Received: 25 May 2017 | Revised: 4 August 2017 | Accepted: 8 August 2017

DOI: 10.1111/1755-0998.12713

RESOURCE ARTICLE

WILEY MOLECULAR ECOLOGY RESOURCES

A novel post hoc method for detecting index switching finds no evidence for increased switching on the Illumina HiSeq X

Gregory L. Owens[1] | Marco Todesco[1] | Emily B. M. Drummond[1] | Sam Yeaman[2] | Loren H. Rieseberg[1]

[1]Department of Botany and Beaty Biodiversity Centre, University of British Columbia, Vancouver, BC, Canada

[2]Department of Biological Sciences, University of Calgary, Calgary, AB, Canada

Correspondence
Gregory L. Owens, University of British Columbia, Vancouver, BC, Canada.
Email: gregory.owens@alumni.ubc.ca

Abstract

High-throughput sequencing using the Illumina HiSeq platform is a pervasive and critical molecular ecology resource, and has provided the data underlying many recent advances. A recent study has suggested that "index switching," where reads are misattributed to the wrong sample, may be higher in new versions of the HiSeq platform. This has the potential to invalidate both published and in-progress work across the field. Here, we test for evidence of index switching in an exemplar whole-genome shotgun data set sequenced on both the Illumina HiSeq 2500, which should not have the problem, and the Illumina HiSeq X, which may. We leverage unbalanced heterozygotes, which may be produced by index switching, and ask whether the undersequenced allele is more likely to be found in other samples in the same lane than expected based on the allele frequency. Although we validate the sensitivity of this method using simulations, we find that neither the HiSeq 2500 nor the HiSeq X has evidence of index switching. This suggests that, thankfully, index switching may not be a ubiquitous problem in HiSeq X sequence data. Lastly, we provide scripts for applying our method so that index switching can be tested for in other data sets.

KEYWORDS
barcode, bioinformatics/phyloinformatics, genomics/proteomics, index hopping, sequencing

1 | INTRODUCTION

High-throughput sequencing, primarily through the Illumina HiSeq platform, has revolutionized molecular ecology. In fact, 50% of original articles in a recent issue of *Molecular Ecology* (Vol 26, Issue 2) included Illumina-derived sequence data. Researchers can now explore questions that were completely unanswerable before current sequencing technologies, using approaches such as genome scans, genome assembly and high-density genetic mapping (e.g., Gould & Stinchcombe, 2017; Li et al., 2017; Standage et al., 2016). With the central role that sequencing plays, it is alarming that a recent preprint suggests increased index switching on the new HiSeq 4000 and HiSeq X machines (Sinha et al., 2017).

To prepare DNA for Illumina sequencing, DNA strands are fragmented, and adapter sequences are attached to the ends of these fragments. These adapters contain the sequence that binds to the flow cell, a primer sequence for amplification during sequencing and, potentially, a barcode index for linking reads to individual samples. Indexes are required when multiplexing samples within a single sequencing lane, and can be included in adapters at one or both ends of the DNA fragments. As the output of a single sequencing lane has increased, multiplexing has become increasingly common. This is especially true in molecular ecology, where researchers often aim to maximize sample size using low-coverage whole-genome data (Buerkle & Gompert, 2013). For example, a single lane on the HiSeq 4000 can sequence 200 stickleback genomes (~460 MB) to 1×

coverage. Consequently, it is critical that samples are correctly demultiplexed or the resulting sequence data will contain mixes of reads from unexpected and unpredictable sources.

A recent preprint by Sinha et al. (2017) reports high levels of index switching in a single-cell RNAseq experiment. They dual indexed (i.e., barcodes on both adapters) all samples using a Nextera XT kit and found that samples that shared a single index had greater similarity in gene expression levels than expected. The authors attributed this to index switching, and showed that controls containing adapters and index primers but no template DNA still had reads assigned to them, receiving 5%–7% of the average number of reads of samples with template DNA as a result of index switching. They proposed that index switching occurs during cluster generation (before sequencing) when free index primers replicate already indexed library fragments. These newly copied fragments will then carry one wrong index and be misattributed to another sample. Importantly, they find that this only occurs on the Illumina HiSeq 4000, which uses a patterned flow cell and a new exclusion amplification (ExAmp) chemistry, and not in the NextSeq 500, which does not. Both the HiSeq 4000 and HiSeq X use a patterned flow cell and the cBot 2 system for cluster generation, suggesting that the problem may occur in both machines. Illumina has acknowledged that index switching can occur and is higher in machines that use a patterned flow cell, but suggests total index switching is <2% of reads (Illumina Inc, 2017).

In the light of the potential problems, we explored a set of whole-genome-sequenced samples, half of which were sequenced on the HiSeq 2500, which does not use the patterned flow cell and ExAmp chemistry, and half on the HiSeq X, which does. We have developed a novel method for detecting index switching in genomic data sets and show that in our samples, index switching is not significantly enriched in the HiSeq X.

2 | METHODS

2.1 | Study species and library preparation

To identify whether index switching was detectable in an average whole-genome sequence data set, we analysed a set of 323 wild *Helianthus annuus* (common sunflower) whole-genome sequence samples. Plants were grown from field-collected seeds obtained from 28 populations located across the Midwestern USA and Southern Canada. Genomic DNA was extracted from frozen leaf tissue using either a modified CTAB protocol (based on Murray & Thompson, 1980), the DNeasy Plant Mini Kit or the DNeasy 96 Plant Kit (Qiagen, Hilden, Germany). DNA was sheared to an average fragment size of 350 bp using a Covaris M220 ultrasonicator (Covaris, Woburn, MA, USA), following the manufacturer's recommendations. Seven hundred and fifty nanograms of sheared DNA was used as starting material to prepare paired-end whole-genome shotgun Illumina libraries, using a protocol largely based on Rowan, Patel, Weigel, & Schneeberger (2015) the TruSeq DNA Sample Preparation Guide from Illumina (Illumina, San Diego, CA, USA) and Rohland & Reich (2012). End-repairing of the sheared DNA fragments was performed using the NEBNext End Repair Module (NEB, Ipswich, MA, USA). The fragments were then A-tailed using Klenow Fragment (3′–>5′exo-) from NEB and ligated to 24-bp-long, nonbarcoded adapters with a 3′ T-overhang (Table S1) using the Quick Ligation Kit from NEB. After each enzymatic step, the reactions were purified using 1.6 volumes of a solution of paramagnetic SPRI beads (MagNA), prepared according to Rohland & Reich (2012). An enrichment step was then performed using KAPA HiFi HotStart ReadyMix (Roche, Basel, Switzerland) and short, nonindexed primers that do not extend the adapters (Table S1). The reactions were then purified using 1.6 volumes of MagNA beads. The sunflower genome contains a very large amount of highly repetitive sequences derived from the recent expansion of two retrotransposon families (Staton et al., 2012). To reduce the representation of repetitive sequences, the enriched libraries were treated with a duplex-specific nuclease (DSN; Evrogen, Moscow, Russia), following the protocols reported in Shagina et al. (2010) and Matvienko et al. (2013) with modifications. The fragments were then further amplified using Kapa HiFi HotStart ReadyMix and primers (to a final concentration of 0.4 μM each) to complete the adapters and add a six-bp index to the P7 adapter (Table S1). The sequence of the completed adapters is identical to that of Illumina's TruSeq adapters.

After amplification, the libraries were purified twice with 1.6 volumes of MagNA beads, quantified using a QuBit dsDNA Broad Range Assay Kit (Invitrogen, Carlsbad, CA, USA) and analysed on a 2100 Bioanalyzer instrument using a High Sensitivity DNA Analysis Kit (Agilent, Santa Clara, CA, USA). The libraries were then quantified on an iQ5 Real Time PCR Detection System (Bio-Rad, Hercules, CA, USA) using Maxima SYBR Green qPCR Master Mix (ThermoFisher Scientific, Waltham, MA, USA) to determine molarity, and pools consisting of ten libraries each were prepared. All libraries were sequenced at the Genome Quebéc Innovation Center; 156 libraries were sequenced on a HiSeq 2500 instrument and 165 were sequenced on a HiSeq X instrument (Illumina, San Diego, CA, USA). Importantly, samples were multiplexed within lanes in a random manner without regard to population ID.

2.2 | Variant calling

We aligned all samples to the *H. annuus* XRQ genome using BWA (version 0.7.9a), removed PCR duplicates using samtools, realigned potential indels using GATK and called variants using FREEBAYES (version 1.1.0) (Badouin et al., 2017; Garrison & Marth, 2012; Li & Durbin, 2010; Li et al., 2009; McKenna et al., 2010). In all cases, we used default parameters. For this analysis, we selected di-allelic SNPs with QUAL >30, and outside of regions known to contain transposable elements, using vcflib (https://github.com/ekg/vcflib).

2.3 | Testing for index switching

To identify whether index switching is increased in samples sequenced on the HiSeq X, we leveraged the fact that individual samples in our data set were either sequenced on the HiSeq X or the HiSeq 2500. Therefore, we cannot only estimate index switching

rates on the HiSeq X, but also tell if it is higher than for previous technology. Crucially, the method proposed here does not require index switching to actually alter genotype calls. Whether index switching changes genotype calls would depend on the samples sequenced, the overall depth and the genotyping program, and therefore, genotype changes may not be a reliable signal of the level of index switching occurring.

Previous work has suggested that index switching is occurring for 1%–10% of reads depending on factors during library preparation and sequencing (Sinha et al., 2017). This low level means that, for our data set, at a single locus, an allele acquired because of index switching is likely to only have one read, given moderate overall read depth. We looked for these unbalanced heterozygotes (i.e., one read for allele 1, many reads for allele 2) with ≥5 reads total and asked if the rare allele (i.e., the undersequenced allele) was found in other samples sequenced in the same lane (which we refer to as "allele sharing"). It should be noted that these unbalanced heterozygotes need not actually be called as heterozygotes. Depending on the base quality scores and the overall sequencing depth, they may be called as homozygotes. We then calculated $\hat{p}$, the binomial probability that the rare allele should be found in one or more of the samples based on f, the allele frequency for all samples sequenced with that machine (excluding the unbalanced focal individual) and n, the number of other samples with genotypes in the lane (1).

$$\hat{p} = 1 - (1 - f)^{2n}. \quad (1)$$

We then plotted $\hat{p}$, the binomial probability of cases where the allele is present in at least one copy in the other samples from the lane, against p, the observed proportion of cases with allele sharing. We fit a line to this relationship using a generalized additive model in the *stat_smooth* command from *ggplot2* (Wickham, 2016). If index switching is not occurring, we expect a straight line at $\hat{p} = p$. Alternatively, if index switching is occurring, we expect $\hat{p} > p$ indicating greater sharing of undersequenced alleles within a lane than expected by chance. These proportions were calculated independently for HiSeq 2500 and HiSeq X samples, using the first 500,000 variable sites in the genome.

As a control, for each unbalanced heterozygote, we calculated the p using the same number of genotyped samples sequenced using the same machine, but not the same lane. This control should not show evidence of index switching.

It is important to note that if samples were sorted into sequencing lanes based on a genetic grouping (e.g., species or population), we would find $\hat{p} > p$ in the absence of index switching. In our data set, this is not the case, as samples were randomly assigned into lanes.

2.4 | Simulations

To explore the sensitivity of our measure of index switching, we bioinformatically switched reads in our vcf file, randomly selecting 0, 0.1, 0.5, 1, 5 or 10 per cent of reads at each site across all individuals to be switched. Switched reads were removed from the individual (i.e., reducing read depth) and added to another individual sequenced in the same lane (i.e., increasing read depth). We then recalculated genotypes simply by assigning samples containing reads for both alleles as heterozygotes. These simulations were run through the same algorithm to detect index switching.

Our initial simulation called heterozygotes regardless of allelic balance, while FreeBayes and other genotyping programmes may call a genotype as homozygous if there is an extreme bias in read depth between alleles. To approximate this, we also ran the same simulation but required heterozygotes to have allele balance >20% (i.e., >20% of reads must come from each allele).

3 | RESULTS

We fail to find evidence that index switching is occurring in our data set. For samples sequenced on both machines, the observed proportion of allele sharing within a lane tracked the predicted proportion closely (Figure 1, Figure S1). This was consistent with the pattern seen in our control that used samples from different lanes. Despite this, we find that our method is able to identify index switching in the simulated data set. In particular, we find elevated allele sharing around $\hat{p} = 0.2$, even when index switching only represents 1% of reads (Figure 2). In our data set, $\hat{p} = 0.2$ corresponds to rare alleles

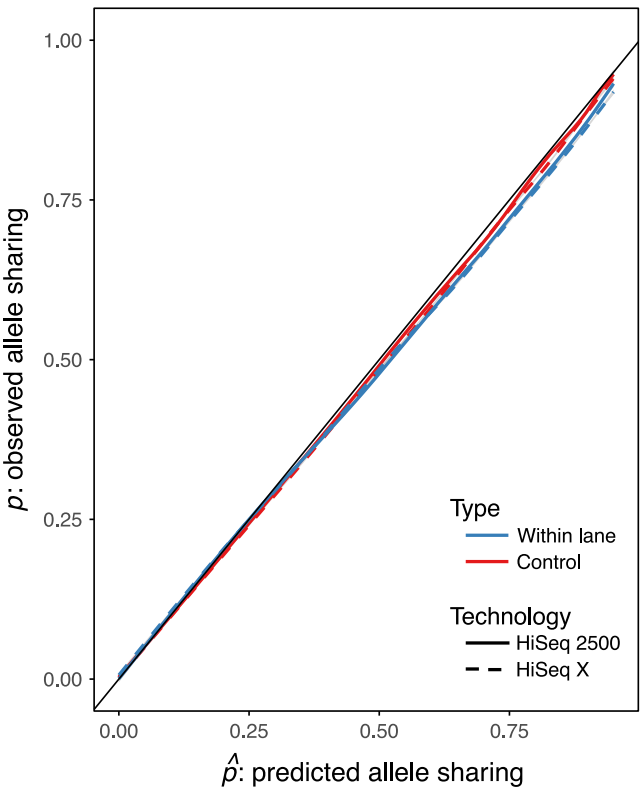

FIGURE 1 (a) The relationship between predicted allele sharing and observed allele sharing for samples sequenced on the HiSeq 2500 (solid line) and HiSeq X (dashed line). Allele sharing was calculated for samples sequenced together in a lane (blue) and for a control group sequenced in different lanes (red) [Colour figure can be viewed at wileyonlinelibrary.com]

(minor allele frequency <5%). This makes sense because common alleles are expected to have high allele sharing even in the absence of index switching which makes the signal more difficult to observe. When using simulations that required allelic balance to call heterozygotes, the same pattern was produced (Figure S2).

4 | DISCUSSION

Widespread, undetected index switching represents a nightmare scenario for molecular ecologists worldwide. Here, we show that in one exemplar data set, index switching is not higher in samples sequenced on the new Illumina patterned flow cells and is likely below 1% of reads. Furthermore, we provide a way to visualize index switching for sequenced genomic data sets.

4.1 | Post hoc index switching detection

The method presented here detects the sharing of rare alleles within lanes that index switching could produce. Previous methods for detecting index switching relied on specific a priori library preparation, while the method described here can be applied to data already sequenced before the issue of index switching was brought to light (Sinha et al., 2017); however, it does rely on the random assignment of samples to sequencing lanes. If genetically related samples are clustered in lanes, then a false positive will occur. While our method relies on visual inspection of data, future programmes could use the same allele sharing information to specifically estimate the amount of index switching in a Bayesian or likelihood model.

Index switching presents similar issues as contamination, a problem long recognized in human genomic studies (Jun et al., 2012). Contamination is when DNA of multiple samples is combined during DNA extraction or library preparation and also results in misattributed reads. This issue has been tackled in a human genomics context, but has unique challenges for nonmodel organisms. Studies using nonmodel organisms often work with much lower coverage per sample, have less accurate or no prior allele frequency information and poorer prior expectations of overall heterozygosity when compared to human studies. Nevertheless, methods for removing contamination may be repurposed to incorporate information from entire sequencing lanes to remove index switching (Flickinger, Jun, Abecasis, Boehnke, & Kang, 2015).

4.2 | Why don't we find index switching?

Our results are clearly different from Sinha et al. who found index switching affecting 5%–10% of reads. One possibility could be that this is caused by differences in sequencing library preparation. Sinha et al. used cDNA as starting material and the Nextera tagmentation technology from Illumina to fragment the DNA and tag the fragments

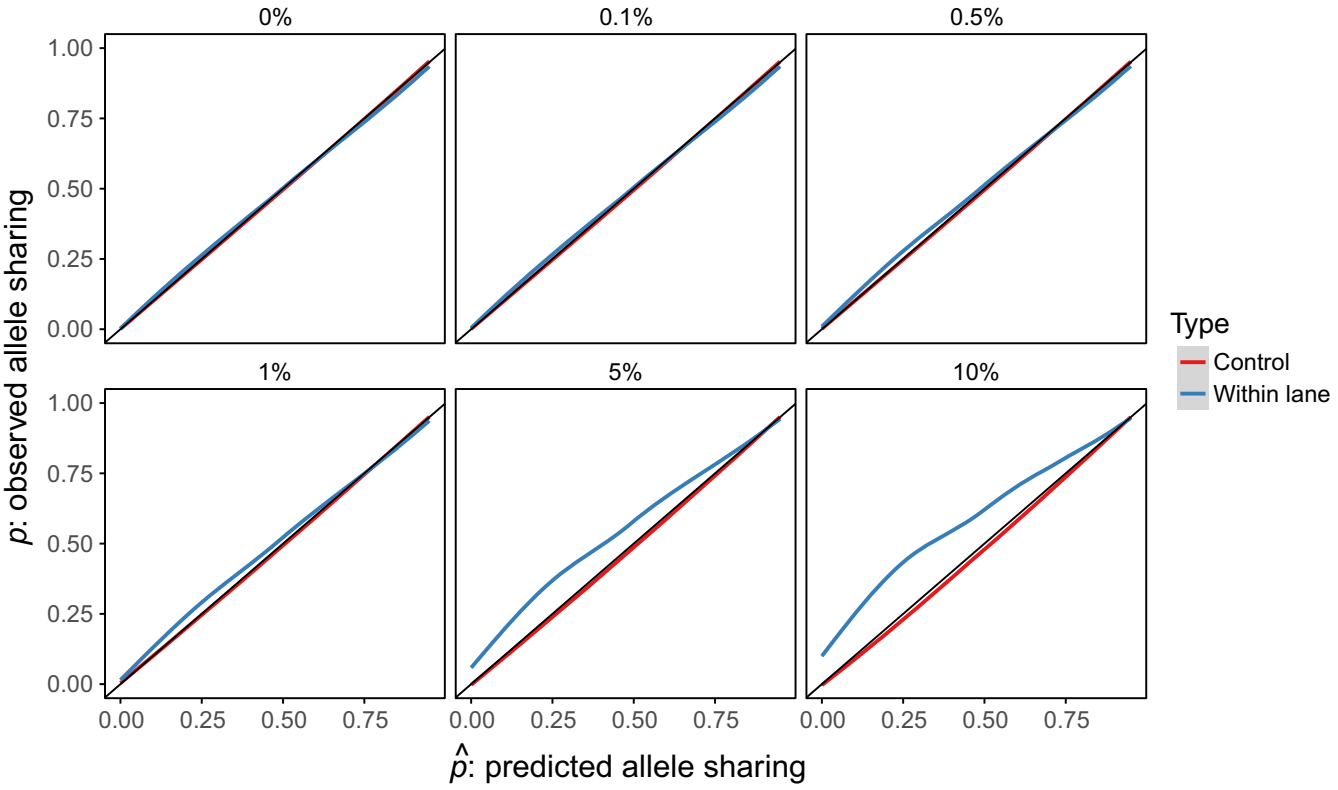

FIGURE 2 The relationship between predicted allele sharing and observed allele sharing with different degrees of simulated index switching. The 0% index switching test controls for the recalling of genotypes that occur during simulated index switching. Allele sharing was calculated for samples sequenced together in a lane (blue) and for a control group sequenced in different lanes (red) [Colour figure can be viewed at wileyonlinelibrary.com]

with adapters, whereas we used genomic DNA sheared using ultrasonication and then added the adapters to the fragments via enzymatic ligation. Furthermore, our protocol included a depletion step, to reduce repetitive elements in the genome that is not present in the Nextera XT protocol. However, the final step of library preparation is substantially equivalent between the two approaches; DNA fragments with short adapters at their extremities are PCR-amplified using primers that complete the adapters and add unique sequence indices, allowing pooling of different samples in a single flow cell. Given that carry-over of free-indexed primers from this step is the likely cause of index switching during the ExAmp procedure (Sinha et al., 2017), the two approaches can be confidently compared for the purpose of investigating the occurrence of index switching.

Another possible difference between the two experiments is that while the Nextera XT kit uses dual indices (i.e., both the P5 and P7 adapters are indexed), we used only a single index on the P7 adapter. This has the potential to halve index switching in our data set, assuming that switching occurs equally from both adapters. If the unindexed P5 adapter was to be replaced in our data set, this would not result in index switching because no index is present. For a dual-indexed library, it would result in index switching.

Finally, the main difference we noticed between our libraries and the one shown in Figure 4B of Sinha et al. (2017) is the large amount of free adapters/primers that are found in the latter (compare with the Bioanalyzer plot for one of our libraries in Figure 3a). Our enhanced clean-up efficiency could be due to fact that while the Nextera XT kit recommends a single clean-up step with 0.6 volumes of Agencourt AMPure XP beads, we performed two rounds of clean-up with 1.6 volumes of MagNA beads (the maximum size of the fragments that are removed during beads clean-up is, approximately, inversely proportional to the ratio of bead solution that is added to the reaction—smaller volumes of beads should therefore be more efficient at removing free adapter/primers). However, a single clean-up with 1 volume of MagNA beads was sufficient to completely remove primers/adapters from our libraries (Figure 3b). MagNA and AMPure XP beads have been shown to have comparable recovery efficiency and size discrimination (Rohland & Reich, 2012), and this is confirmed by our experience. While it is possible that, because of their different design, libraries produced using the Nextera XT protocol simply contain a much larger amount of free adapters/primers that cannot be efficiently removed with one single clean-up step, we did not directly test this.

4.3 | When is index switching confounding?

Certain kinds of experiments are more likely to be affected by index switching. Gene expression quantification using RNAseq is especially sensitive because highly expressed genes can bleed into other samples, homogenizing expression measures with lanes. In cancer genomics, low-frequency alleles represented by a minority of reads are both important and can be produced by index switching. Similar issues can occur in Pool-seq experiments used in molecular ecology, where index switching could affect estimation of allele frequencies, slightly homogenizing differences among pools sequenced in the same lane.

For high-coverage genomic sequencing of diploid organisms, index switching can produce unbalanced heterozygotes, where one allele is

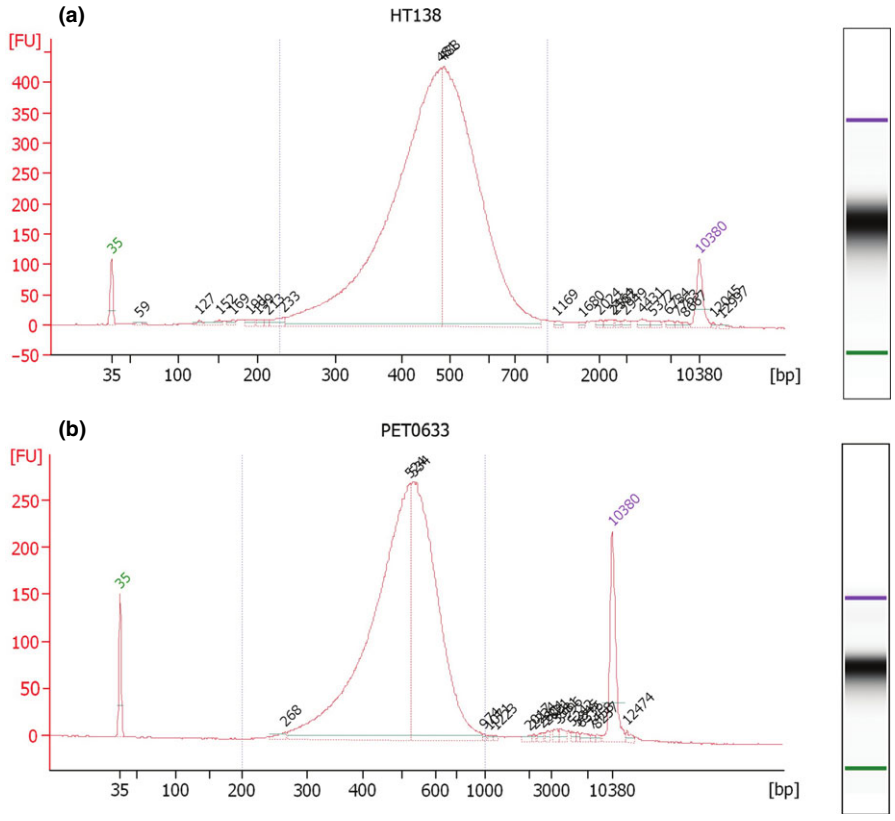

FIGURE 3 Bioanalyzer plots for representative whole-genome shotgun sequencing libraries used in this study, after the final amplification and clean-up step. The plot shows the abundance of fragments of different sizes in the library (measured in fluorescence units, FU). The peaks at 35 bp (green) and 10,380 bp (purple) are internal standards. Free index primers should appear as a peak at ~50 bp. (a) Library that underwent two rounds of clean-up after PCR amplification, each using 1.6 volumes of MagNA beads. (b) Library that underwent a single round of clean-up with 1 volume of MagNA beads [Colour figure can be viewed at wileyonlinelibrary.com]

represented by one or two reads and the other by many reads. These present a genotyping challenge because unbalanced heterozygotes can also be produced naturally by stochastic sampling of alleles or via PCR bias during library prep. Future genotyping programmes may use haplotype information of reads along with sequencing lane identity to detect when index switching is occurring and remove contaminants. In low-coverage genome sequencing, identifying individual instances of index switching may be impossible and will result in an increased rate of false heterozygote genotype calls (when an index is switched among alternate homozygotes) and slightly increased quality scores for heterozygotes miscalled as homozygotes (when all sequenced reads represent only one allele of the heterozygote and an index is switched from a homozygote with the same allele).

Which samples are multiplexed in a lane has a large effect on whether index switching is a problem. If each sample represents a distinct, distantly related species, then misattributed reads are unlikely to align to a reference genome. If all samples are from a single population, misattributed reads are more likely to carry alleles already present. In the worst case scenario, samples of closely related species or distantly related populations with misattributed reads could be mistakenly inferred as novel alleles. This could reduce divergence estimates like F_{ST} or confuse phylogenetic signals. Although stringent allelic balance cut-offs for heterozygous genotypes would remove the false heterozygotes from index switching, it may also remove true heterozygotes or miscall them as homozygotes, especially at lower (<10) average read depth.

Whether low levels of index switching result in incorrect genotype calls is partially determined by the genotyping programme. In this study, we used FreeBayes, which, similar to GATK, uses a Bayesian model incorporating information across samples including allele balance. False heterozygote calls from index switching may be less likely with Bayesian methods than more simplistic genotyping models but could also result in the removal of real variants because of skewing of data set wide allelic balance, a parameter often used for filtering. In the same vein, joint genotyping allows for variants to be called at lower coverage if they are found in other individuals in the sample set, but this is also the exact pattern produced by index switching.

4.4 | Best practices to avoid index switching

Although we failed to detect index switching here, it may be prudent to employ techniques for avoiding the issue. Two main suggestions have been proposed, and we support both: (i) using dual index barcodes, so that both indices are unique to a sample and (ii) thoroughly cleaning library preparations to remove free primers. If sequencing is already completed and index switching is suspected, then bioinformatic filters can be applied although none are ideal. Filtering for allelic balance (i.e., the fraction of nonreference reads in heterozygotes) is an obvious choice and is often used to filter variant sets (Li, 2014), but low levels of index switching will produce a small fraction of heterozygotes so this filter may not be effective. Raising minimum depth to call genotypes should prevent unbalanced heterozygotes from being called as heterozygous, but may drastically reduce genotyping rates. In our data set, unbalanced heterozygotes were called as homozygotes exclusively when total genotype depth was ≥14 reads (Figure S3), although this exact threshold will be contingent on the specific data set and genotyping program. Simply requiring more stringent allelic balance in individual genotypes will undoubtedly undercall true heterozygotes, potentially a worse problem, although future genotyping programmes may take this information into account intelligently (e.g., Flickinger et al., 2015). Beyond this, researchers should be more aware of what samples are multiplexed together, a process that is often determined by the sequencing facility without regard to sample identity.

5 | CONCLUSION

We have failed to find evidence for index switching here, but we certainly do not make the claim that it cannot or does not happen. However, we would like to make two points: (i) index switching does not always occur and (ii) vigilance is necessary. With greater attention to this problem, research laboratories and companies can spend time and effort creating molecular protocols to reduce this issue and bioinformatic programs to detect or remove misattributed reads. Like all genotyping methods, errors are inevitable, but by better understanding their source, we can sort signal from noise.

AUTHOR CONTRIBUTIONS

G.O. project conception, methods development, bioinformatics, and original draft preparation; E.D. and M.T. sequence data collection; G.O., E.D., M.T., S.Y. and L.R. reviewing and editing manuscript.

DATA ACCESSIBILITY

All scripts and data used in this manuscript are available on github (https://github.com/owensgl/index_investigator).

ORCID

Gregory L. Owens http://orcid.org/0000-0002-4019-5215
Marco Todesco http://orcid.org/0000-0002-6227-4096

REFERENCES

Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., … Legrand, L. (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature*, *546*(7656), 148–152.

Buerkle, A. C., & Gompert, Z. (2013). Population genomics based on low coverage sequencing: How low should we go? *Molecular Ecology*, *22* (11), 3028–3035.

Flickinger, M., Jun, G., Abecasis, G. R., Boehnke, M., & Kang, H. M. (2015). Correcting for sample contamination in genotype calling of DNA sequence data. *The American Journal of Human Genetics*, *97*(2), 284–290.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN]

Gould, B. A., & Stinchcombe, J. R. (2017). Population genomic scans suggest novel genes underlie convergent flowering time evolution in the introduced range of *Arabidopsis thaliana*. *Molecular Ecology*, *26*, 92–106. https://doi.org/10.1111/mec.13643

Illumina Inc. (2017). Effects of index misassignment on multiplexing and downstream analysis.

Jun, G., Flickinger, M., Hetrick, K. N., Romm, J. M., Doheny, K. F., Abecasis, G. R., … Kang, H. M. (2012). Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *The American Journal of Human Genetics*, *91*(5), 839–848.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*, 589–595.

Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*, *30*(20), 2843–2851.

Li, Z., Guo, B., Yang, J., Herczeg, G., Gonda, A., Balázs, G., … Merilä, J. (2017). Deciphering the genomic architecture of the stickleback brain with a novel multilocus gene-mapping approach. *Molecular Ecology*, *26*, 1557–1575. https://doi.org/10.1111/mec.14005

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, *25*, 2078–2079.

Matvienko, M., Kozik, A., Froenicke, L., Lavelle, D., Martineau, B., Perroud, B., & Michelmore, R. (2013). Consequences of normalizing transcriptomic and genomic libraries of plant genomes using a duplex-specific nuclease and tetramethylammonium chloride. *PLoS ONE*, *8*(2), e55913.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303.

Murray, M. G., & Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*, *8*(19), 4321–4325.

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, *22*(5), 939–946.

Rowan, B. A., Patel, V., Weigel, D., & Schneeberger, K. (2015). Rapid and inexpensive whole-genome Genotyping-by-Sequencing for crossover localization and fine-scale genetic mapping. *G3*, *5*(3), 385–398.

Shagina, I., Bogdanova, E., Mamedov, I. Z., Lebedev, Y., Lukyanov, S., & Shagin, D. (2010). Normalization of genomic DNA using duplex-specific nuclease. *BioTechniques*, *48*(6), 455.

Sinha, R., Stanley, G., Gulati, G. S., Ezran, C., Travaglini, K. J., Wei, E., … Weissman, I. L. (2017). Index switching causes "spreading-of-signal" among multiplexed samples in Illumina HiSeq 4000 DNA sequencing. *bioRxiv*. https://doi.org/10.1101/125724

Standage, D. S., Berens, A. J., Glastad, K. M., Severin, A. J., Brendel, V. P., & Toth, A. L. (2016). Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. *Molecular Ecology*, *25*, 1769–1784.

Staton, S. E., Bakken, B. H., Blackman, B. K., Chapman, M. A., Kane, N. C., Tang, S., … Burke, J. M. (2012). The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements. *The Plant Journal*, *72*(1), 142–153.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer-Verlag.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Owens GL, Todesco M, Drummond EBM, Yeaman S, Rieseberg LH. A novel post hoc method for detecting index switching finds no evidence for increased switching on the Illumina HiSeq X. *Mol Ecol Resour*. 2018;18:169–175. https://doi.org/10.1111/1755-0998.12713