Nanopore Adaptive Sampling: A tool for enrichment of low abundance species in metagenomic samples

Samuel Martin[1], Darren Heavens[1], Yuxuan Lan[1], Samuel Horsfield[1], Matthew D Clark[2], Richard M Leggett[1]

[1] Earlham Institute, Norwich, UK [2] Natural History Museum, London, UK

Nanopore sequencing devices enable software-controlled enrichment (termed "adaptive sampling") of DNA molecules. By aligning molecules to a reference database as they travel through the pore, molecules that align poorly can be ejected, thus freeing the pore to sequence a new molecule. We determined the effect that molecule length has on the efficiency of enrichment by sequencing a synthetic bacterial mock with targeted molecule lengths (mean values 1.7kbp, 4.7kbp, 10.6kbp, 12.8kbp). We performed 5 experiments, in which we enriched for each species in the mock for one hour, and performed a control run for comparison.

A Mathematical Model of Enrichment:

We developed a simple mathematical model to predict the enrichment of a single species in a metagenomic sample, using factors such as species abundance and mean molecule length.

$$e_x = \frac{R + CS}{x_{\mathrm{ab}}R + (1 - x_{\mathrm{ab}})DS + CS}$$

Where, for a species denoted by x we have e_x is the predicted enrichment of x, x_{ab} is the abundance of x, R is the mean molecule length (in bases), C is the time for a pore to capture a molecule, D is the time taken for a decision to be made on whether a molecule should be sequenced or ejected from the pore (seconds), and S is the sequencing rate (bases per second). We also developed a shiny app, where researchers can examine how adjusting parameter values affects the model. The app is available at https://sr-martin.shinyapps.io/model_app/.

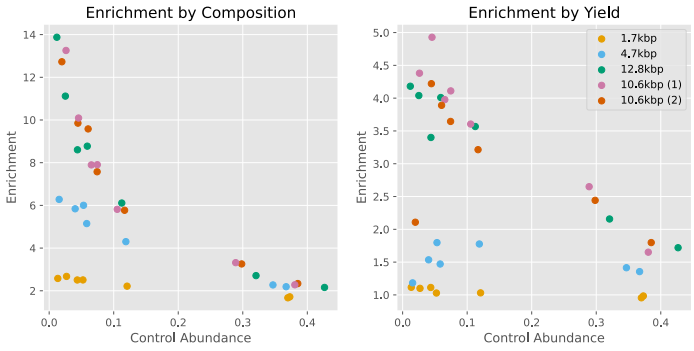

Fig.1: Enrichment values observed in terms of composition and yield. Key values indicate the average read length for each experiment.

We can measure enrichment as the relative change in either yield or composition against a control run. Control runs were performed on the same flowcell during the experiments by having half of the pores perform adaptive sample, and half perform normal sequencing.

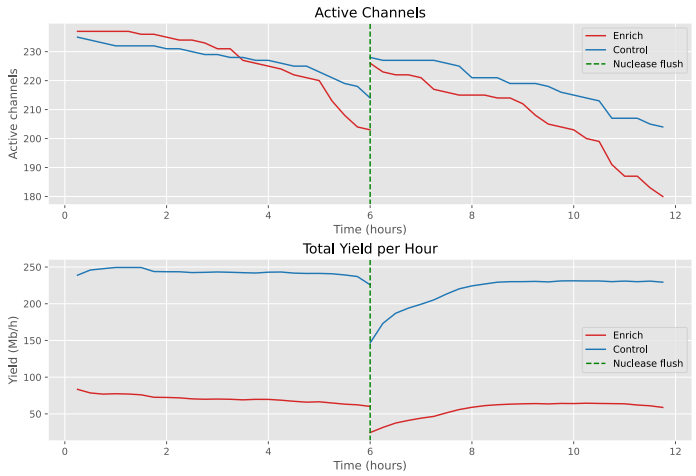

Fig.3: The effect of adaptive sampling on the number of active channels and total yield over time.

Key Findings:

We found that enrichment is highest for low abundance species and when average molecule lengths are higher. When molecule lengths are around 2kbp, there is little improvement in yield by performing adaptive sampling, but for long molecule lengths we found improvements of up to 5x against a control run (Fig.1). Enrichment by composition correlated well with the predictions of our model (Fig.2). Enrichment by yield is less than that predicted by enrichment by composition; this may be due to reduced pore efficiency that is associated with repeated ejection of molecules. We performed further experiments to determine the effect that adaptive sampling has on the number of active channels and total yield, and included a nuclease flush of the flowcell after 6 hours (Fig.3).

Paper: S Martin et al. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biology* **23**, 11 (2022), https://doi.org/10.1186/s13059-021-02582-x

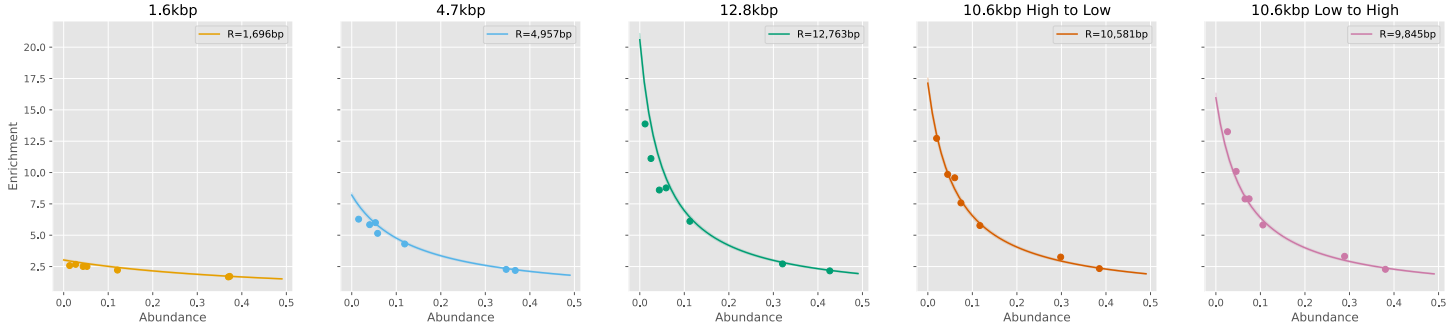

Fig. 2: Observed enrichment plotted with predicted enrichment curve for each experiment. The variables $C = 0.5s, D = 1.0s$, and $S = 420$bps were kept constant. We found high correlation between our observations and the model's predictions, with an overall Pearson's r value of 0.9825

Samuel Martin
Postdoctoral Scientist
samuel.martin@earlham.ac.uk

Earlham Institute, Norwich Research Park, Norwich, Norfolk, NR4 7UZ, UK
www.earlham.ac.uk