SEPARATION AND IDENTIFICATION OF BACTERIOPHAGES USING SINGLE-MOLECULE DNA BARCODES

Nathaniel Wand[1,2], Stephen J. W. Busby[3], Iain B. Styles[1,4], Robert K. Neely[1,2]
[1]PSIBS Doctoral Training Centre, [2]School of Chemistry, [3]School of Biosciences, [4]School of Computer Sciences, University of Birmingham, B15 2TT Birmingham, United Kingdom
Email: now348@bham.ac.uk

KEY WORDS: DNA Methyltransferases, single-molecule imaging, DNA combing, optical mapping

DNA methyltransferases can be used to label DNA sequence-specifically with, for instance, fluorescent probes. This is a novel technology which allows DNA to be targeted without damage, with high efficiency, and crucially allows sequence context. If fluorophore-labelled DNA is combed onto a surface, single molecules of DNA can be used as a 'barcode', and read and analysed much like any other barcode [1].

Currently the leading approaches for identification of microorganisms are next generation sequencing methods. Typically short sequence reads are used and can make sequence assembly difficult, particularly in complex and repetitive genomes. Using longer sequence reads is still relatively expensive and reduces overall coverage. The identification of organisms within a mixed population is particularly challenging. The DNA barcode can act as a complementary technique, allowing direct visualisation of the DNA sequence. As well as acting as a scaffold for sequence assembly, this barcode can be used for bacteriophage strain typing [2].

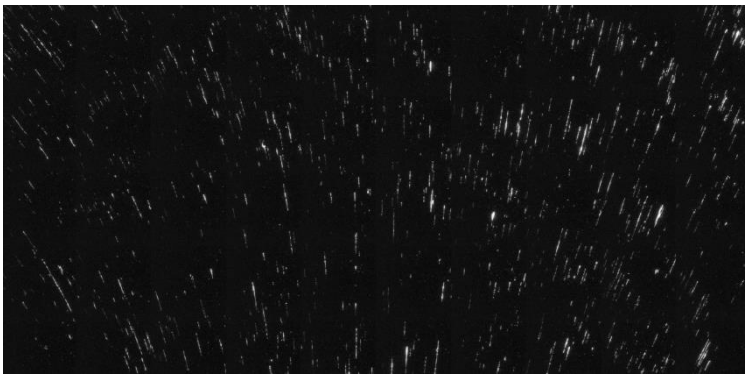

Figure 1: A tiled fluorescence microscopy image of DNA 'fluorocodes'. T7 bacteriophage DNA is labelled with Atto647N at TCGA sites and then deposited by molecular combing

Here a new approach is described which can rapidly separate and identify mixtures of genomic DNA. Individual DNA barcodes are extracted from scans such as Figure 1 and the similarity between barcodes calculated using the normalised cross correlation. Similar DNA barcodes may then be separated using t-Distributed Stochastic Neighbour Embedding (t-SNE), in which similar objects are clustered in 2d-space. An average barcode can be derived and compared to a library of known sequences to identify the DNA. Here we report that a mixture T7 and λ bacteriophage DNA (both around ~50kb) has been separated and identified. Results from modelling suggest the same methods may be used to identify larger genomes, such as *E. Coli* (~2000kb).

[1] C. Vranken, J. Deen, L. Dirix, T. Stakenborg, W. Dehaen, V. Leen, J. Hofkens, and R. K. Neely, "Super-resolution optical DNA mapping via DNA methyltransferase-directed click chemistry," *Nucleic Acids Res.*, **42**, e50 (2014)
[2] A. Grunwald, *et al,* "Bacteriophage strain typing by rapid single molecule analysis," *Nucleic Acids Res.*, **43**, e117 (2015)