

## GENOME WATCH

# Playing the 'Next-Generation Game'

Helena Seth-Smith

Advances in single-molecule DNA sequencing are enabling research into the fine resolution of DNA structure, and rapid, direct sequencing of pathogen genomes.

New sequencing technologies are continuing to develop, meaning that many distinct sequencing platforms are now available, each suited to answering different genomic questions. A series of recently published studies used one such platform to study the methylation patterns in bacterial genomes, which can provide important biological insights.

Single-molecule, real-time (SMRT) sequencing, as pioneered by Pacific Biosciences with their RS sequencing platform, detects the kinetics of a single DNA polymerase as it incorporates nucleotides during DNA synthesis. Modifications to individual bases can be detected during routine sequencing of native DNA, because the polymerase pauses during base incorporation. Methylation is the most common form of DNA modification and is important for protecting bacteria from infecting bacteriophages through the activity of methyltransferases (MTases), which are often found as part of restriction–modification (RM) systems. From a practical point of view, information on the methylation status of genomic DNA has the potential to inform protocol improvements for use with bacteria that prove difficult to transform in laboratory settings.

Two publications from laboratories at Pacific Biosciences have provided a proof

of principle for using this platform to identify methylated sites in bacterial genomes<sup>1,2</sup>. Clark *et al.* first confirmed the utility of the approach by sequencing plasmids isolated from bacterial strains in which MTases were overexpressed. The authors were able to determine the target consensus sites of several MTases, as well as noting the specificity of the reactions through monitoring suspected off-target activities that were occasionally produced by some MTases. The first bacterial genome-wide methylation screens were carried out by Murray *et al.* By analysing several bacterial species, this team determined new genome sequences and also gained a picture of the effect of native MTases *in vivo*.

The 'methylome' of the haemolytic–uremic syndrome-linked *Escherichia coli* O104:H4 str. C227-11, isolated during the outbreak in Germany in 2011, has also been studied in detail<sup>3</sup>. This strain contains genes encoding ten putative adenine-specific MTases and a putative cytosine-specific MTase. Within the genome, 51,972 bases (1.9% of the total) were detected as being methylated, of which 49,311 were found to be N<sup>6</sup>-methyladenine (m<sup>6</sup>A). The majority of these (39,324; 79%) were found in the consensus GATC, which may be targeted by three Dam-like MTases in this strain. Surprisingly, these sites do not represent all the GATC sites within the genome, as a handful of loci (20 confirmed) with this consensus were found to be unmethylated. Many of these unmethylated sites were in the vicinity of genes involved in the phosphotransferase transport system, raising intriguing questions about the regulation of this activity.

*E. coli* O104:H4 str. C227-11 contains the prophage  $\Phi$ Stx104, which is one of the defining factors of the German outbreak strains and carries the Shiga toxin as well as an MTase, *M.EcoGIII*. The target consensus of *M.EcoGIII* was

identified as CTGCAG, in which the adenine is modified to m<sup>6</sup>A in 96% of the 2,486 target sites in the genome. The presence of *M.EcoGIII* affects the growth rate of the strain and also has transcriptional consequences: one-third of the genes in the genome were found to be differentially expressed when the RM locus was deleted, as determined by RNA sequencing carried out on an Illumina platform. Thus, the role of MTases might extend beyond that of protection against foreign DNA.

In a further development using the same third-generation sequencing technology, it has been found that plasmids and small genomes can be sequenced directly without the need for a standard library protocol<sup>4</sup>. A group at the Sanger Institute, Cambridge, UK, have developed a methodology using random hexamers to prime from directly extracted DNA, meaning that genomic data (including m<sup>6</sup>A sites) could be obtained from samples containing as little as 1 ng of DNA and in only 8 hours. This has clear potential for applications in outbreak situations, particularly when plasmid identification might be useful.

As these studies demonstrate, next-generation technologies are illuminating many aspects of genome biology above and beyond the sequence of the four DNA bases.

Helena Seth-Smith is at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.  
e-mail: [microbes@sanger.ac.uk](mailto:microbes@sanger.ac.uk)  
doi:10.1038/nrmicro2956

1. Clark, T. A. *et al.* Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res.* **40**, e29 (2012).
2. Murray, I. A. *et al.* The methylomes of six bacteria. *Nucleic Acids Res.* **40**, 11450–11462 (2012).
3. Fang, G. *et al.* Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nature Biotech.* **30**, 1232–1239 (2012).
4. Coupland, P. *et al.* Direct sequencing of small genomes on the Pacific Biosciences RS without library preparation. *BioTechniques* **53**, 365–372 (2012).

**Competing interests statement**  
The author declares no competing financial interests.

