Taxon-specific training of Nanopore basecalling models: *Mycobacterium tuberculosis*

This manuscript (permalink) was automatically generated from mbhall88/tubby-manuscript@c959710 on November 25, 2021.

Authors

- **Michael B. Hall**
  0000-0003-3683-6208 · mbhall88 · mbhall88
  European Bioinformatics Institute, Hinxton, Cambridge, CB10 1SD, UK

- **Zamin Iqbal**
  0000-0001-8466-7547 · iqbal-lab · ZaminIqbal
  European Bioinformatics Institute, Hinxton, Cambridge, CB10 1SD, UK · Funded by Grant XXXXXXXX

Abstract

Nanopore-based single molecule sequencing has changed the landscape of sequence analysis, with its combination of long reads, portable hardware and real-time results. As DNA strands pass through a nanopore, an electrical signal is measured, whose form depends on local DNA sequence, molecular damage,and methylation. This can be viewed as an opportunity to measure methylation and damage, but is also a challenge for accurate basecalling of the DNA strand. The highest accuracy basecallers are now based on neural networks, trained on a range of genomes. Wick *et al.* recently showed that they could achieve higher basecalling accuracy in *Klebsiella pneumoniae* if they trained the neural net on that species and associated family. This raises a fundamental challenge for data sharing. If every research team trains their own basecaller for their species, it will become very difficult to co-analyse sequence from different studies, as they will all have very different error-profiles. We therefore are seeking to crowd-source a survey to investigate which taxa benefit from such bespoke basecalling models. Importantly, these models would be made available to others to ensure consistency in studies relating to the relevant taxa. We begin this effort by providing methods, results, and a custom model from training on *Mycobacterium tuberculosis* data.

Keywords

Author notes

Abbreviations

Impact statement

Data summary

Introduction

Methods

Results

Discussion

Funding information

Acknowledgements

Author contributions

Conflicts of interest

Ethical statement

References