Bacterial comparative genomics and transcriptomics using long nanopore reads

Long reads enable complete assembly of bacterial genomes, allowing us to compare genomes between strains and species, and to monitor the transcriptome-wide response to different stimuli.

Contact: publications@nanoporetech.com  More information at:  www.nanoporetech.com and publications.nanoporetech.com

---

Single-contig assembly of bacterial genomes using long 1D² reads and Canu

We extracted genomic DNA from a culture of E. coli strain NEB 5-alpha, prepared a 1D² library, using an SQK-LSK338 kit, and sequenced this library on a FLO-MIN107 flowcell. We basecalled with Albacore 2.0.2 and assembled 8,000 of the 1D² reads (160 Mb) with Canu 1.5, using an error-correction rate of 0.04. We obtained a single-contig assembly and constructed a Mummer plot of this assembly against the reference sequence of the NEB 5-alpha strain (Fig. 1a) for confirmation that our assembly was correct. We also constructed a Mummer plot against the K12 MG1655 strain, to reveal any large-scale structural differences (Fig. 1a). Several differences were evident, and we investigated these further.

---

Nucleotide-level resolution of breakpoints by comparison to the annotated reference

We performed global alignment of our assembly to the MG1655 reference genome using MAUVE, to detect and render any structural variants (insertions, deletions, translocations and inversions) and to locate the resulting breakpoints. By comparing the assembly to the annotated reference, we were able to identify affected genes and mobile elements (Fig. 2). The easy availability of fully-closed bacterial genome assemblies using long nanopore reads will simplify the identification and characterisation of bacterial strains, establishing their correct phylogenetic placement, and will allow us to monitor the exchange of genetic material – such as antibiotic-resistance cassettes – between strains and species.

---

Comparative transcriptomics: PCR-free sequencing of full-length cDNAs allows us to monitor changes in expression levels under different growth conditions

To investigate transcriptomic changes under different growth conditions, we grew E. coli strain NEB 5-alpha cells in LB broth, and measured the optical density at regular intervals, to identify the growth stage of the culture. We took samples of the culture in both log-growth phase and stationary phase, transferring the cells immediately to a solution of RNAlater, before extraction of total RNA using Qiagen's RNeasy kit. In our experience, this kit provides a simple and efficient way to obtain full-length RNAs from gram negative bacteria. We added a polyA tail to the total RNA using E. coli PolyA cDNA protocol and sequenced each library on a FLO-MIN006 flowcell, generating 3 Gb of sequence per library. Using BWA-mem, we aligned the cDNA reads to our NEB 5-alpha assembly and generated a Circos plot to display the differences in expression level (Fig. 3a). We converted the alignment bam files to bigWig files using deepTools, and visualised the alignments in IGV alongside a genome annotation from strain MG1655, which we added by gff liftover. There were differences in the expression levels of many genes. Among the genes with the most significantly different expression levels, we found csgA/E and mhpD to be more highly expressed in stationary phase. These genes are involved in biofilm formation and quorum-sensing pathways, respectively. Conversely, we found the expression of metE and mhpD to be substantially higher in log phase. These genes are involved in energy transfer and amino-acid synthesis (Figs. 3b–e).

---

Contact: publications@nanoporetech.com  More information at:  www.nanoporetech.com and publications.nanoporetech.com

© 2017 Oxford Nanopore Technologies. All rights reserved.