# The genome sequence of *E. coli* OP50

Robin C. May[1], Nicholas J. Loman[1], Anthony S. Haines[1], Mark J. Pallen[1], Claudia Boehnisch[1], Charles W. Penn[1], Choong Hoon Lee[2] and Jihyun Kim[2]

[1] School of Biosciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT. U.K., [2]Functional Genomics Program, School of Science, University of Science and Technology, Yuseong, Daejeon 305-333, Korea

Correspondence to: Robin May r.c.may@bham.ac.uk

The vast majority of *C. elegans* strains worldwide are maintained on lawns of *Escherischia coli* OP50. As a resource for the *C. elegans* community, and potentially for evolutionary biologists interested in mapping changes in this strain following parallel evolution in laboratories around the world, we have sequenced the genome of *E. coli* OP50 using Solexa sequencing.

Two cultures of *E. coli* OP50, both taken from our laboratory stock (which was originally obtained approximately five years ago from the group of Ronald Plasterk) were grown overnight in liquid broth and DNA then isolated using the QIAgen genomic DNA isolation kit. Subsequent Solexa sequencing (performed at the Genome Center at Washington University) on the pooled DNA yielded a sequencing depth of approximately 100-fold.

Pairwise comparison of the genome to other *E. coli* genome sequences using myXBase (http://my.xbase.ac.uk/) identified the *E. coli* strain REL606 as the closest genetic match. Since REL606 is an *E. coli* B strain, this confirms the widespread annotation of *E. coli* OP50 as an *E. coli* B strain.

Preliminary mapping of the OP50 genome to REL606 as a reference identified a total of 977 putative single nucleotide polymorphisms. 509 of these are predicted to have little effect (132 falling in intergenic regions, 377 being synonymous substitutions). 451 are missense mutations, 15 introduce premature stops into (putative) proteins and two are in tRNA sequences. Interestingly, we identified missense mutations in two putative uracil permeases/transporters, which may explain the reported uracil auxotrophy of OP50. Finally, several of the missense or nonsense mutations lie in chemotaxis or flagellar genes, suggesting that OP50 is likely to have lost the capacity for directional motility after decades in laboratory culture. In line with this, our preliminary experiments indicate that OP50 is non-motile on soft agar.

The genome sequence for *E. coli* OP50 is available at the NCBI Short Read Archive under accession numbers SRX012490 or SRA010042 (http://www.ncbi.nlm.nih.gov/sites/entrez?db=sra&term=SRX012490&report=full) and via the NCBI Genome Project site (accession 41499, http://www.ncbi.nlm.nih.gov/sites/entrez?Db=genomeprj&cmd=ShowDetailView&TermToSearch=41499). In addition a spreadsheet listing all identified mutations relative to REL606 is available from our webpage (http://www.biosciences.bham.ac.uk/labs/may/May_lab_pages/Worms.html). We very much hope that our preliminary analysis will encourage others to produce genome data on "their" *E. coli* OP50 in order to facilitate the analysis of evolutionary changes in this lineage and to enable a more detailed characterization of this strain.

Note added by Sydney Brenner: OP50 is a mutant of *E. coli* B (Berkeley strain). The Berkeley strain grew rI and rII mutants of T4 as r+.