Supplementary Materials to ‘Axe: rapid, competitive sequence read demultiplexing using a trie’

Kevin D. Murray$^1$ and Justin O. Borevitz$^1$

$^1$ARC Centre of Excellence in Plant Energy Biology, Research School of Biology, ANU, Canberra, Australia

Supplementary Methods

Validation experiments

To quantify the performance of axe relative to similar tools, 10 million 100bp paired end reads were simulated from a random 1Mbp genome using Mason2 (Holtgrewe, 2010). Sets of index sequences of various sizes (see results) were drawn from the set of all 8-mers with a minimum hamming distance of 3. Sample frequencies were drawn from a gamma distribution with a shape parameter of 2; read pairs are randomly assigned a sample from these sample frequencies. Index sequences are inserted into the 5' end of sequences and errors added with a frequency of $10^{-2.5}$ (PHRED quality of 25). Combinatorial index sets were generated using the same process for each read.

These datasets were used to benchmark all operational modes of axe, alongside previous read demultiplexing software flexbar, fastx and AdapterRemoval. The precise versions and parameters for these programs, and the workflow which performs the simulations reported here, are available at https://github.com/kdmurray91/axe-experiments.

References

Holtgrewe, M.