Edited by DFB on 29/6/22

Laboratory strains of *Escherichia coli* K-12: not such perfect role models after all.

Douglas F. Browning [1]*, Jon L. Hobman [2], Stephen J.W. Busby[3].

[1] School of Biosciences, College of Health and Life Sciences, Aston University, Aston Triangle, Birmingham, B4 7ET, UK.

[2] School of Biosciences, University of Nottingham, Sutton Bonington Campus, Sutton Bonington, Loughborough, LE12 5RD, UK.

[3] Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Birmingham, B15 2TT, UK.

***For correspondence: d.browning@aston.ac.uk**

Key words: *Escherichia coli* K-12, genomic analysis, laboratory-based evolution, F plasmid, bacteriophage lambda.

Abbreviations: ACT, Artemis Comparison Tool; BLAST, Basic Local Alignment Search Tool; CDS, coding sequence; CGE, Center for Genomic Epidemiology; CGSC, Coli Genetic Stock Centre; ONT, Oxford Nanopore Technologies; SNV, single nucleotide variant.

Abstract

Escherichia coli K-12 was originally isolated 100 years ago and since then, it has become an invaluable model organism and a cornerstone of molecular biology research. However, despite its apparent pedigree, since its initial isolation, *E. coli* K-12 has been repeatedly cultured, passaged, and mutagenized, resulting in an organism that carries extensive genetic changes. To understand more about the evolution of this important model organism, we have sequenced the genomes of two ancestral K-12 strains, WG1 and EMG2, considered to be the progenitors of many key laboratory strains. Our analysis confirms that these strains still carry genetic elements such as bacteriophage lambda (λ) and the F plasmid, but also indicates that they have undergone extensive lab-based evolution. Thus, scrutinizing the genomes of ancestral *E. coli* K-12 strains, leads us to question whether *E. coli* K-12 is a sufficiently robust model organism for 21st century microbiology.

DATA SUMMARY

All supporting data are provided within the article or through supplementary data files. Supplementary Figs. S1 to S14 and Supplementary File S1 are available with the online version of this article. All genome sequence data has been deposited in NCBI GenBank under Bioproject ID PRJNA848777. The assembled and annotated genomes of WG1 and EMG2 have been deposited with the accession numbers, CP099590 and CP099591 (WG1) and CP099588 and CP099589 (EMG2).

Impact Statement

Since its isolation in 1922, *Escherichia coli* K-12, has become arguably the premier model organism for contemporary science. The adoption of *E. coli* K-12 by many microbiologists across the globe, means that it has a complex pedigree, and, although many *E. coli* K-12 strains have been sequenced, little is known about the early versions of K-12, which still carry the F plasmid and bacteriophage λ. To understand more about the lab-based evolution that has shaped this important model organism, we have sequenced two ancestral K-12 strains, WG1 and EMG2, that are considered to be the progenitors of many of the laboratory strains used today.

INTRODUCTION

Escherichia coli K-12 was originally isolated in 1922 from a convalescent diphtheria patient and, later in the 1940s, adopted by Charles Clifton and Edward Tatum as a model organism (1-3). Since then, *E. coli* K-12 has become the "workhorse" of molecular biology, becoming arguably the premier model organism in science today. MG1655 was the first *E. coli* K-12 strain to have its genome sequence published, followed by W3110, resulting in an explosion of genomic research and comparative genomics (4, 5). However, despite its prestige, *E. coli* K-12 was stored on agar plates, stabs or slopes before cryopreservation became established, and has been repeatedly subcultured and mutagenized (Fig. 1), resulting in an organism which carries extensive genetic changes and has lost the ability to produce many surface-associated structures (3). For example, *E. coli* K-12 lab strains are unable to synthesize O antigen on their lipopolysaccharide and no longer carry the F plasmid or bacteriophage λ (3, 6-9). One major strength of using *E. coli* K-12 strains for cloning and heterologous gene expression is that K-12 strains cannot establish in the human gut (10, 11), and, thus, even so-called "wild type" *E. coli* K-12 strains, like MG1655 and W3110, are very different from commensal or environmental isolates (3, 4, 12, 13). To understand more about the evolution of this important model organism, we have sequenced the genomes of two *E. coli* K-12 strains, WG1 and EMG2, the proposed ancestors of key laboratory strains (Fig. 1) (1, 2). Our analysis confirms that these strains carry genetic elements such as phage λ and the F plasmid, but indicates that they have also undergone extensive mutational alternation during their evolution in laboratories.

METHODS

Bacterial strains and whole genome sequencing

E. coli K-12 strains WG1 and EMG2 were obtained for the Coli Genetic Stock Centre (CGSC), strain numbers CGSC#5073 and CGSC#4401, respectively (1, 2). Each strain was sequenced using the enhanced sequencing option from MicrobesNG (https://microbesng.com/), which uses a combination of Illumina and Oxford Nanopore Technologies (ONT). Cell cultures were grown in LB medium and the cell pellet was isolated by centrifugation and resuspended in the cryo-preservative in a Microbank™ tube (Pro-Lab Diagnostics UK, United Kingdom). Approximately $2x10^9$ cells were used for high molecular weight DNA extraction using Nanobind CCB Big DNA Kit (Circulomics, Maryland, USA). DNA was quantified with the Qubit dsDNA HS assay in a Qubit 3.0 (Invitrogen). Long read

genomic DNA libraries were prepared with the Oxford Nanopore SQK-LSK109 kit with Native Barcoding EXP-NBD104/114 (ONT, UK), using 400-500 ng of high molecular weight DNA. Twelve to twenty-four barcoded samples were pooled in a single sequencing library and loaded on a FLO-MIN106 (R.9.4 or R.9.4.1) flow cell in a GridION (ONT, UK). Illumina reads were adapter trimmed using Trimmomatic 0.30 with a sliding window quality cutoff of Q15 (14). Unicycler v0.4.0 was used for genome assembly (15) and Prokka 1.11 to annotate contigs (16). Sequence data has been deposited at DDBJ/ENA/GenBank with the accession numbers CP099590 and CP099591 for WG1 and CP099588 and CP099589 for EMG2.

Bioinformatic analysis of genome sequences

For single nucleotide variant (SNV) calling, reads from EMG2 were aligned to the WG1 reference genome using BWA-Mem and processed using SAMtools 1.2. Variants were called using VarScan with two thresholds, sensitive and specific, where the variant allele frequency is greater than 90% and 10% respectively. The effects of variants were predicted and annotated using SnpEff. Draft genomes were visualized using Artemis (17), and comparisons between *E. coli* K-12 genomes were made using the Basic Local Alignment Search Tool (BLAST) at NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi), the Artemis Comparison Tool (ACT) (18) and the Proksee Server (https://proksee.ca/) (19). Genome representations were drawn using the Proksee Server (19) and ACT (18). Plasmid replicons were detected in draft genomes with PlasmidFinder 2.1 (5), using the software at the Center for Genomic Epidemiology (CGE) (http://www.genomicepidemiology.org/). Insertion sequences were located using ISfinder (https://www-is.biotoul.fr/blast/resultat.php) (20).

RESULTS

Comparison of the WG1 and EMG2 genomes

Whole genome sequencing of WG1 and EMG2 resulted in draft genome sequences, each comprising of 2 contigs; the larger contig, Contig 1, is the chromosomal sequence and the smaller, Contig 2, is the F plasmid (Figs. 2 and 3, Supplementary Fig. S1: Tables 1 and 2). Since both strains carry bacteriophage λ and the F plasmid, their genomes are slightly bigger than other sequenced *E. coli* K-12 strains, such as MG1655 and W3110 (Table 1) (4, 12). Comparison of the genomes of both WG1 and EMG2 with those of MG1655 and W3110 indicated that, unlike W3110, no major chromosomal rearrangements had occurred in these

strains (Supplementary Fig. S2) (19, 21). However, we identified a number of obvious regions of difference (Fig. 2 and Supplementary Fig. S1 and S3). For example, both EMG2 and W3110 have lost the cryptic prophage CPZ-55, and EMG2 has lost the *gatYZABDR* locus, which is involved in galactitol metabolism (22) (Fig. 2 and Supplementary Fig. S4). Interestingly, the *gatYZABDR* genes appear to have been a hotspot for insertion sequence element disruption in both MG1655 and W3110, which affects expression of this region (Supplementary Fig. S5) (22). Similarly, the region upstream of *flhDC* locus, which controls flagella production, also seems to have been targeted by different transposable elements (Fig. 2 and Supplementary Fig. S6) (23, 24). Note that strains that have been stored in agar stabs for many years accumulate deleterious mutations due to wholesale transposition of insertion sequences (25-27). As insertion of elements into this region influences motility, it is likely that the sequence heterogeneity found in this region produces a spectrum of effects (23, 24). For WG1, we detected the loss of cryptic prophage CP4-6 and a large deletion of the lipopolysaccharide O-antigen biosynthetic cluster, previously termed *rfb*-51 (Supplementary Figs. S1, S3 and S4) (8). Note that EMG2, MG1655 and W3110 carry the alternative *rfb*-50 mutation (an IS*5* disruption of the rhamnose transferase gene *wbbL*), which appears to be common to most *E. coli* K-12 strains (28), and so do not produce O-antigen either (Supplementary Figs. S4c) (8, 9). Loss of O-antigen production seems likely to be an adaptation to laboratory life, with both the first *E. coli* strain NCTC 86 (isolated in 1885), and commonly used B strains (*e.g.*, BL21(DE3)), all being rough in nature (13, 29, 30). In addition to these differences, WG1 also carries additional genes, encoding an LPS export ABC transporter permease (*lptG*), an acyl-carrier protein (*acpP*) and a NAD-dependent epimerase/dehydratase (*oleD*), which are flanked by IS*5* elements (Fig. 2 and Supplementary Fig. S7).

The F Plasmid

Comparison with MG1655 confirmed that both WG1 and EMG2 both carry the F plasmid, however, the two versions of F differ markedly in size, with that from EMG2 (99158 bp) similar in size to the previously sequence F plasmid (AP001918.1: 99159 bp), whilst F from WG2 is considerable smaller (67408 bp) (Table 2: Fig. 3 and Supplementary Fig. S8). This can be attributed to the loss of a large section of F in WG1, carrying the AIDA-I like autotransporter adhesin genes *ycbB* and *ychA*, the *ompP* omptin and the IncFIB replicon (Table 2: Fig. 3 and Supplementary Figs. S9) (7, 31). Surprisingly, F from WG1 carries additional DNA that is not found on F, which includes an IncFII RepA protein

(Supplementary Figs. S8 and S9). As in the previously sequenced F plasmid (AP001918.1), EMG2 F carries an IS*3* insertion in the *finO* gene, which leads to constitutive F transfer (7, 32, 33). However, this insertion sequence is absent from the WG1 F (Fig. 3; Supplementary Figs. S8 and S9), suggesting that conjugative transfer is regulated in this plasmid and that the insertion of IS*3* must have occurred in the immediate ancestor of EMG2. Thus, it is clear that F plasmids from both EMG2 and F have undergone significant lab-based evolution, resulting in two very different plasmids.

Bacteriophage λ

Comparison with MG1655 indicated that, as expected, both WG1 and EMG2 carry the bacteriophage λ prophage integrated between the *bioA* and *ybhC* genes (Fig. 2 and Supplementary Figs. S1 and S10). However, comparison with the previous sequenced λ genome (NC_001416) identified some differences in λ from WG1 and EMG2, in particular with the genes encoded tail fibres J, Stf and Tfa (Supplementary Fig. S11). Of note is *stf* (side tail fibre), which, in λ (NC_001416), carries a frame shift disrupting the gene into two ORFs (*orf-401* and *orf-314*) (34, 35). Bacteriophage λ carrying this lesion (λ PaPa) forms larger λ plaques (6, 35). Thus, as *stf* remains intact in WG1 and EMG2, it is likely that both strains would produce a small plaque phenotype (6, 35).

Similarities and differences between WG1 and EMG2

Single nucleotide variant calling showed that *E. coli* K-12 strains WG1 and EMG2 also differ in a number of key genes involved in important cellular functions (Supplementary File S1). For example, in EMG2, the gene encoding the major sigma factor σ^{70} (*rpoD*), carries a substitution, which results in Try at position 571 (Supplementary Fig. S12a). This is also found in MG1655 and W3110, whilst most *E. coli* strains carry His at this position. Substitutions at σ^{70} residue 571 have been shown to affect transcription at the *lac*, *araBAD*, *merT*, *merR* and the P22 phage *ant* promoters, as well as interfering with σ^{70} binding to core RNA polymerase and its ability to compete with alternative sigma factors (36-40). Conversely, in WG1 the gene encoding the α subunit of RNA polymerase, carries a mutation which results in a Gly to Arg substitution at position 311 (Supplementary Fig. S12b). This alteration affects expression from both the *merT* and *merR* promoters and the anaerobically activated *pepT* promoter in *Salmonella enterica* serovar Typhimurium (38, 41) (note that α in *E. coli* and *S. enterica* serovar Typhimurium are identical). As for many K-12 strains, both

WG1 and EMG2 carry a truncation in *rpoS*, which encodes the stress and stationary phase sigma factor σ^S (Supplementary Fig. S12c). (Note that the *rpoS* gene in MG1655 is the pseudo revertant *rpoS* 33Q allele)(4, 12). Additionally, *E. coli* K-12 strains also carry changes in genes that influence translation. Like MG1655 and W3110, EMG2 carries a mutation in the gene encoding release factor RF2 (*prfB*) (Thr at position 246) and a mutation in *rpsG* (30S ribosomal protein S7), which results in C-terminal extension of the S7 protein product (Supplementary Fig. S12d and e). Both substitutions have been shown to affect translation, with the mutation in RF2 resulting in poor termination at UGA stop codons and the trans-translational tagging of S7 with the SsrA peptide (42-46). Thus, it is clear that, for both EMG2 and WG1, adaptation to a laboratory lifestyle has resulted in strains, with altered transcription and translation machineries, that likely impact on global gene expression.

Our analysis also indentifies mutations in genes involved in metabolism and cellular homeostasis (Supplementary File S1). Similar to MG1655 and W3110, EMG2 carries a frame shift in *rph* (previously termed *rph-1)* that results in a truncation of RNase PH, which affects the expression of *pyrE*, manifesting in a pyrimidine starvation phenotype (Supplementary Fig. S12f) (47, 48). Like other K-12 strains, EMG2 also carries a mutation in *ilvG*, which produces a truncated protein product that affects branch chained amino acid biosynthesis (49) (Supplementary Fig. S12g). Whilst these mutations are absent from WG1, WG1 carries lesions in *mdtF* (an AcrB efflux pump homologue) and *nfi* (DNA repair endonuclease V), both of which result in truncated products (Supplementary Fig. S11h and i). Thus, WG1 is likely compromised in both drug efflux and DNA damage repair (50, 51).

DISCUSSION

The use of *E. coli* K-12 has shaped biological knowledge and research over the last century (52). Fred Neidhardt's comment that 'All cell biologists have at least two cells of interest: the one they are studying and *E. coli*' (53) still holds true for many scientists, with *E. coli* K-12 still the cornerstone of molecular biology and microbiology. However, it is clear that adaptation to the laboratory lifestyle has resulted in *E. coli* K-12 strains which have alterations in transcription, translation, general metabolism and cellular homeostasis. As *E. coli* K-12 strains EMG2, MG1655 and W3110 share many common alterations (*e.g.*, in *rpoD, prfB, rpsG, rph* (*rph-1*), *wbbL* (*rfb-50*), *prfB* and *ilvG*) this indicates that they share a similar lineage and that many of these mutations were fixed in their common ancestral strain (Fig. 1). On the other hand, WG1 carries alterations in different genes (e.g., *rfb-50, rpoA, mdtF* and

nfi), suggesting that it is distinct from these strains (Fig. 1). It is worth noting that WG1 is similar to *E. coli* strains NCM3722 (54) and LS5218 (55). Strain NCM3722 (CGSC#12355) was first detailed by Sydney Kustu (48) and LS5218 is an industrial strain used for the production of fatty acid derived products (55). Both strains carry bacteriophage λ, a smaller version of the F plasmid (Tables 1 and 2: Supplementary Figs. S13 and S14) and contain many of the mutations carried by WG1 (54, 55).

In addition to lineage specific mutations, it is clear that WG1 and EMG2 have undergone their own lab-based evolution events, such as loss of cryptic prophages and gene disruption. The suggestion is that the selection of particular traits by microbiologists has driven lab-based evolution. Hence, IS inactivation of *finO* in F made plasmid transfer easier to study, larger plaques enabled the intricacies of λ lysogeny to be examined and lack of O-antigen enhances plasmid transformation (6, 7, 30, 35). Thus, our interpretation of *E. coli* biology has been inadvertently biased. Moreover, many other laboratory strains, handed down for generations, are as yet unsequenced, so it is unclear what other changes lie within those strains.

Heterogeneity in bacterial lab strains and plasmids has been observed many times and we are at a stage when even the same *E. coli* K-12 stock strains can produce different outcomes, calling reproducibility into question (27, 48, 56-59). It is clear that there are significant major differences between K-12 and other commensal *E. coli* strains, and these differences became fixed in the ancestors of the very widely used MG1655 and W3110 strains. Given the different mutations seen in WG1 compared to EMG2, it seems likely that identical or similar mutations will be present in other K-12 lineages. However, due to the extensive genetic systems that have been developed, demonstration of safe use, and lack of ability to colonize humans, *E. coli* K-12 strains will justifiably continue to be widely used (10, 11, 52). We think it is important that there is an awareness of the mutations present in K-12 strains, and the effects of these mutations on the physiology and metabolism of these strains. An understanding of the conditions that might select for mutants in laboratories, and the use of cost effective and accurate sequencing of laboratory stocks should help to prevent further undetected mutations arising in K-12 strains, which could compromise our understanding of fundamental biological processes. Thus, it is hoped that the next century will continue to provide more insight into the complex biology and evolution of this versatile organism. Indeed, appreciation of various K-12 strains, as well differences between various bacterial families, is sure to enhance our understanding of life.

Funding information

This work was generously supported by BBSRC research grants BB/R017689/1 to DFB and SJWB and BBSRC BB/E01044X/1 to JLH.

Acknowledgements

We thank MicrobesNG for genome sequencing, in particular Andrew Holmes and Emily Jane Richardson for bioinformatics and SNV calling.

Author contributions

DFB, JLH and SJWB conceived the study, selected samples, carried out bioinformatic analyses and wrote the manuscript. All authors read and approved the final version of the manuscript.

Conflicts of interest.

The authors declare that there are no conflicts of interest.

Ethical statement.

No ethical clearance was required for this study.

Consent to publish

All authors give their consent to publish.

References.

1. Bachmann BJ. Pedigrees of some mutant strains of *Escherichia coli* K-12. Bacteriol Rev. 1972;36(4):525-57.
2. Clowes RC, Hayes W. Experiments in Microbial Genetics. . Oxford ; Edinburgh: Blackwell Scientific Publications.; 1968.
3. Hobman JL, Penn CW, Pallen MJ. Laboratory strains of *Escherichia coli*: model citizens or deceitful delinquents growing old disgracefully? Molecular microbiology. 2007;64(4):881-5.
4. Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of *Escherichia coli* K-12. Science (New York, NY). 1997;277(5331):1453-62.
5. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, et al. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrobial agents and chemotherapy. 2014;58(7):3895-903.
6. Casjens SR, Hendrix RW. Bacteriophage lambda: Early pioneer and still relevant. Virology. 2015;479-480:310-30.
7. Koraimann G. Spread and Persistence of Virulence and Antibiotic Resistance Genes: A Ride on the F Plasmid Conjugation Module. EcoSal Plus. 2018;8(1).
8. Liu D, Reeves PR. *Escherichia coli* K12 regains its O antigen. Microbiology (Reading). 1994;140 (Pt 1):49-57.
9. Stevenson G, Neal B, Liu D, Hobbs M, Packer NH, Batley M, et al. Structure of the O antigen of *Escherichia coli* K-12 and the sequence of its *rfb* gene cluster. Journal of bacteriology. 1994;176(13):4144-56.
10. Smith HW. Survival of orally administered *E. coli* K 12 in alimentary tract of man. Nature. 1975;255(5508):500-2.
11. Anderson ES. Viability of, and transfer of a plasmid from, *E. coli* K12 in human intestine. Nature. 1975;255(5508):502-4.
12. Hayashi K, Morooka N, Yamamoto Y, Fujita K, Isono K, Choi S, et al. Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. Molecular systems biology. 2006;2:2006.0007.
13. Dunne KA, Chaudhuri RR, Rossiter AE, Beriotto I, Browning DF, Squire D, et al. Sequencing a piece of history: complete genome sequence of the original *Escherichia coli* strain. Microbial genomics. 2017;3(3):mgen000106.
14. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England). 2014;30(15):2114-20.
15. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS computational biology. 2017;13(6):e1005595.
16. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics (Oxford, England). 2014;30(14):2068-9.
17. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinformatics (Oxford, England). 2000;16(10):944-5.
18. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. ACT: the Artemis Comparison Tool. Bioinformatics (Oxford, England). 2005;21(16):3422-3.
19. Grant JR, Stothard P. The CGView Server: a comparative genomics tool for circular genomes. Nucleic acids research. 2008;36(Web Server issue):W181-4.
20. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. Journal of computational biology : a journal of computational molecular cell biology. 2000;7(1-2):203-14.
21. Hill CW, Harnish BW. Inversions between ribosomal RNA genes of *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America. 1981;78(11):7069-72.
22. Nobelmann B, Lengeler JW. Molecular analysis of the *gat* genes from *Escherichia coli* and of their roles in galactitol transport and metabolism. Journal of bacteriology. 1996;178(23):6790-5.

23. Zhang Z, Kukita C, Humayun MZ, Saier MH. Environment-directed activation of the *Escherichia coli flhDC* operon by transposons. Microbiology (Reading). 2017;163(4):554-69.
24. Barker CS, Prüss BM, Matsumura P. Increased motility of *Escherichia coli* by insertion sequence element integration into the regulatory region of the *flhD* operon. Journal of bacteriology. 2004;186(22):7529-37.
25. Naas T, Blot M, Fitch WM, Arber W. Insertion sequence-related genetic variation in resting *Escherichia coli* K-12. Genetics. 1994;136(3):721-30.
26. Naas T, Blot M, Fitch WM, Arber W. Dynamics of IS-related genetic rearrangements in resting *Escherichia coli* K-12. Mol Biol Evol. 1995;12(2):198-207.
27. Freddolino PL, Amini S, Tavazoie S. Newly identified genetic variations in common *Escherichia coli* MG1655 stock cultures. Journal of bacteriology. 2012;194(2):303-6.
28. Kuhnert P, Nicolet J, Frey J. Rapid and accurate identification of *Escherichia coli* K-12 strains. Applied and environmental microbiology. 1995;61(11):4135-9.
29. Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, et al. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). Journal of molecular biology. 2009;394(4):644-52.
30. Browning DF, Wells TJ, Franca FL, Morris FC, Sevastsyanovich YR, Bryant JA, et al. Laboratory adapted *Escherichia coli* K-12 becomes a pathogen of *Caenorhabditis elegans* upon restoration of O antigen biosynthesis. Molecular microbiology. 2013;87(5):939-50.
31. Yen MR, Peabody CR, Partovi SM, Zhai Y, Tseng YH, Saier MH. Protein-translocating outer membrane porins of Gram-negative bacteria. Biochimica et biophysica acta. 2002;1562(1-2):6-31.
32. Cheah KC, Skurray R. The F plasmid carries an IS3 insertion within finO. J Gen Microbiol. 1986;132(12):3269-75.
33. Yoshioka Y, Ohtsubo H, Ohtsubo E. Repressor gene finO in plasmids R100 and F: constitutive transfer of plasmid F is caused by insertion of IS3 into F *finO*. Journal of bacteriology. 1987;169(2):619-23.
34. Sanger F, Coulson AR, Hong GF, Hill DF, Petersen GB. Nucleotide sequence of bacteriophage lambda DNA. Journal of molecular biology. 1982;162(4):729-73.
35. Hendrix RW, Duda RL. Bacteriophage lambda PaPa: not the mother of all lambda phages. Science (New York, NY). 1992;258(5085):1145-8.
36. Siegele DA, Hu JC, Gross CA. Mutations in *rpoD*, the gene encoding the sigma 70 subunit of *Escherichia coli* RNA polymerase, that increase expression of the *lac* operon in the absence of CAP-cAMP. Journal of molecular biology. 1988;203(1):29-37.
37. Siegele DA, Hu JC, Walter WA, Gross CA. Altered promoter recognition by mutant forms of the sigma 70 subunit of *Escherichia coli* RNA polymerase. Journal of molecular biology. 1989;206(4):591-603.
38. Caslake LF, Ashraf SI, Summers AO. Mutations in the alpha and sigma-70 subunits of RNA polymerase affect expression of the *mer* operon. Journal of bacteriology. 1997;179(5):1787-95.
39. Jishage M, Kvint K, Shingler V, Nyström T. Regulation of sigma factor competition by the alarmone ppGpp. Genes & development. 2002;16(10):1260-70.
40. Cicero MP, Sharp MM, Gross CA, Kreuzer KN. Substitutions in bacteriophage T4 AsiA and *Escherichia coli* sigma(70) that suppress T4 motA activation mutations. Journal of bacteriology. 2001;183(7):2289-97.
41. Lombardo MJ, Bagga D, Miller CG. Mutations in *rpoA* affect expression of anaerobically regulated genes in *Salmonella typhimurium*. Journal of bacteriology. 1991;173(23):7511-8.
42. Schaub RE, Hayes CS. Deletion of the RluD pseudouridine synthase promotes SsrA peptide tagging of ribosomal protein S7. Molecular microbiology. 2011;79(2):331-41.
43. Pavlov MY, Freistroffer DV, Dincbas V, MacDougall J, Buckingham RH, Ehrenberg M. A direct estimation of the context effect on the efficiency of termination. Journal of molecular biology. 1998;284(3):579-90.
44. Collier J, Binet E, Bouloc P. Competition between SsrA tagging and translational termination at weak stop codons in *Escherichia coli*. Molecular microbiology. 2002;45(3):745-54.

45. Dinçbas-Renqvist V, Engström A, Mora L, Heurgué-Hamard V, Buckingham R, Ehrenberg M. A post-translational modification in the GGQ motif of RF2 from *Escherichia coli* stimulates termination of translation. The EMBO journal. 2000;19(24):6900-7.
46. Uno M, Ito K, Nakamura Y. Functional specificity of amino acid at position 246 in the tRNA mimicry domain of bacterial release factor 2. Biochimie. 1996;78(11-12):935-43.
47. Jensen KF. The *Escherichia coli* K-12 "wild types" W3110 and MG1655 have an *rph* frameshift mutation that leads to pyrimidine starvation due to low *pyrE* expression levels. Journal of bacteriology. 1993;175(11):3401-7.
48. Soupene E, van Heeswijk WC, Plumbridge J, Stewart V, Bertenthal D, Lee H, et al. Physiological studies of *Escherichia coli* strain MG1655: growth defects and apparent cross-regulation of gene expression. Journal of bacteriology. 2003;185(18):5611-26.
49. Lawther RP, Calhoun DH, Gray J, Adams CW, Hauser CA, Hatfield GW. DNA sequence fine-structure analysis of *ilvG* (*IlvG+*) mutations of *Escherichia coli* K-12. Journal of bacteriology. 1982;149(1):294-8.
50. Schuster S, Vavra M, Greim L, Kern WV. Exploring the Contribution of the AcrB Homolog MdtF to Drug Resistance and Dye Efflux in a Multidrug Resistant *E. coli* Isolate. Antibiotics (Basel, Switzerland). 2021;10(5).
51. Weiss B. Endonuclease V of *Escherichia coli* prevents mutations from nitrosative deamination during nitrate/nitrite respiration. Mutat Res. 2001;461(4):301-9.
52. Zimmer C. Microcosm: *E. coli a*nd the New Science of Life. New York, NY: Pantheon Books; 2008.
53. Neidhardt FC, Curtiss R. *Escherichia coli* and *Salmonella* : cellular and molecular biology. Washington, D.C: American Society for Microbiology Press; 1996.
54. Brown SD, Jun S. Complete Genome Sequence of *Escherichia coli* NCM3722. Genome announcements. 2015;3(4).
55. Rand JM, Gordon GC, Mehrer CR, Pfleger BF. Genome sequence and analysis of *Escherichia coli* production strain LS5218. Metab Eng Commun. 2017;5:78-83.
56. Nahku R, Peebo K, Valgepea K, Barrick JE, Adamberg K, Vilu R. Stock culture heterogeneity rather than new mutational variation complicates short-term cell physiology studies of *Escherichia coli* K-12 MG1655 in continuous culture. Microbiology (Reading). 2011;157(Pt 9):2604-10.
57. Spira B, de Almeida Toledo R, Maharjan RP, Ferenci T. The uncertain consequences of transferring bacterial strains between laboratories - *rpoS* instability as an example. BMC microbiology. 2011;11:248.
58. Hooton SPT, Pritchard ACW, Asiani K, Gray-Hammerton CJ, Stekel DJ, Crossman LC, et al. Laboratory Stock Variants of the Archetype Silver Resistance Plasmid pMG101 Demonstrate Plasmid Fusion, Loss of Transmissibility, and Transposition of Tn*7/pco/sil* Into the Host Chromosome. Frontiers in microbiology. 2021;12:723322.
59. Jishage M, Ishihama A. Variation in RNA polymerase sigma subunit composition within different stocks of *Escherichia coli* W3110. Journal of bacteriology. 1997;179(3):959-63.

Table 1. Comparison of the genomes of different *E. coli* K-12 laboratory strains.

	WG1	**EMG2**	**MG1655**	**W3110**	**NCM3722**	**LS5218**
Accession N°	CP099590 CP099591	CP099588 CP099589	NC_000913.3	NC_007779.1	CP011495.1 CP011496.1	MVJG000000 00.1
Genome size [a]	4,735,495	4,774,480	4,641,652	4,646,332	4,745,591	4,699,198
Plasmid	F plasmid	F plasmid	None	None	F plasmid	F plasmid
Total N° CDSs [b]	4431	4457	4285	4213	4539	4368
GC content	50.75 %	50.73 %	50.79 %	50.8 %	50.76 %	50.72 %

[a] Genome size (bp) includes the F plasmid for WG1, EMG2, NCM3722 and LS5218.

[b] Number of coding sequences (CDSs) is as predicted by each genome annotation.

435 **Table 2. Comparison of the F plasmid from different *E. coli* K-12 laboratory strains.**

	WG1	**EMG2**	**F Plasmid**	**NCM3722**	**LS5218**
Accession	CP099591	CP099589	AP001918.1	CP011496.1	CM007715.1
F Plasmid size	67,408 bp	99,158 bp	99,159 bp	67,545 bp	67,502 bp
Total N° CDSs [a]	73	98	105	79	83
GC content	51.66 %	48.17 %	48.17 %	51.67 %	51.67 %
Plasmid Replicons [b]	IncFIA, IncFIC(FII)	IncFIA, IncFIB, IncFIC(FII)	IncFIA, IncFIB, IncFIC(FII)	IncFIA, IncFIC(FII)	IncFIA, IncFIC(FII)

436

437 [a] Number of CDSs is as specified by genome annotation.

438 [b] Plasmid replicons were detected using PlasmidFinder 2.1 using software at the CGE (5).

439

440

FIGURE LEGENDS

Fig. 1. The pedigree of *Escherichia coli* K-12 strains. The figure details the pathway of *E. coli* K-12 evolution from its isolation in 1922 to the generation of MG1655 and W3100 strains (1, 3, 4, 12). Blood agar indicates selection on blood agar plates; UV, irradiation with ultraviolet light; EMB-gal, selection for utilization of galactose on eosin methylene blue indicator plates. Dotted lines represent uncertain evolutionary lineage events.

Fig. 2. Genome comparison of different *E. coli* K-12 strains. The figure shows the comparison of the WG1 chromosome (contig 1) and F plasmid (contig 2) with the genomes of EMG2, MG1655 (NC_000913.3) and W3110 (NC_007779.1), using the Proksee Server (19). The outer two rings display the genes and features of the WG1 genome, with selected genes and differences labelled. The green, brown and blue rings illustrate the BLAST results when the genome sequences of *E. coli* K-12 strains EMG2, MG1655 (NC_000913.3) and W3110 (NC_007779.1), respectively, are compared to that of WG1.

Fig. 3. Comparison of the F plasmids from different *E. coli* K-12 strains. The figure shows the comparison of the F plasmid (AP001918.1) with that from EMG2 and WG1 using Proksee (19). The outer two rings display the genes and features of the F plasmid, with selected genes labelled. The green and brown rings illustrate the BLAST results when the F plasmid sequences from EMG2 and WG1, respectively, are compared to the original F plasmid sequence.

Fig. 1. Browning *et al.* (2022)

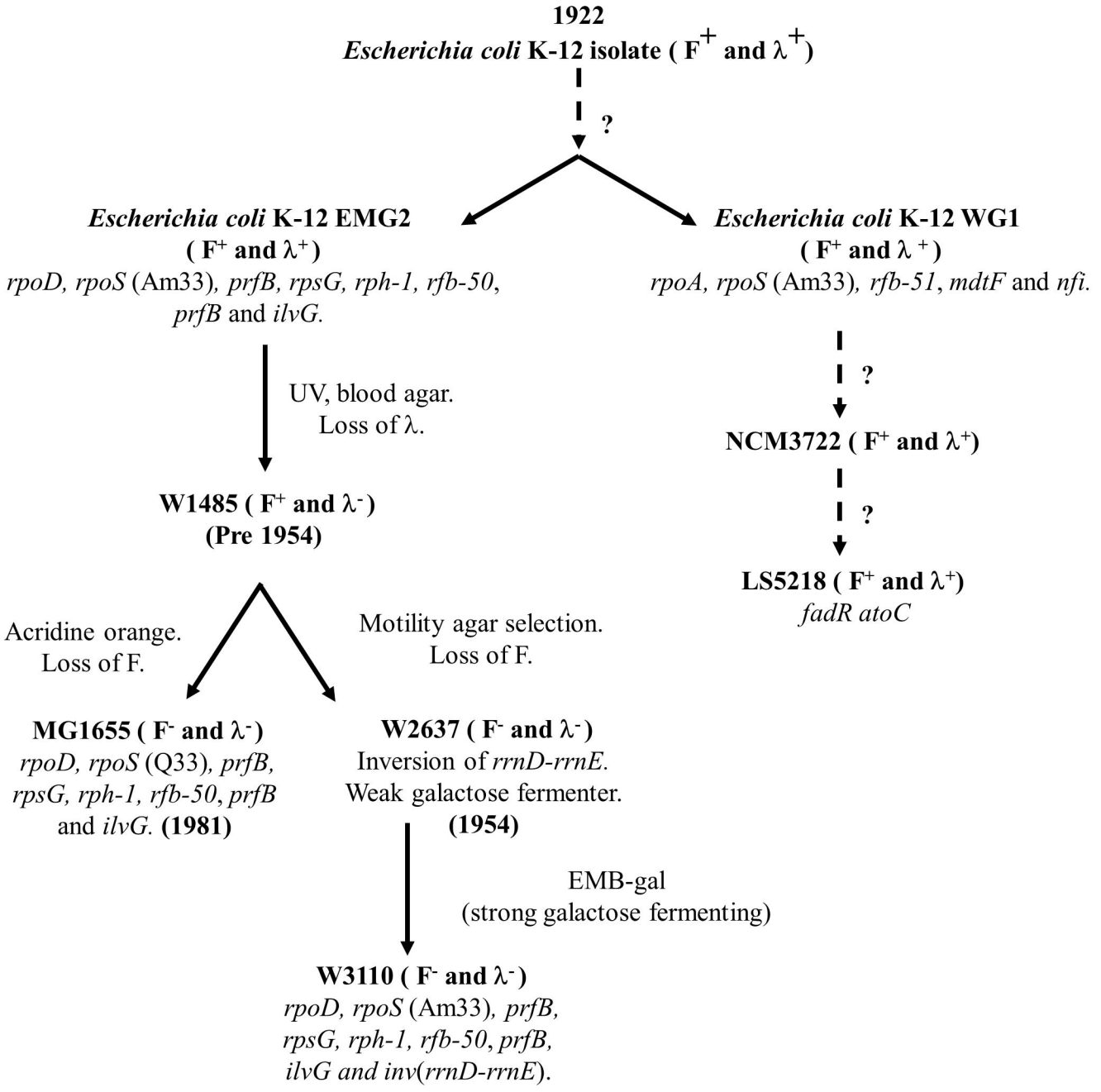

Fig. 2. Browning *et al.* (2022)

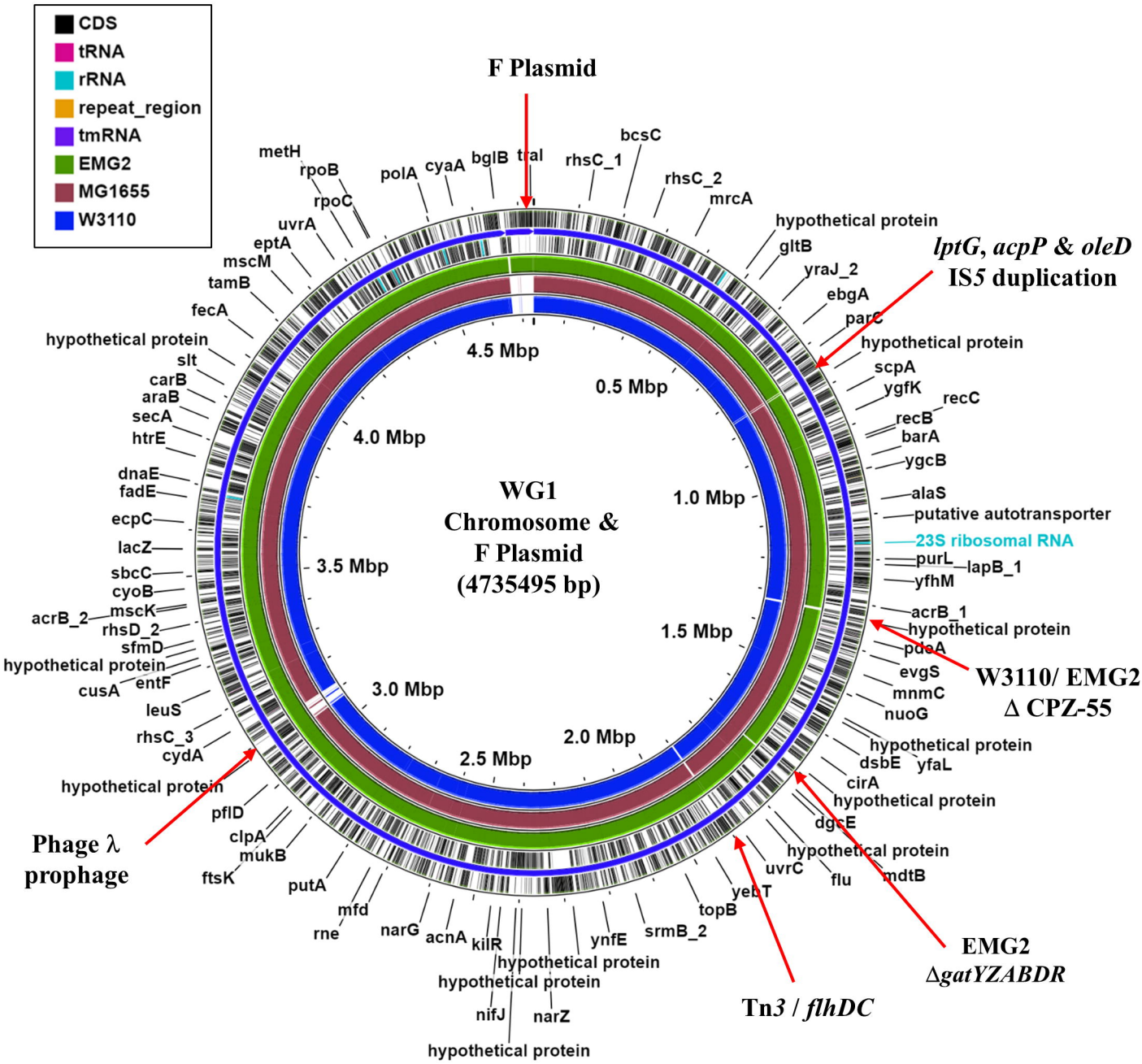

Fig. 3. Browning *et al.* (2022)