exists principally as a harmless component of natural gut microflora of animals including  
17 humans and, while it can cause disease, it might seem surprizing that this single lineage within the  
18 enormous complexity of the prokaryote tree of life has become so influential.

19  
20 From early discoveries, understanding of some of the world’s most common diseases has been  
21 driven by the exchange of bacterial isolates between experimental laboratories. In 1920, the archives  
22 of the Lister Institute became the National Culture Type Collection (NCTC). This collection, which  
23 contains isolates that date back to the very infancy of infectious disease research, underpins a wealth  
24 of internationally recognised standard microbiology procedures and has driven understanding of the  
25 enormous genotypic and phenotypic diversity within and between bacterial species. In addition to  
26 reference strains, the NCTC collection contains a wealth of information including isolate records, lab  
27 books and personal communications from some of Europe’s most influential microbiologists. But  
28 until now, the identity of Escherich’s original *E. coli* strain, the most famous bacterium of all, has  
29 remained obscure.

30  
31 In 2015, whilst searching through the NCTC and Lister Institute document archives, we were able to  
32 piece together the journey of the first *E. coli* isolate through the laboratories of some of history’s  
33 most famous microbiologists (**Figure 1**). Around 1900, Escherich’s brother-in-law, Meinhard von  
34 Pfaundler, sent the “*Bacterium coli commune*” strain from “Professor Escherich’s laboratory” to

35 Herbert Durham at the Pathological Laboratory of the University of Cambridge [2]. Durham gave the  
36 original Escherich strain to Albert MacConkey and Harriette Chick, then at the Royal Commission  
37 on Sewage Disposal in Liverpool, who acknowledged the gift in their 1900 and 1901 studies  
38 published in the Thomson-Yates Laboratories Reports [3-6]. MacConkey then moved to the Lister  
39 Institute in November 1901 [7] and later was joined by Harriet Chick, who after graduating from  
40 London University, went on to become the first woman to be appointed a research fellow at the  
41 Lister Institute, much to the displeasure of some of her male colleagues [8]. Strain collections that  
42 accompanied MacConkey and Chick from Liverpool formed the basis of MacConkey's landmark  
43 1905 paper describing *B. coli* and lactose-fermenting bacteria [5]. The *B. coli* strain was  
44 subsequently used by Arthur Boycott, then assistant pathologist at the Lister Institute, and referred to  
45 as 'the original Escherich strain' in his work on typhoid fever published in 1906 [9]. This 'stock  
46 strain' was archived as *B. coli communis* at the Lister Institute in 1911 by Francis Bainbridge [10],  
47 and the institute's collections became the NCTC in 1920 [11], under the curation of Ralph St John-  
48 Brooks and Mabel Rhodes. The 'Original Escherich strain' was assigned the name of NCTC86  
49 (**Figure 2**) and *Bacterium coli commune* was officially reclassified as *Escherichia coli* in 1919 [12]  
50 in honour of Theodor Escherich, who had died in 1911 in Vienna [13].

51

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

61

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

70 Based on these findings it is likely that, in 1885, Escherich isolated a non-pathogenic strain that was  
71 part of the natural gut microbiota of the new-born subject, rather than the cause of neonatal dysentery.  
72 The question remains, how did a harmless commensal bacterium go on to be central for  
73 understanding fundamental biological processes such as bacterial conjugation [14], phage genetics  
74 [15], the topography of gene structure [16], and genomics [17]? In fact, the reason for its early  
75 discovery and its adoption as a model organism are probably the same. *E. coli* is ubiquitous in  
76 multiple environments and acquiescent to laboratory culture and this has led to it becoming  
77 arguably the most well understood organism on Earth.

78

- 79 1. Escherich T (1885) Die Darmbakterien des Neugeborenen und Säuglings. Fortschr Med 3: 515-  
80 522; 547-554.
- 81 2. Durham HE (1901) Some theoretical considerations upon the nature of agglutinins, together with  
82 further observations upon *Bacillus typhi abdominalis*, *Bacillus enteritidis*, *Bacillus coli*  
83 *communis*, *Bacillus lactis aerogenes*, and some other bacilli of allied character. J Exp Med 5:  
84 353-388.
- 85 3. MacConkey A (1900) The differentiation and isolation from mixtures of the *Bacillus coli*  
86 *communis* and *Bacillus typhosus* by the use of sugars and the salts of bile. In: Boyce R,  
87 Sherrington CS, editors. The Thompson Yates Laboratories Report: The University Press of  
88 Liverpool.
- 89 4. Chick H (1901) The distribution of *B. coli commune*, Part II. In: Boyce R, Sherrington CS, editors.  
90 The Thompson Yates Laboratories Report: The University Press of Liverpool.
- 91 5. Macconkey A (1905) Lactose-Fermenting Bacteria in Faeces. J Hyg (Lond) 5: 333-379.
- 92 6. MacConkey AT (1908) Bile Salt Media and their advantages in some Bacteriological  
93 Examinations. J Hyg (Lond) 8: 322-334.
- 94 7. G.F.P. (1931) Obituary. Dr. A.T. MacConkey. Nature 127: 980-981.
- 95 8. Chick J (2011) War on Disease: a History of the Lister Institute. BMJ 343: d8209.
- 96 9. Boycott AE (1906) Observations on the Bacteriology of Paratyphoid Fever and on the Reactions  
97 of Typhoid and Paratyphoid Sera. J Hyg (Lond) 6: 33-73.
- 98 10. Bainbridge FA (1911) The Action of Certain Bacteria on Proteins. J Hyg (Lond) 11: 341-355.
- 99 11. Russell JE (2014) Historical Perspectives. The National Collection of Type Cultures.  
100 Microbiologist: Society for Applied Microbiology.
- 101 12. Castellani A, Chalmers AJ (1919) Manual of Tropical Medicine, 3rd edition: William Wood and  
102 Company, New York.
- 103 13. Friedmann HC (2006) Escherich and *Escherichia*. Adv Appl Microbiol 60: 133-196.
- 104 14. Lederberg J, Tatum EL (1946) Gene recombination in *Escherichia coli*. Nature 158: 558.
- 105 15. Susman M (1995) The Cold Spring Harbor Phage Course (1945-1970): a 50th anniversary  
106 remembrance. Genetics 139: 1101-1106.
- 107 16. Benzer S (1961) On the Topography of the Genetic Fine Structure. Proc Natl Acad Sci U S A 47:  
108 403-415.
- 109 17. Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, et al. (1997) The complete  
110 genome sequence of *Escherichia coli* K-12. Science 277: 1453-1462.
- 111
- 112 18. Supplementary material for the genomic characterisation of *E. coli* NCTC86 can be found on  
113 <http://www.sheppardlab.com/nctc86>