Title: Dynamic, adaptive sampling during nanopore sequencing using Bayesian experimental design

Abstract:
Nanopore sequencers can select which DNA molecules to sequence by rejecting a molecule after analysis of a small initial part. Until recently, selection was based on predetermined regions of interest that remain constant throughout an experiment. Sequencing therefore could not be re-focused on molecules likely contributing most to experimental success.
We present BOSS-RUNS, an algorithmic framework and software to generate dynamically updated decision strategies. We quantify uncertainty at each genome position with real-time updates from observed data. For each DNA fragment, we decide whether the expected decrease in uncertainty that it would provide warrants fully sequencing it, thus optimizing information gain.
We show how this can mitigate coverage bias both within a microbial community, which leads to improved variant calling; for example, low-coverage sites of a species at 1% abundance were reduced by 87.5%, with 12.5% more single-nucleotide polymorphisms detected.
Further, we now expand our method to allow for true de novo enrichment, i.e. without requiring prior information about sample composition. In this reference-free approach we construct and incrementally update genome assemblies in real-time. These are used to inform the decision process regarding rejecting DNA fragments from well-assembled or over-represented regions and focusing on extending contigs instead. Thus, coverage biases can be mitigated in a similar manner without the need for input genomes.
Overall, these data-driven updates to molecule selection are applicable to many sequencing scenarios, such as enrichment of regions with increased divergence or low coverage, or even true de novo enrichment of unknown species in mixed samples.