A simple library preparation modification significantly reduces barcode crosstalk in ONT multiplexed sequencing

Qing Dai[1], Claudia K. Gunsch[1], and Joshua A. Granek[2,*]

[1]Civil and Environmental Engineering, Pratt School of Engineering, Duke University, Durham NC, 27708, USA
[2]Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham NC, 27708, USA
[*]joshua.granek@duke.edu

ABSTRACT

Sample multiplexing is a standard component of high-throughput sequencing, but it introduces a potential point of error. We identified a problem with sample multiplexing in Oxford Nanopore standard ligation-based barcoding protocol that results in a relatively high rate of read misassignment due to barcode crosstalk. This barcode crosstalk is particularly pernicious because it masquerades as sample cross-contamination. We have quantified the rate of read misassignment due to crosstalk. To address this problem we devised Post-Ligation Pooling, a simple protocol modification that reduces misassignment due to crosstalk by more than an order of magnitude. We have also compared Post-Ligation Pooling to the Short Fragment Buffer wash, an optional step recently introduced by Oxford Nanopore, and to a library preparation that combines Post-Ligation Pooling with the Short Fragment Buffer wash, which produces the lowest crosstalk. Our results highlight the need to recognize the potential for barcode crosstalk and use our modified protocol to mitigate it in when working in metagenomics, doing clinical assays, using low-DNA samples, and any other application where cross-contamination can produce misleading results.

Introduction

Index barcodes are an important tool for high-throughput sequencing. They enable sample multiplexing, allowing multiple samples to be pooled and sequenced simultaneously. This allows large-scale studies by not only lowering per-sample costs, but also by providing a mechanism for managing batch effects. However, index barcodes add a potential point for error: misassignment, when reads are attributed to the wrong sample. There are two major causes of misassignment: demultiplexing error and barcode crosstalk (also known as barcode hopping or index misassignment).

Demultiplexing error occurs during analysis of sequence data. In demultiplexing, each sequence read is assigned to a sample according to its barcode sequence. A demultiplexing error occurs when a read is assigned to the wrong sample, or more commonly, not assigned to any sample at all. The most common cause of demultiplexing error is error in the barcode sequence which results in a mismatch with the expected barcode sequence. Barcode crosstalk occurs during the library preparation process when the incorrect barcode is attached to nucleic acid fragment (e.g. the barcode for sample A is attached to a DNA fragment from sample B).

There has been a great deal of work done to reduce demultiplexing error: designing error correcting barcode sequences, improving basecalling and demultiplexing algorithms and post hoc correction. This includes several demultiplexing algorithms designed specifically for Oxford Nanopore Technologies (ONT) long-read sequences: the signal-level demultiplexer Deepbinner achieves a mean precision of ~98.5% and recall of ~92%, with about 0.3% of reads consistently misclassified and 7–8% left unclassified[1] and MysteryMaster[2] further reduce false-assignment rates to below 0.5%.

Because most demultiplexing error results in reads simply not being assigned to any sample, its primary effect is a reduction of usefuly data. Barcode crosstalk is much more pernicious because reads from one sample are incorrectly assigned to another. The barcode crosstalk therefore has an effect that is equivalent to cross-contamination of samples. Mitigation of crosstalk is critical for both accuracy and reproducibility because even minimal crosstalk can cause false positives in diagnostic metagenomics, inflate diversity estimates, or distort quantitative analyses[3,4]. The magnitude of barcode crosstalk varies among sequencing platforms and library preparation chemistries. On Illumina patterned flow-cell instruments (e.g., HiSeq 4000, NovaSeq), index hopping rates typically range from 0.1–6%, depending on library type and indexing strategy[5,6]. Mitigation of barcode crosstalk requires experimental solutions - it is not amenable to computational solutions. Barcode crosstalk cannot be detected or quantified by algorithmic advances alone. This also requires carefully designed experiments. One example of crosstalk mitigation is the implementation of unique dual index barcodes, which reduce the rate of misassignment of reads

generated on Illumina patterned flow-cell from 0.1–6% to below 0.01%[6] (it should be noted that unique dual indices do not reduce crosstalk, but they do allow detection of reads with crosstalk).

Despite computational advances in demultiplexing and error correction, the experimental origins of barcode crosstalk have received comparatively little attention. While physical and biochemical contributors—such as adapter carryover, bead cleanup efficiency, and library pooling strategies—are less systematically studied. In this work, we quantify misassignment due to crosstalk in the ONT library preparation protocol and demonstrate that simple modifications can substantially reduce crosstalk in multiplexed ONT sequencing. By recognizing the potential for this problem and using a modified protocol, researchers can improve the reliability of their multi-sample long-read sequencing.

Results

Real-sample metagenomics validates protocol improvements

In our research on microbiomes of the built-environment we use ONT sequencing to generate shotgun metagenomic data. We observed what first appeared to be cross-contamination: negative controls (molecular biology-grade "water blank" and no-input "empty blank") seemed to be composed of a mix of a true sample (p-trap water sample) and positive controls (DCS *lambda* and *phiX*), and even the positive controls seemed to contain DNA from the true sample (Figure 1A).

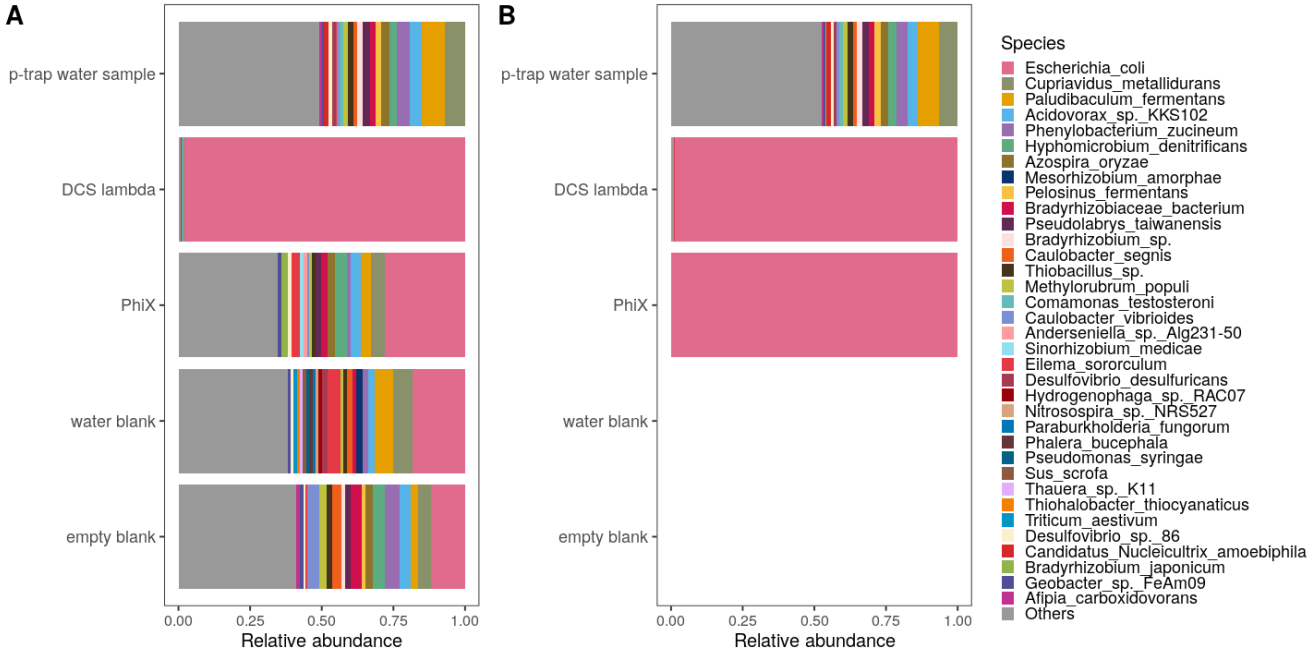

Figure 1. Taxonomic classification of libraries prepared with (A) standard protocol shows apparent cross-contamination between samples, while (B) libraries prepared with the modified PLP protocol are free from the apparent cross-contamination. Note that the positive controls (DCS *lambda* and *PhiX*) are classified here as *Escherichia coli* because they are both bacteriophages that are found in the *E. coli* genome.

Cross-contamination is a well known problem in microbiome research[3], but careful examination of our experimental procedures eliminated common sources such as "kitome" and "splashome". We ultimately explored barcode-crosstalk as a possible source of the apparent cross-contamination. In the standard protocol, samples are pooled immediately after index barcoding, before sequencing adapters ligation. We hypothesized that barcode crosstalk might occur if excess index barcode is ligated onto sample DNA during the adapter ligation step. We tested ligating adapter before pooling samples to see whether this would reduce the opportunity for this barcode crosstalk. A modified post-ligation pooling (PLP) protocol—where barcoded libraries are only pooled after final adapter ligation— substantially reduced the apparent cross-contamination (Figure 1B), confirming barcode crosstalk as the source of the apparent cross-contamination. The PLP strategy's near elimination of barcode crosstalk is not only clear from the elimination of p-trap and DCS *lambda* signals from the "blank" samples, but by looking at the absolute number of sequence reads in this blanks (where no reads are expected). PLP reduces reads in the water by one to two orders of magnitude: 1535 reads were demultiplexed to the empty blank in multiplexed libraries prepared following the Standard protocol, while the PLP prepared empty blank produced only 7 reads (Table S1). For water blank, the Standard

protocol produced 1139 reads, compared with only 90 reads with the PLP protocol. *PhiX* serves as another partial negative control (inadvertently): *PhiX* genomic DNA sequenced poorly, so most of the *phiX* data from the library prepared with the standard protocol resembles the standard water blank. The PLP library preparation appears to eliminates this contamination, as with the the blanks. We hypothesize that the poor sequencing yield of *phiX* arises from genome structure: because the *phiX* genome is circular, only the small fraction of genomic DNA with double-strand breaks was a substrate for barcode and adapter ligation during library prep.

We also used the DCS *lambda* and *PhiX174* genomes provided by ONT to build a custom reference and directly mapped the samples with Minimap2. According to the results summarized in Table S2, PLP consistently improved target specificity compared with the Standard protocol. For the DCS *lambda* sample, PLP yielded 2660 DCS *lambda*–mapped reads out of 2836 total (about 94% on-target), whereas the Standard protocol produced 6918 DCS *lambda* reads out of 8274 total (about 84% on-target). For the *PhiX* sample, PLP generated 94 *phiX*-mapped reads among 187 total (about 50% on-target) and no DCS *lambda* reads, while the Standard protocol produced 548 *phiX* reads and 48 misassigned DCS *lambda* reads out of 2719 total (about 20% on-target). Together, these results show that PLP produces cleaner positive-control libraries with a higher proportion of reads mapping to the intended spike-in genome and markedly reduced cross-assignment between DCS *lambda* and *phiX*.

Evaluation of barcode crosstalk using ATCC reference genomes across library preparation protocols

Our results from the environmental and control samples demonstrate that the PLP library preparation substantially reduces barcode crosstalk in a real-world metagenomic context. A limitation with environmental data is that we do not know the true absence or presence of every organism a priori, so we cannot quantify misassignment rates as precisely as with a defined mock community. We designed a minimal experiment to accurately quantify crosstalk resulting from the standard protocol and how effectively the PLP modification reduces it. The minimal experiment consisted of four distinct DNA samples: purified DNA from three different bacterial isolates, *Mycoplasma hominis*, *Phocaeicola vulgatus*, *Corynebacterium striatum*, and the DCS *lambda* (ONT's standard positive control) used above. In addition to comparing the Standard library preparation with our PLP preparation, we also included a revised protocol recently released by ONT which uses a Short Fragment Buffer cleanup (SFB), and a novel protocol that combines the PLP strategy with ONT's SFB revised protocol (SFB+PLP).

	Standard vs PLP		SFB vs PLP		SFB vs PLP+SFB	
	Standard	PLP	SFB	PLP	SFB	PLP+SFB
3ng Flongle	ML	PC	MP	CL	LC	PM
20ng MinION	MP	CL	CP	LM	ML	PC

Table 1. Summary of pairwise comparisons for different input amounts and protocols. The pairwised protocol samples are pooled together in adapter cleaning step and sequenced on the same flow cell. The single-letter codes indicate the species of input gDNA: M = Mycoplasma hominis, P = Phocaeicola vulgatus, C = Corynebacterium striatum, and L = DCS *lambda*.

To eliminate the possibility that the performance of a library preparation method was confounded by DNA type or run-specific effects, we designed the comparison using a combinatorial, Latin-square–inspired design stratified by input amount: for each protocol comparison (Standard vs PLP, SFB vs PLP, SFB vs PLP+SFB), the four genomes were split into two complementary two-species mixtures, with different pairings used for the 3 ng Flongle and 20 ng MinION runs; within each run, the paired mixtures were prepared with the two protocols being compared, pooled at the adapter-cleaning step, and sequenced on the same flow cell 1. By sequencing both library types on the same flow cell, we confirmed that PLP's improved performance was not an artifact of run-to-run variability – the reduction in cross-sample contamination was directly attributable to the library preparation method rather than differences in flow cell output.

It is worth noting that the total sequencing yield differed between the Standard and PLP runs (with the Standard flow cell producing more reads overall). Differences in output can influence the absolute number of contaminant reads, which is why the minimal experiment was designed to pool both protocols on a single flow cell for a balanced comparison. The Standard method (red) exhibits substantially higher barcode crosstalk/misassignment in every sample, whereas Short Fragment Buffer cleanup (SFB, green) and the post-ligation pooling strategy (PLP, blue) each reduce misassigned reads by an order of magnitude (Figure 2. Combining SFB with PLP (purple) virtually eliminates cross-assignment, with blank and control samples showing near-zero spurious reads and Tables S6 and S5). The contrast is clear: multiplexing with the PLP protocol greatly mitigated index misassignment and carry-over contamination relative to the Standard library prep. This improvement in did not come at the cost of detecting true community members, and it suggests that PLP can improve the fidelity of microbiome analyses, and other applications that are sensitive to cross-contamination, by ensuring reads stay assigned to their correct sample of origin.

Across all experiments, barcode crosstalk was markedly lower with the new library prep strategies, especially PLP method (See Table 2). Under the original Standard protocol, we detected a significant fraction of reads with barcode crosstalk – on

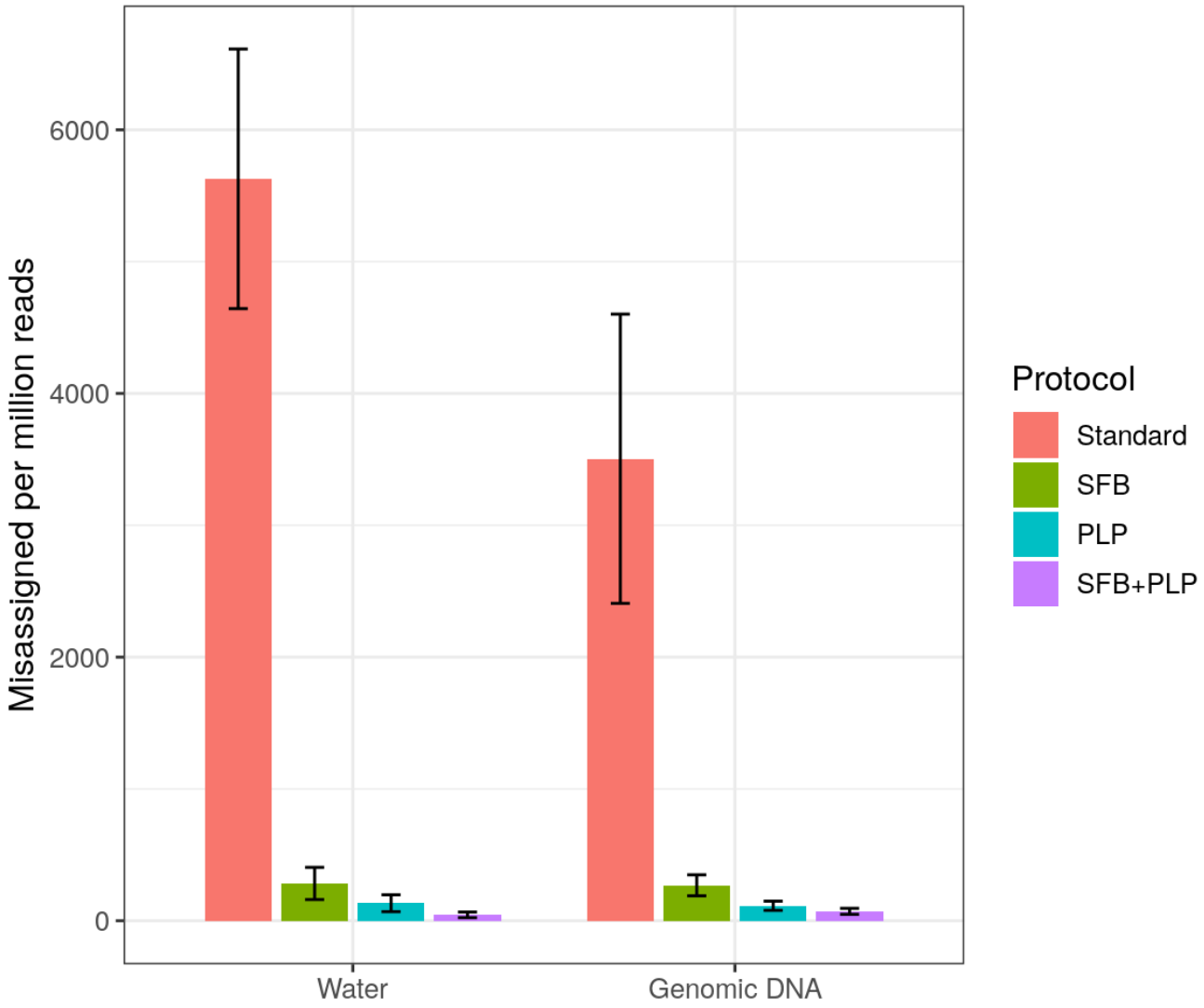

Figure 2. Misassigned reads per million mapped reads across different treatment.

the order of thousands reads per million mapped (0.5–0.6% in worst cases, such as in blanks and control DNA samples). In contrast, the PLP method reduced barcode crosstalk by one to two orders of magnitude. Even when used alone, PLP improved the demultiplexing precision to about 99.96%, compared to ~99% with the Standard approach (averaged across all barcoded samples). The Short Fragment Buffer (SFB) cleanup,ONT's recently introduced post-barcoding wash, also substantially improved performance over the Standard protocol, improved performance over the Standard protocol, reducing the barcode crosstalk rate from 0.362% to 0.064%. Notably, combining the SFB and PLP protocols (SFB + PLP) achieved the best results: barcode crosstalk was nearly absent, with many samples and batches showing zero or only a few dozen barcode crosstalk reads per million (precision >99.98%). This corresponds to F1 scores 0.99982 for SFB+PLP, compared to 0.988 for the Standard method. Across methods, precision was already high for the Standard protocol (99.0%), so the absolute gains in precision with SFB, PLP, and SFB+PLP were relatively modest. In contrast, recall improved markedly from 98.6% with the Standard protocol to ≥99.9% with all three modified protocols, accompanied by a >20-fold reduction in the fraction of misassigned reads (from 0.362% to 0.017%). This pattern indicates that the main benefit of the modified workflows is preventing reads from "leaking" out of their correct barcode and being assigned to other samples, rather than dramatically changing the purity of reads entering each barcode. In summary, PLP dramatically improved barcode assignment fidelity, especially in combination with SFB, yielding demultiplexing performance comparable to the best indexing practices.

Method	Mean Recall (%)	Mean Precision (%)	Mean F1 (%)	Misassigned reads (%)	number of samples
Standard	98.588	99.024	98.805	0.362	4
SFB	99.865	99.950	99.908	0.064	8
PLP	99.979	99.959	99.969	0.023	8
SFB+PLP	99.981	99.983	99.982	0.017	4

Table 2. Performance of different library prep treatments, only assigned mapped reads included

Discussion

Accurate barcode assignment is critical in multiplexed sequencing studies, particularly in metagenomics, clinical diagnostics, and other multi-sample applications. Even minor barcode misassignments can produce serious false positives, for example, reporting a pathogen in the wrong sample[7]. Our optimized SFB+PLP workflow virtually eliminated such artifacts. In blanks, no spurious reads were observed with SFB+PLP, whereas the Standard workflow often produced contaminant sequences, such as unrelated microbes, in blanks. Our findings align with reports from other studies, which have found barcode crosstalk rates on the order of a few percent for conventional ONT library kits(e.g. ~2–3% in ligation or tagmentation workflows)[8]. Together, our results highlight a substantial improvement in sequencing integrity.

Mechanistically, these improvements are consistent with a simple intuition: barcode crosstalk is most likely when barcoded fragments from different samples, together with any residual free barcode oligonucleotides, share a common ligation environment. In the standard ONT native barcoding workflow (Figure 3 left side), individual samples are pooled immediately after the barcode ligation step, before final adapter attachment. This creates an opportunity for barcode "leakage" or carry-over between libraries. Any residual, unligated barcodes from one sample can associate with DNA from another sample during the pooled adapter ligation. ONT's own documentation notes that using Short Fragment Buffer rather than ethanol for the post-barcoding clean-up removes excess barcode oligonucleotides more efficiently and reduces physical barcode crosstalk, implying that residual free barcodes contribute to misassignment events[9]. Such events are analogous to tag jumping reported in on the Illumina platform[3]. ONT's revised SFB protocol addresses this risk from a chemical angle by improving the removal of excess barcode oligonucleotides before pooling. Our PLP workflow (Figure 3 right side) implements a complementary design change: PLP workflow keeps samples physically isolated through all ligation steps and pools fully barcoded, adapter-ligated libraries afterwards. This re-ordered pooling strategy removes the shared ligation environment altogether and therefore eliminates the main opportunity for barcode leakage. Importantly, this re-ordered pooling strategy requires only minimal deviations from the standard native barcoding protocol and can be implemented on existing kits without purchasing additional reagents such as the SFB expansion. In practice, laboratories already familiar with the ONT native barcoding workflow can adopt PLP simply by changing the position of the pooling step and adjusting volumes accordingly, making the approach straightforward to roll out at scale. The consequences were evident: the Standard protocol's pooled ligation yielded measurable cross-contamination, with the water blank samples showing up to 5,000–6,000 spurious reads per million. Consistent with this, our SFB-inclusive library preparation showed far fewer misassigned reads. In summary, PLP addresses the root cause of barcode misassignment (premature pooling), and SFB further minimizes carryover of index residues, both contributing to the substantial reduction in apparent cross-sample contamination.

Our findings corroborate and extend earlier reports of index misassignment in both short-read and long-read sequencing.

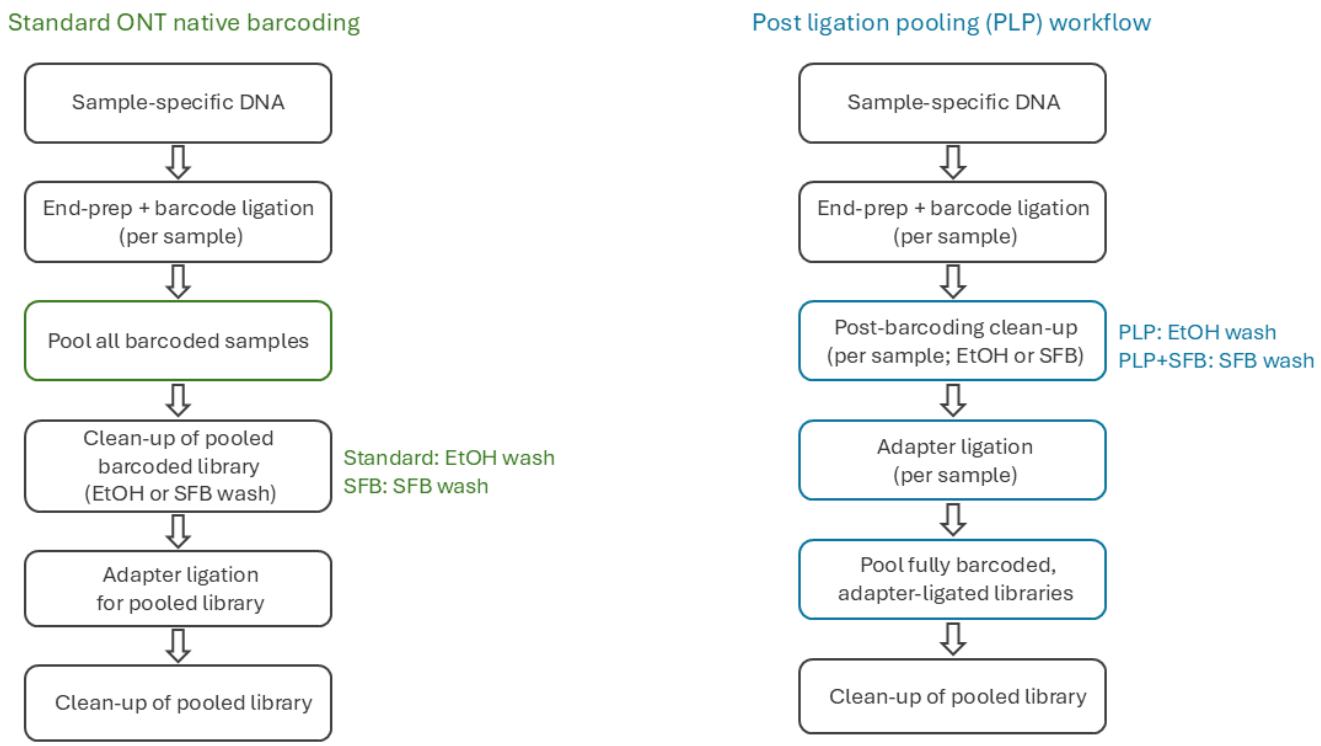

Figure 3. Comparison of pooling strategy in ONT native barcoding workflows

Barcode crosstalk is not unique to ONT; for example, in Illumina platforms even a well-designed indexing scheme can suffer a small percentage of reads swapping between samples. Reported index hopping rates on Illumina vary from as low as ~0.1% up to nearly 1% in certain workflows[6], and problematic library preparations (e.g. those involving pooled PCR amplification of tagged amplicons) have shown tag jumping artifacts on the order of a few percent of reads[3]. Such levels of cross-talk can severely inflate apparent diversity or introduce false positives if not controlled[3]. Even in Illumina's patterned flow cells, a low index hopping probability (~0.3–0.9%) can lead to a surprisingly high fraction of spurious "phantom" molecules in low-input single-cell samples (sometimes >8% of reads, and even >80% in near-empty libraries)[10]. In our ONT dataset, the Standard protocol's misassignment rate (~0.5% in the worst case) was of the same order of magnitude as some of these Illumina artifacts. This level of error is unacceptable for some applications. Encouragingly, the PLP+SFB method brought ONT misassignment down to ~0.01%, effectively on par with Illumina's best-in-class UDI workflows (which achieve <0.01%, often only a handful of swaps per run[6]). Thus, by implementing PLP and SFB + PLP we reduce ONT barcoding errors into a regime that is competitive with other sequencing technologies' multiplexing performance.

This improvement is especially significant in low-input or low-biomass contexts, where trace amounts of contamination can dominate the signal. Salter et al.[11] showed that even "blank" controls frequently contain detectable sequences from reagent-derived contamination, and that these artifacts are more prominent when starting biomass is low. More recently it has been recognized that "splashome", cross-contamination between adjacent wells in multiwell sample plates, can produce misleading results, particularly when working with low-biomass samples[12]. While the source of the problem is different, read misassignment due to barcode crosstalk has the potential to produce similarly misleading results. With the PLP protocol (and SFB wash), even libraries prepared from just a few nanograms of DNA can be multiplexed with confidence, knowing that barcodes will not leak into or from other samples. This allows for higher sensitivity in cases where biological signals are intrinsically rare—such as single-cell sequencing or environmental DNA detection—without the risk of misassignment due to barcode crosstalk. Effectively, our protocol elevates the reliability of ONT multiplexing to a level suitable for sensitive, low-input applications.

More broadly, by dramatically reducing barcode misassignment, the PLP and SFB+PLP workflows enhance the suitability of ONT sequencing for high-accuracy applications. For instance, Penter et al.[13] used multiplexed long-read sequencing to profile somatic mutations and T-cell receptor repertoires across cancer samples. In such settings, reliable detection of low-frequency variants or rare clonotypes depends critically on barcode specificity: even low levels of barcode crosstalk can, in principle, obscure a patient's true mutational profile or blur immune-repertoire signatures—precisely the type of scenario in which protocol-level reductions in barcode crosstalk become important. Similarly, Steimle et al.[14] used full-length 16S rRNA gene sequencing with SQK-NBD114.96 to compute IgA-coating indices for individual taxa and relate them to EAE severity in a mouse model of multiple sclerosis. In such enrichment-based designs, where the same taxon can be highly abundant in an IgA^+ fraction but nearly absent in the corresponding IgA^- fraction, even modest cross-sample or cross-fraction crosstalk has the potential to bias IgA indices for low-abundance taxa. In both cases, the standard ONT Native Barcoding Kit (SQK-NBD114) was used, meaning that our PLP and SFB+PLP improvements are directly applicable. By ensuring clean sample demultiplexing even at low DNA input, our method extends the reliability of multiplexed ONT sequencing for both translational genomics and microbial discovery pipelines.

In essence, this technical advance enables researchers to confidently scale up multiplexed ONT experiments (e.g., large metagenomic panels or time-series studies) without sacrificing data integrity. Robust barcode deconvolution is no longer a limiting factor. The SFB+PLP protocol thus significantly expands the utility of ONT sequencing across clinical, environmental, and research domains.

Methods

Case study experimental design, library preparation and sequencing

We designed an experiment that combines a realistic complex community with defined positive and negative controls and evaluates two library preparation protocols in parallel. Genomic DNA extracted from a p-trap water sample was used as the "real" environmental community. As positive controls, we included the ONT DCS *lambda* genome and a *phiX* genome standard purchased from NEB (ΦX174 RF II DNA), the latter provided as circular DNA. Two types of negative controls were prepared: an negative control (water blank) sample in which an equal volume of molecular biology–grade water was processed in place of DNA, and an empty blank in which no DNA or water was added and only the kit components and reagents were carried through library preparation. For each of these samples (p-trap, DCS *lambda*, *phiX*, water blank, and empty blank), we prepared two technical libraries from the same DNA: one using the PLP protocol and one using the standard protocol. The PLP and standard libraries were then sequenced on two separate Flongle flow cells, with 10 ng input DNA per library in all cases.

Case study demultiplexing and taxonomy profiling

We use Dorado v1.0.2 (https://github.com/nanoporetech/dorado) to do demultiplexing for all sequencing results. Pipeline can be found in git repo (pending). Taxonomy profiling step is using Minimap-MEGAN pipeline assembled by PacBio (https://github.com/PacificBiosciences/pb-metagenomics-tools/tree/master/Taxonomic-Profiling-Minimap-Megan) with minor adaptation for HPC deploying.

All analyses, including scripts for basecalling, demultiplexing, mapping, taxonomy profiling, figure plotting, and metric computation, are fully reproducible and available in the accompanying GitLab repository (pending).

Experimental design of the defined-genome multiplex for quantifying barcode crosstalk across protocols

To evaluate barcode crosstalk under different library preparation conditions, we used the Oxford Nanopore Technologies (ONT) SQK-NBD114.96 (Native Barcoding Kit) as the reference workflow. Three ATCC bacterial genomic DNA (ATCC 8482 *Phocaeicola vulgatus*, ATCC 23114 *Mycoplasma hominis* and ATCC 6940 *Corynebacterium striatum*) and the positive control DNA (DCS, *lambda* amplicon DNA provided in the ONT kit) were used as samples with known genomic identities, allowing barcode-to-sample traceability for downstream validation.

Each treatment condition included two genomic-DNA (gDNA) samples and one blank (no-template) control. To minimize system-level variation arising from differences in flow cell performance (e.g., available pores, active channel counts, or runtime stability), we implemented a combinatorial, Latin-square–inspired design stratified by input amount. For each protocol comparison (Standard vs PLP, SFB vs PLP, and SFB vs PLP+SFB), the four genomes were partitioned into two complementary two-species mixtures, with different pairings used for the 3 ng Flongle and 20 ng MinION runs. Within each run, the paired mixtures were prepared using the two protocols being compared, pooled after the adapter-cleanup step, and sequenced on the same flow cell.

Barcode quantification experiment library preparation and sequencing

Four library preparation schemes were tested: standard (ethanol wash), post ligation pooling (PLP), SFB wash (SFB), and SFB + PLP (SFB wash and post ligation pooling). The standard condition corresponds to ONT's previous standard protocol, where ethanol washing is used following barcoding. The SFB condition follows ONT's recently updated protocol (as of Oct. 2025) that replaces ethanol washing with the Short Fragment Buffer (SFB) wash after barcoding and pooling step. The PLP and SFB + PLP protocols represent additional modified workflows developed in this study; detailed step-by-step procedures are provided in the Supplementary Information (SI), including an optimized high-yield version of each protocol.

For each comparison, paired experiments were performed and sequenced concurrently on separate flow cells to isolate the effects of the washing method: Standard vs PLP, SFB vs PLP, and SFB vs SFB + PLP. In total, three MinION and three Flongle flow cells were used. Each Flongle library was prepared from 3 ng of input DNA per sample (12 ng in total), while each MinION library used 20 ng of input DNA per sample (80 ng in total).

Barcode quantification experiment demultiplexing and crosstalk analysis

The demultiplexing steps are the same as case study using Dorado v1.0.2.

Reference genomes were constructed by merging the reference genomic sequences of the three ATCC strains and the DCS DNA obtained from the ONT documentation and ATCC database. This composite reference was used to index Minimap2[15], which performed mapping of all reads under default parameters optimized for ONT long reads.

Alignment results were processed using Samtools[16], from which we extracted for each sample the number of reads mapped to each contig and the number of unmapped reads. Based on these mapping statistics, we calculated precision, recall, F1 score, the fraction of unmapped reads, and the misassignment rate (reads mapping to a genome inconsistent with their assigned barcode).

For the unfiltered dataset, all metrics were computed using only assigned reads—that is, reads successfully demultiplexed into one of the barcode bins, excluding unclassified reads. For the filtered dataset, calculations of precision, recall, and F1 additionally incorporated unmapped and unclassified reads, treating them respectively as false negatives or unassigned cases when computing summary statistics.

References

1. Wick, R. R., Judd, L. M. & Holt, K. E. Deepbinner: Demultiplexing barcoded oxford nanopore reads with deep convolutional neural networks. *PLOS Comput. Biol.* **14**, e1006583, DOI: 10.1371/journal.pcbi.1006583 (2018).
2. Khezri, A. *et al.* Mysterymaster: scraping the bottom of the barrel of barcoded oxford nanopore reads. *BMC Bioinforma.* **26**, DOI: 10.1186/s12859-025-06266-2 (2025).
3. Schnell, I. B., Bohmann, K. & Gilbert, M. T. P. Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Mol. Ecol. Resour.* **15**, 1289–1303, DOI: 10.1111/1755-0998.12402 (2015).

4. van der Valk, T., Vezzi, F., Ormestad, M., Dalén, L. & Guschanski, K. Index hopping on the illumina hiseqx platform and its consequences for ancient dna studies. *Mol. Ecol. Resour.* **20**, 1171–1181, DOI: 10.1111/1755-0998.13009 (2019).
5. Costello, M. *et al.* Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics* **19**, DOI: 10.1186/s12864-018-4703-0 (2018).
6. MacConaill, L. E. *et al.* Unique, dual-indexed sequencing adapters with umis effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics* **19**, DOI: 10.1186/s12864-017-4428-5 (2018).
7. Xu, Y. *et al.* Detection of viral pathogens with multiplex nanopore minion sequencing: Be careful with cross-talk. *Front. Microbiol.* **9**, DOI: 10.3389/fmicb.2018.02225 (2018).
8. Sauvage, T., Cormier, A. & Delphine, P. A comparison of oxford nanopore library strategies for bacterial genomics. *BMC Genomics* **24**, DOI: 10.1186/s12864-023-09729-z (2023).
9. Ligation sequencing amplicons - native barcoding kit 24 v14 (sqk-nbd114.24). *Nanopore technical documents* (2025).
10. Farouni, R., Djambazian, H., Ferri, L. E., Ragoussis, J. & Najafabadi, H. S. Model-based analysis of sample index hopping reveals its widespread artifacts in multiplexed single-cell rna-sequencing. *Nat. Commun.* **11**, DOI: 10.1038/s41467-020-16522-z (2020).
11. Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, DOI: 10.1186/s12915-014-0087-z (2014).
12. Olomu, I. N. *et al.* Elimination of "kitome" and "splashome" contamination results in lack of detection of a unique placental microbiome. *BMC Microbiol.* **20**, DOI: 10.1186/s12866-020-01839-y (2020).
13. Penter, L. *et al.* Integrative genotyping of cancer and immune phenotypes by long-read sequencing. *Nat. Commun.* **15**, DOI: 10.1038/s41467-023-44137-7 (2024).
14. Steimle, A. *et al.* Gut microbial factors predict disease severity in a mouse model of multiple sclerosis. *Nat. Microbiol.* **9**, 2244–2261, DOI: 10.1038/s41564-024-01761-3 (2024).
15. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574, DOI: 10.1093/bioinformatics/btab705 (2021).
16. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, DOI: 10.1093/gigascience/giab008 (2021). Giab008, https://academic.oup.com/gigascience/article-pdf/10/2/giab008/36332246/giab008.pdf.

Acknowledgements

This work was supported by the Engineering Research Centers Program of the National Science Foundation under NSF Cooperative Agreement No. EEC-2133504. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation.

Author contributions statement

QD conceived the experiments, QD conducted the experiments, QD analysed the results, and all authors (QD, CKG, and JAG) reviewed the manuscript.

Additional information

Accession codes

The data have been deposited with links to BioProject accession number PRJNAxxxxxx (ACCESSION NUMBER PENDING) in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/).

Competing interests

The authors declare no competing interests.