Supplement

Supplementary methods

Comparison of different demultiplexing/filtering options

We compared four alternative filtering strategies across all five libraries:

- Guppy only: all raw reads as demultiplexed by guppy
    - Guppy + Deepbinner [1]: raw reads as demultiplexed by guppy with unclassified reads reassigned where possible by Deepbinner
    - Guppy + Deepbinner + Filtlong: as for option two but additionally processed with Filtlong using the following filters:
        - 90% of reads (based on highest quality score as judged by kmer match to Illumina reads) or maximum of 250Mbp (~ 50X depth), whichever resulted in fewer reads.
        - Only reads > 1000bp
        - –trim and –split 100
        - min_mean_q 25
    - Guppy + Deepbinner + random subsampling: as for option two but Rasusa[2] was used to subsample reads to ~ 30X coverage.

Using Guppy and Guppy+Deepbinner to reclaim unclassified reads provided the most complete assemblies (i.e. chromosome and all contigs circularised 36/57 and 37/57 respectively). Adding quality and length based filtering to Guppy + Deepbinner resulted in a slightly worse performance (33/57 assemblies circularised). Random sub-sampling to 150Mb with Rasusa provided the least complete assemblies (30/57) although in some cases the sampling threshold was more than the total number of sequenced bases for the sample. Using only the Guppy and Guppy+Deepbinner read filtering strategies we additionally compared assemblies created with Unicycler’s --mode set to ‘normal’ and ‘bold’. As expected, there were more complete assemblies using bold (36/57 (Guppy only) 37/57 (Guppy + Deepbinner)) compared with normal modes (29/57 (Guppy only) 31/57 Guppy + Deepbinner).

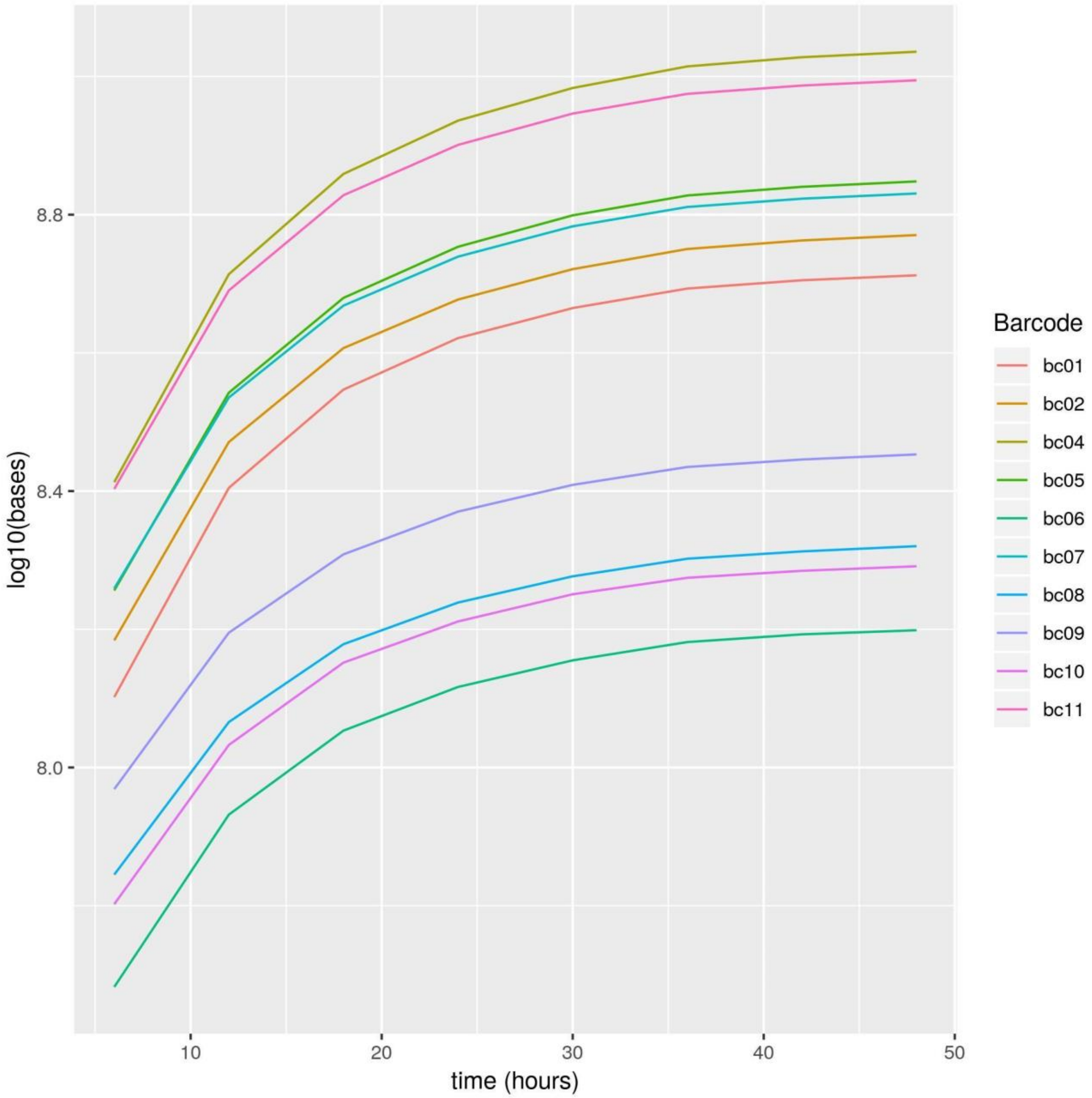


***Figure S1**: Log base output per barcode over time for library 1.*

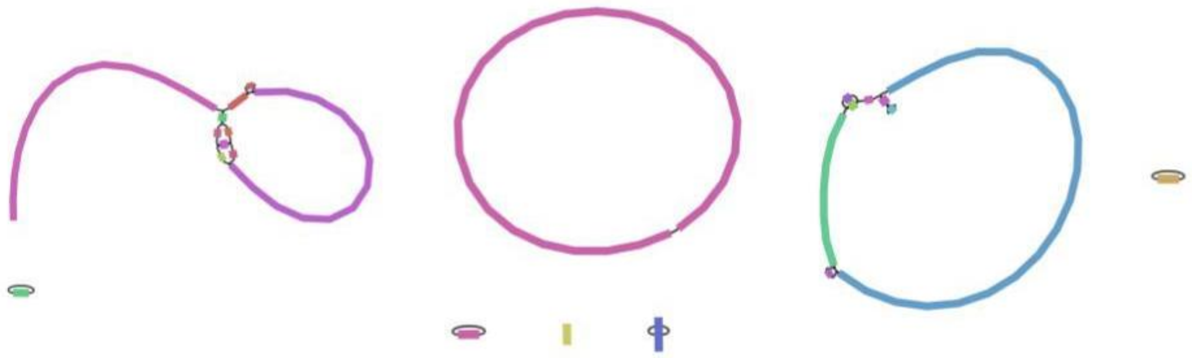
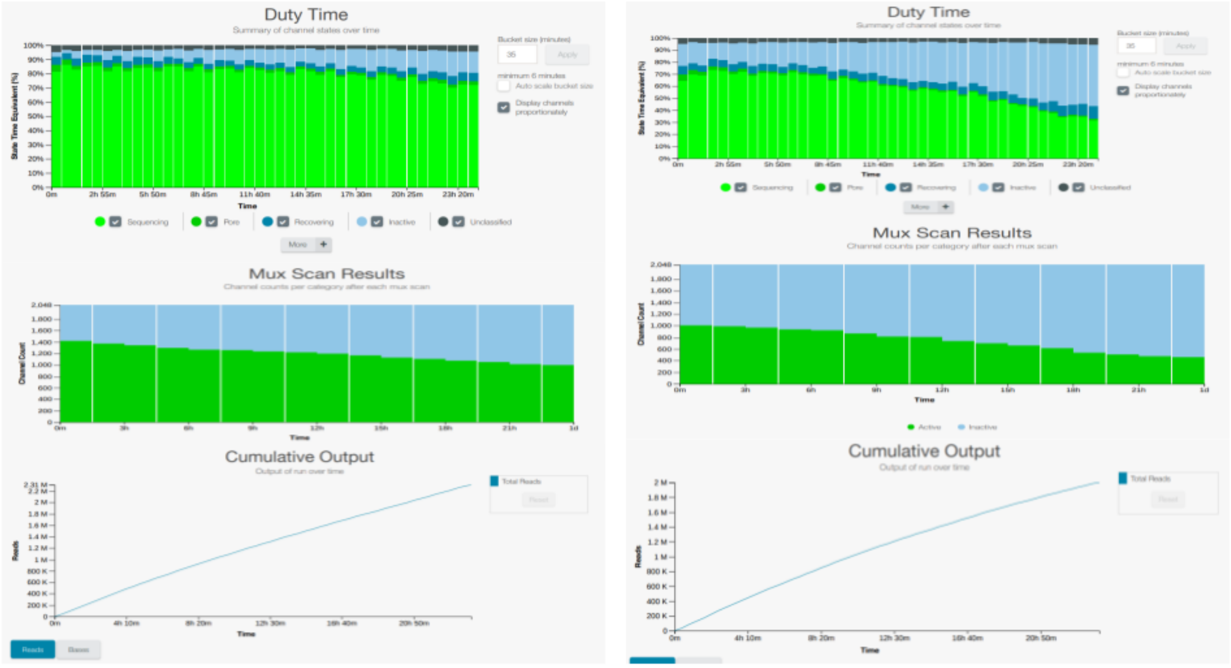


Figure S2*: Graphs of the three incomplete assemblies from library 1 (left to right: barcodes bc02, bc04, bc05, isolates blc-23, blc-24, blc25) at 24 hours.*



Bucket size (minutes) 35 Apply

minimum 6 minutes

□ Auto scale bucket size

☑ Display channels proportionately

***Figure S3**: screenshots from the MinKnow reports from library 3 (left) and library 4 (right). These two libraries were run on the same flow cell with the ONT wash kit used following the end of library 3 sequencing at 24 hours before loading library 4.*

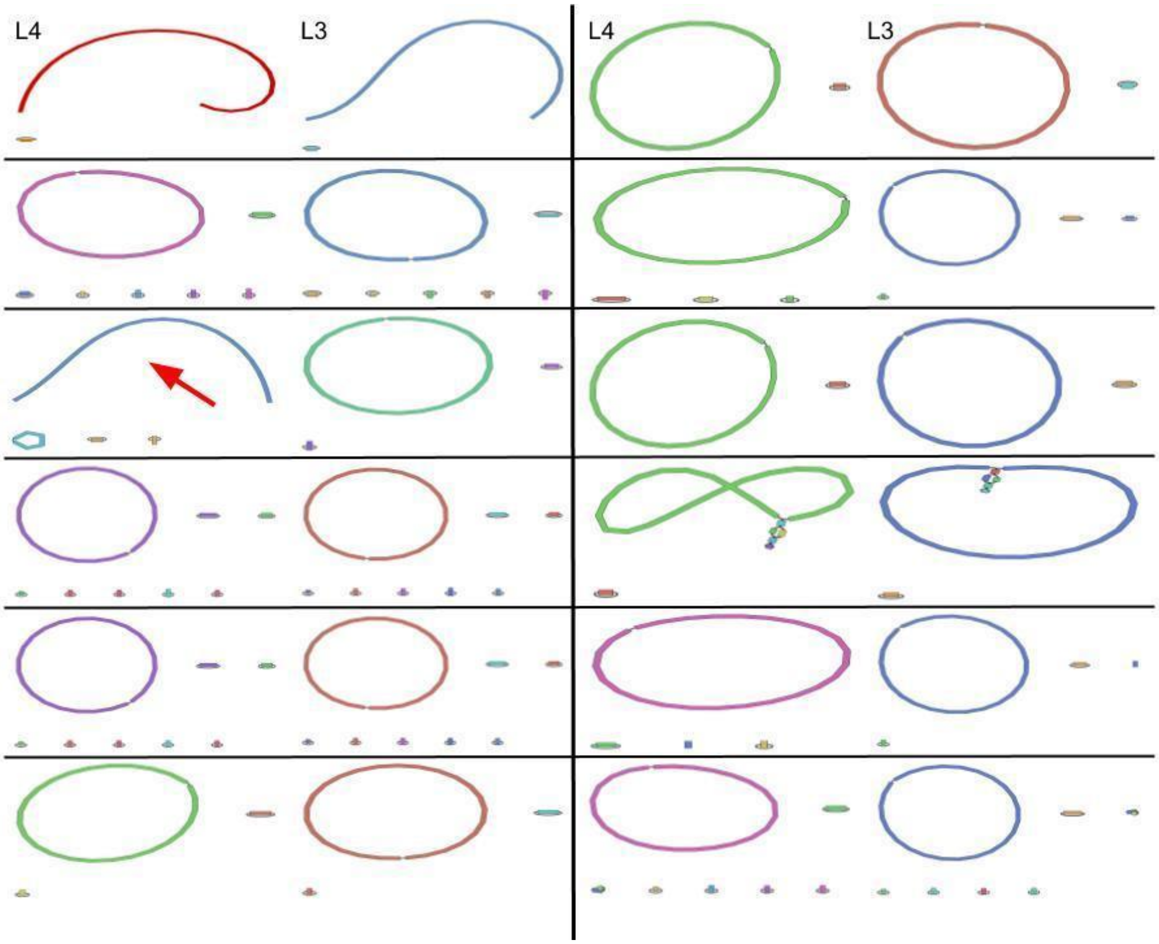


Figure S4: *Each box represents the graph of a single isolate sequenced with different barcodes in libraries 3 (L3) and 4 (L4). The left panel contains isolates blc-44-49 and the right L4 isolates blc-50-55 (see table 1). Red arrow denotes the assembly with the major structural difference (blc-46). NB. Bandage randomly colours and bends contigs.*

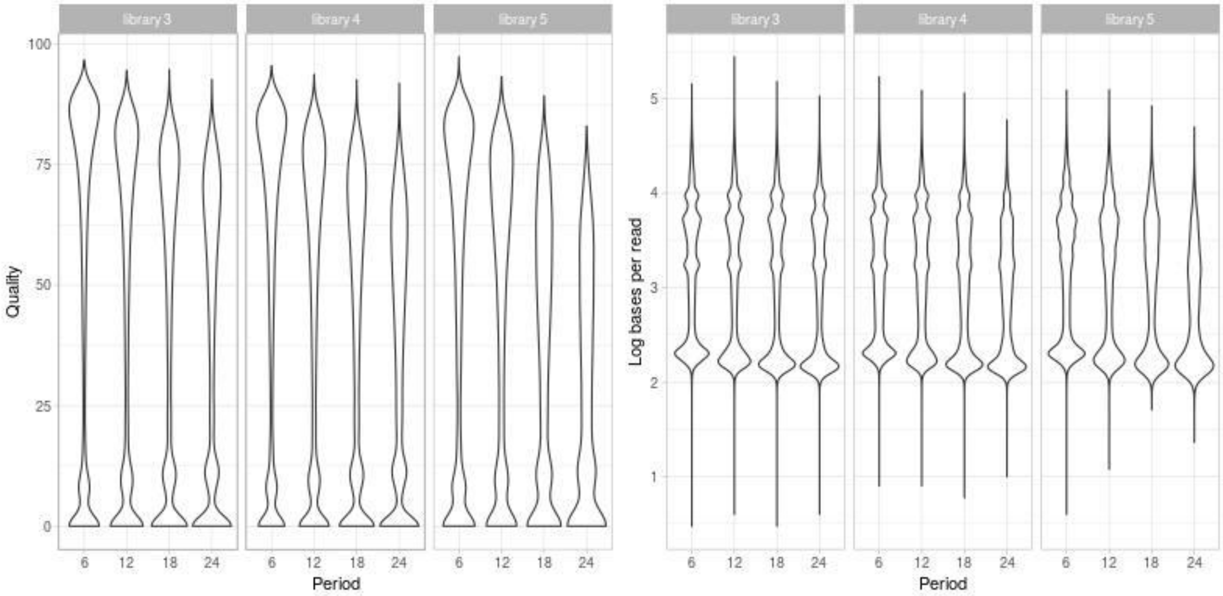


***Figure S5**: Read length/quality violin plots for libraries 3, 4 and 5 which were all sequenced consecutively on the same flowcell. Period 6 is time >= 0 hours < 6 hours, period 12 is time >=6 hours <12 hours etc.*

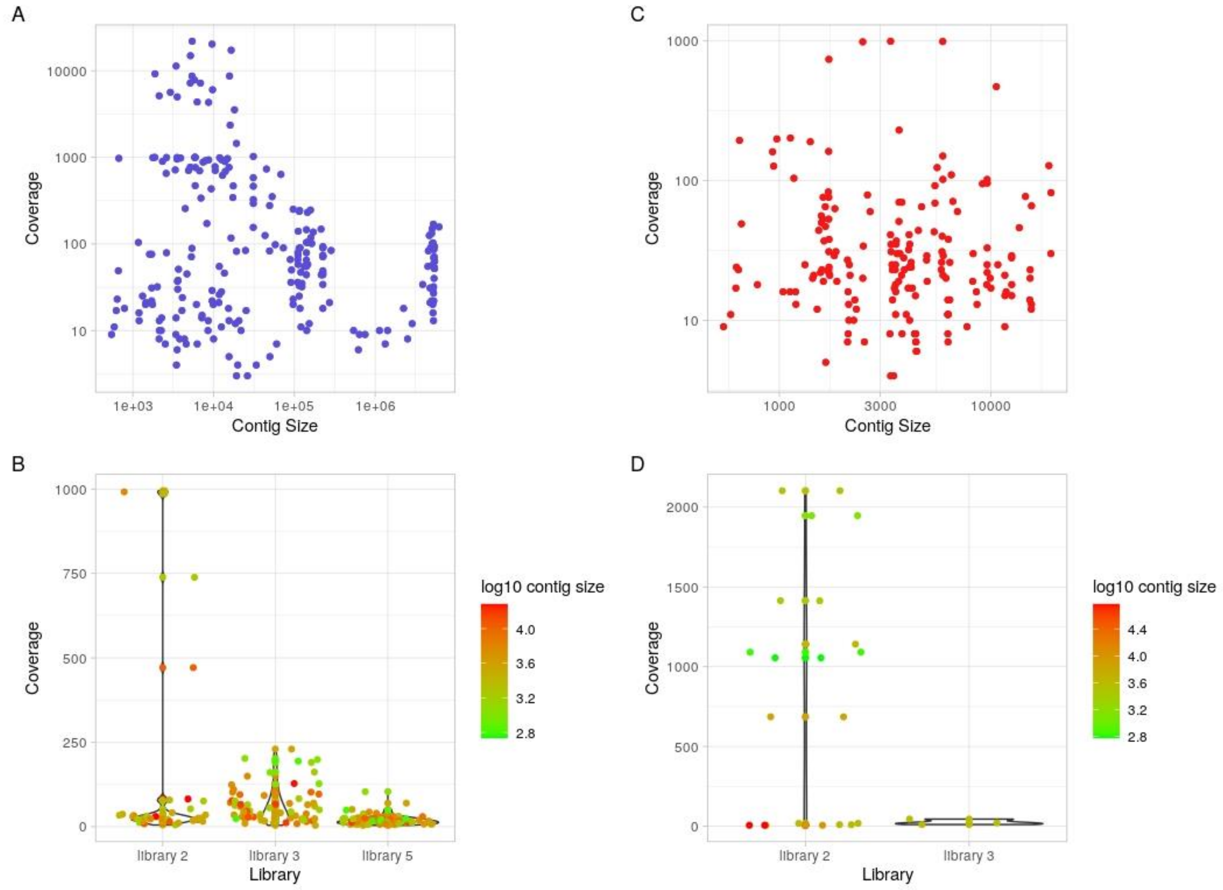


Figure S6*: Size and coverage of contigs from Flye long-read only assemblies which A) matched to a contig in the hybrid assembly (match defined as an alignment of at least 100 kbp or ¼ of the replicon using the analysis.py script from [3]) or C) which did not map to the hybrid assembly. In general, ‘true’ small contigs were present at high coverage compared to the spurious unmapped small contigs. B - Violin plot of the coverage of the unmapped contigs for each of the libraries (library 1 = 0, library 2 =36, library 3 = 67, library 5 = 78) using reads as demultiplexed by Guppy alone (there were no unmapped contigs for library 1). Log10 contig size is indicated by the colour gradient as shown. D - Violin plot of coverage of unmapped contigs using reads demultiplexed by taking the consensus of Deepbinner and Guppy. Flye assemblies from 1 and 5*

produced no spurious contigs using this method and there were also substantially fewer for libraries 2 and 3 (n=22 and 3 respectively).

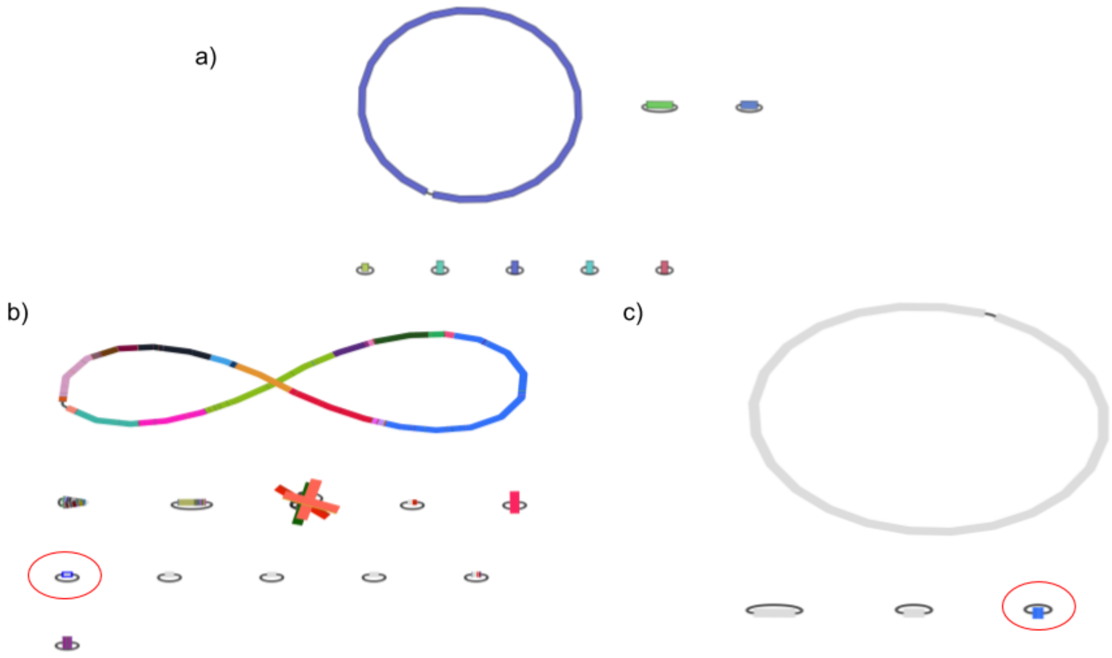


***Figure S7**: a) Hybrid assembly of blc-48 (barcode 02, library 3) showing the likely ground truth of a chromosome and 7 plasmids. b) Flye assembly of the same isolate. Colours represent blast hits from the short-read only assembly (i.e. replicons coloured grey like the one highlighted with a red ring are not present in the short read assembly and are presumed to be spurious). c) The contig highlighted with the red ring in b) was blasted against all other hybrid assemblies created from the same library (3), revealing a close match to a plasmid from blc-51 (barcode 05) and demonstrating likely cross contamination between barcodes in the same library. Note this does not represent between library contamination because library 3 was run on a new flow cell.*

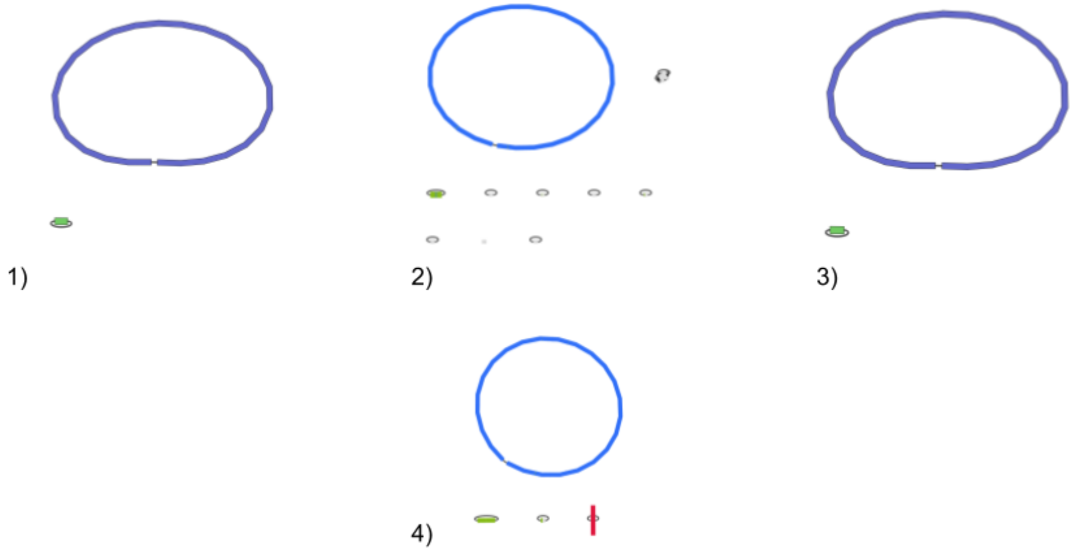


***Figure S8**: 1) Hybrid assembly of blc-50 (barcode04, library 3) representing the likely ground truth (1 chromosome, 1 plasmid). 2) Flye assembly of the same isolate - colours show blast hits using the against the hybrid assembly demonstrating the presence of the true chromosome and plasmid but also seven other likely spurious replicons. 3) Flye assembly using only reads assigned to barcode 04 by both Guppy and Deepbinner (1 chromosome, 1 plasmid). In general such a consensus demultiplexing strategy greatly improved long-read only assemblies. 4) blc-49 (barcode 03, library 3) using consensus demultiplexing of Guppy and Deepbinner - colours represent blast hits against the short-read only assembly. There is a spurious additional part of which is green (ie truly belongs to the plasmid on the left) and part is grey (ie is not represented in the hybrid assembly). This demonstrates that whilst consensus demultiplexing is useful to improve the accuracy of multiplexed long read only assembly, it is likely still prone to within-library contamination.*

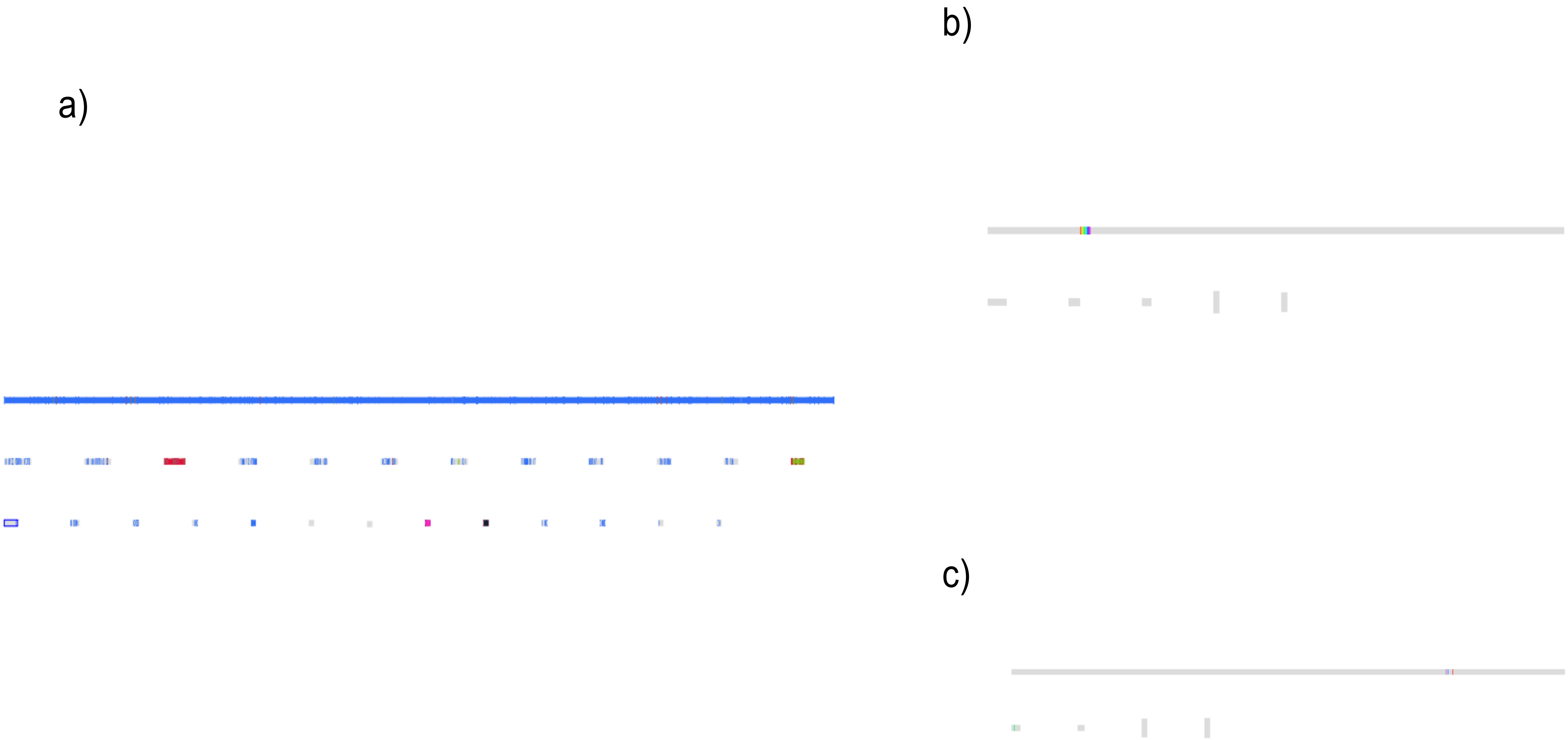

a)
b)
c)

***Figure S9**: The problem of within library contamination affecting long-read only assemblies is not unique to the rapid barcoding kit nor to this study. The hybrid assembly of RBHSTW-0000059 from the REHAB[4] dataset (c) was blasted against its long-read assembly (a). Coloured regions of contigs in a) show sections of the long read assembly also found in the hybrid assembly, whereas grey contigs are not found in the hybrid assembly and therefore likely represent contamination. b) shows a blast search using one of the contaminant contigs from the long-read assembly of RBHSTW-0000059 (bottom row, far left, highlighted with blue outline) against another isolate (MGH78578) sequenced on the same flowcell in the same run. A near perfect match on the chromosome is demonstrated, indicating the likely true origin of this contig whereas there is no convincing match against the hybrid assembly of RBHSTW-0000059 shown in c), confirming the likely within library contamination.*

Supplementary table S1 is provided in separate spreadsheet file to aid viewing

***Table S1**: Summary of isolates sequenced in each library. * reference strain MGH78578 ** Sequencing data for library 3 is available on figshare (https://doi.org/10.6084/m9.figshare.11816532) and was not uploaded to NCBI to avoid duplication (the isolates are the same as those in library*

Library	Human reads	Total Reads	% human reads
1	553	1648038	0.02
2	147	818091	0.02
3	357	1586627	0.02
4	241	1328083	0.02
5	62	331211	0.02

Table S2*: Reads binned as human by centrifuge from each library. Library 2 was run on a flow cell first used to sequence a human pathology specimen for 24 hours which had 2028024/2059966 (98.4%) initial reads binned as human. After washing the reads classified as human appeared within the range of all other libraries (which were sequenced on flow cells on which no human DNA had been loaded). This demonstrates the highly effective removal of DNA by the ONT wash kit.*

References

1. **Wick RR, Judd LM, Holt KE**. Deepbinner: Demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. *PLoS Comput Biol* 2018;14:e1006583.

2. **Hall M**. *Rasusa*. Github. https://github.com/mbhall88/rasusa (accessed 20 February 2020).

3. **Wick RR, Holt KE**. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Res* 2019;8:2138.

4. **De Maio N, Shaw LP, Hubbard A, George S, Sanderson ND, *et al.*** Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microb Genom*;5. Epub ahead of print September 2019. DOI: 10.1099/mgen.0.000294.