LI-Seq: A Cost-Effective, Low Input DNA method for Whole Genome Library Preparation

Authors: Teia M. Schweizer[1*], Matthew G. DeSaix[1], Kristen C. Ruegg[1]

* Corresponding Author: Teia.Schweizer@gmail.com

[1] Department of Biology, Colorado State University, Fort Collins, CO, 80525

Summary

1. Samples from species of high conservation concern are often low in total genomic DNA. Whole Genome Sequencing (WGS) can provide many insights that can be used to aid in species conservation, but current methods for working with low quality and low input samples can be cost prohibitive for population level genomic analyses. Thus, there is an urgent need for a cost-effective method of preparing WGS libraries from low input DNA samples.
2. To bridge the gap between sampling techniques commonly used in conservation genetics that yield low quality and low input DNA and the powerful tool of WGS, we developed LI-Seq, a more efficient method that successfully produces libraries from low quality DNA with as low input as 0.48 ng of DNA, with an average final library size of 300-500 base pairs.
3. Sequencing results suggest no difference in sequencing quality or coverage between low quality, low input and high quality, high input starting material using our protocol. We

conclude that our new method will facilitate high-throughput WGS on low quality, low input samples, thus expanding the power of genomic tools beyond traditional high quality samples.

Keywords: whole genome sequencing, low input DNA, low quality DNA, conservation genomics, cost-effective, population genetics

Introduction

The field of genomics began developing in the 1990s with whole genome shotgun sequencing of bacteria (Weissenbach 2016) and has advanced rapidly with the improvement of High-Throughput Sequencing (HTS) techniques (Goodwin et al 2016). The ability to sequence whole genomes with relative ease has opened new research avenues and made it possible to estimate fundamental population genetic parameters with increasing precision in both model and non-model organisms (Allendorf, Hohenloe, and Luikart 2010; Ouborg et al 2010). Of particular relevance to ecologists and conservation biologists, HTS has made it possible to investigate previously challenging topics such as the genetic basis of local adaptation, patterns of inbreeding across the genome, and how species adapt to changing climate conditions (Kohn et al 2006; Ruegg et al 2018; Allendorf, Hohenloe, and Luikart 2010; Ouborg et al 2010). As a result, genomic tools are revolutionizing the fields of ecology, evolution, and conservation biology.

Despite the proliferation of HTS methods for model organisms (Ekblom and Galindo 2011), there remain a number of technical and financial limitations to the widespread use of genomic approaches in situations where the amount of input DNA maybe limited and costs are a concern. While the cost of sequencing has dramatically decreased in the last two decades (Goodwin et al 2016), it is often still prohibitively high for use in population-level studies where hundreds or thousands of individuals must be sequenced (Fuentes-Pardo and Ruzzante 2017). Methods modified from commercially available whole genome sequencing (WGS) library preparation kits offer low coverage options at a fraction of the cost per individual, making them suitable for population genetics studies (Therkildsen and Palumbi 2017; Kryazhimskiy et al 2014; Baym et al 2015). However, these methods still typically require high quality and high input DNA and are not optimized to efficiently amplify smaller target library sizes. Such high quality and quantity DNA can be difficult to attain when working with threatened, endangered, or cryptic species, where ethical and logistical challenges are often prohibitive (Kohn et al 2006; Ouborg et al 2010). However, samples that yield low quality and quantity DNA have previously found limited use in whole genome studies unless potentially cost-prohibitive library preparation kits or methods are employed (Taylor et al 2020). Given the immense potential benefits of analyzing whole genomes for effective wildlife conservation and management efforts (Funk et al 2012; Ryder 2005; Russello et al 2015), there is an urgent need for a cost-effective method of preparing WGS libraries from low quality and quantity DNA samples.

Low quality and quantity samples are often a hallmark of noninvasive or minimally invasive sampling techniques. Noninvasive genetic sampling methods first gained recognition in 1992 when DNA was successfully extracted from passively-collected hair for a genetic study of an endangered bear species (Taberlet and Bouvet 1992). Since then, noninvasive genetic sampling has been successfully used in genetic studies across myriad taxonomic groups (Stenglein et al 2010; Valiere et al 2003; Roques et al 2014; Regnaut et al 2006). Noninvasive sampling encompasses samples such as saliva, hair, feces, or feathers, collected without capturing, handling, or otherwise disturbing the study organism (Waits et al 2005). Minimally invasive sampling entails capturing or handling a study organism with minimal invasion or tissue collection (e.g. feather pulls and buccal swabs) (Carroll et al 2017). In recent years, noninvasive and minimally invasive sampling methods have gained popularity, especially for use in monitoring threatened and endangered species (Lukacs and Burnham 2005; Fuentes-Pardo and Ruzzante 2017). However, noninvasively or minimally invasively collected samples typically yield lower concentrations of DNA which can limit their use in whole genome studies unless expensive library preparation kits or library preparation methods are employed (Taylor et al 2020). Given the immense potential benefits of analyzing whole genomes for effective wildlife conservation and management efforts (Funk et al 2012; Ryder 2005; Russello et al 2015), there is an urgent need for a cost-effective method of preparing WGS libraries from noninvasively or minimally invasively collected samples.

Although low cost methods for WGS exist (Therkildsen and Palumbi 2017; Kryazhimskiy et al 2014; Baym et al 2015), they still require a prohibitively large amount of high quality input DNA (2.5 ng) for many conservation applications. More specifically, further analysis of these methods reveals that much of the DNA is wasted during the library preparation step due to the fact that the average fragment size produced from these methods is 1kb, but the average fragment size needed for many common sequencing platforms, such as Illumina, is 300-500 base pairs. Thus, DNA above 500 base pairs is often removed prior to sequencing. To bridge the gap between sampling techniques commonly used in conservation genetics that yield low quality DNA and the powerful tool of WGS, we developed a more efficient method that successfully produces libraries from low quality DNA with as low input as 0.48 ng of DNA, with an average final library size of 300-500 base pairs.

We demonstrate the utility of our method for producing high quality sequencing data at a fraction of the cost of traditional library preparation methods using DNA extracted from a single flight feather calamus, or quill, of a small (8-9 grams) passerine bird, the American Redstart (*Setophaga ruticilla*). We compare the sequence data from our low input DNA library (from feather) to those generated from a high input library (from blood) and demonstrate that our method produces comparable sequence quality for both low and high input DNA sources. These results have important implications for conservation genomics research seeking to maximize efficient sequencing from low input DNA samples.

Materials and Methods

Library Preparation for Sequencing

We identified three key parameters in other methods (i.e. Therkildsen and Palumbi 2017) that we could optimize in order to target the ideal fragment distribution and avoid loss of critical DNA when working with low input DNA samples. The three key parameters modified herein were: (1) the ratio of tagmentation transposome (which cleaves DNA and adds an adapter for indices) to input DNA quantity, (2) the duration of tagmentation incubation, and (3) the duration of the indexing PCR elongation time.

In order to compare sequencing quality from high and low input and quality DNA libraries and assess the efficiency of our method, we extracted DNA from 50 high DNA quantity and quality bird blood samples and 50 low DNA quantity and quality feather calamus tips of the American Redstart. For blood, DNA was extracted from between 50-100 µL of whole blood stored in Queen's Lysis Buffer (~80 µL of whole blood plus 300 µL of buffer), using Qiagen DNEasy Blood and Tissue Kit and eluted into 100 µL of provided AE buffer. For extractions from feathers, like other low-quality samples, maximizing DNA yield is critical. Therefore, we followed the Qiagen protocol but with the following modifications. To each sample, we added 10 µL of 1 M Dithiothreitol (DTT) to the initial lysis step to aid in breaking down disulfide bonds found in the keratin of feathers. Flowthrough after the first filtration step when lysate was transferred to the spin column was pipetted back onto the filter for a second centrifugation. Prior to the final elution step, AE buffer was placed in an incubator at 56 ^{o}C. During the final elution step, AE

buffer was left to incubate on the filter for five minutes instead of two. We eluted feather extractions into 400 µL (two rounds of 200 µL elutions through the spin column as recommended by Qiagen protocol for maximum yield). Prior to proceeding with library prep, we concentrated feather DNA extractions using a 1:1 ratio of Serapure beads (Faircloth and Glenn 2014) from 400 µL to 15 µL and eluted into 10mM Tris-Hcl (Figure 1, step 1a).

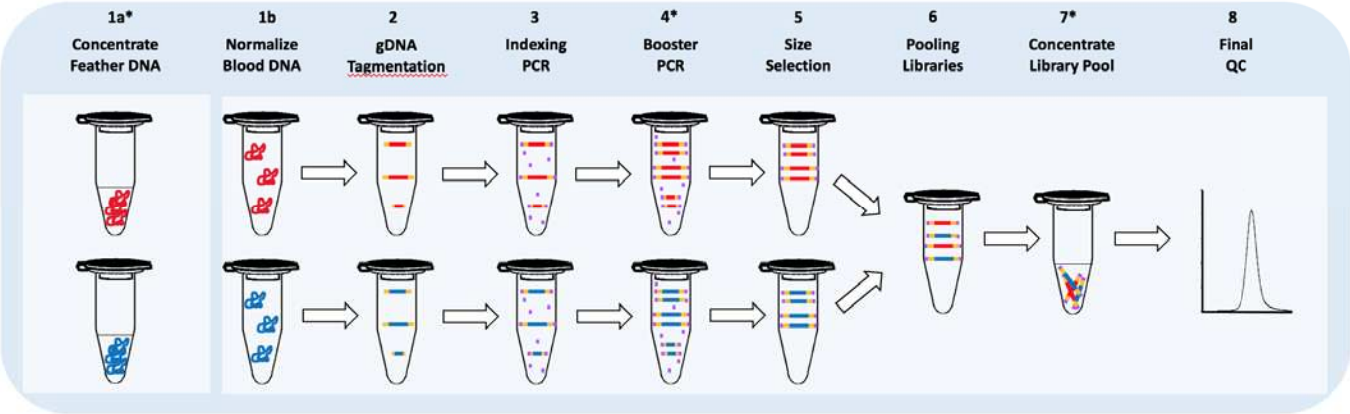

Figure 1. Lab workflow diagram for LI-Seq method. Steps 1 through 8 follow full protocol (see Supplementals). Steps with * after number denote steps where modifications to typical WGS library preparation methods were implemented. Low input samples start with step 1a and high input samples start with step 1b.

To ensure that final library concentrations will be similar across samples, we quantified each DNA extraction using a Qubit dsDNA High Sensitivity Assay Kit (Invitrogen) and normalized each sample to a concentration of 0.48 ng/µL- 4.5 ng/µL, with a target of 2.5 ng/µL (Figure 1, step 1a and 1b). To fragment the DNA and "tag" it with Nextera adapters, we added 2.50 µL of TD Buffer and 0.5 µL of TDE1 Enzyme (Illumina) to 1 µL of normalized DNA and incubated the samples in a thermocycler at 55 ^{o}C for 20 minutes (Figure 1, step 2).

To amplify the tagmented DNA and add Nextera indexing adapters for sequencing, we pipetted 1 µL of each index primer into the appropriate well of tagmented DNA until all samples had a unique pair of dual indexes. We then added 6.0 µL of Kapa Hifi Hotstart Mix (KMM; Kapa Biosystems) before running in a thermocycler as follows: held at 72° for 3 minutes, held at 98° for 2 minutes and 45 seconds, cycled 8 times through 98° for 15 seconds, then 62° for 30 seconds, then 72° for 30 seconds, held at 72° for 1 minute, and then held at 4° until removed from thermocycler (Figure 1, step 3). As per the Therkildsen and Palmubi (2017) method, this indexing PCR had more cycles than the original Illumina protocol and was broken into two stages (Indexing PCR and "Reconditioning PCR, " which was renamed Booster PCR in this method). Additional cycles were added in the Therkildsen and Palumbi (2017) method because the tagmented DNA was not purified prior to Indexing PCR, making the PCR reaction less efficient. To further amplify, or "boost," copies of indexed DNA without using additional Nextera indices, we added 7.6 µL of KMM, 4.4 µL of ultrapure water, and 1.6 µL each of a custom 10uM primer pair (P1=AATGATACGGCGACCACCGA; P2=CAAGCAGAAGACGGCATACGA) to each library. We ran the samples in a thermocycler as follows: held 95° for 5 minutes, cycle 4 times through 98° for 20 seconds, then 62° for 20 seconds, then 72° for 2 minutes, hold at 72° for 2 minutes, and then held at 4° until removed from the thermocycler (Figure 1, step 4).

To purify the PCR product and remove undesirable fragments, we followed standard Ampure bead protocol (Beckman Coulter) using a 0.7:1 bead to DNA ratio which will remove below approximately 320 bp and eluted into 30 µL of 10mM Tris-Hcl (Figure 1, step 5). In order to

avoid overrepresentation of one individual during whole genome resequencing, we then quantified using a Qubit dsDNA High Sensitivity Assay Kit (Invitrogen) and pooled an equal number of copies of each sample into a 1.5 mL tube (Figure 1, step 6). Finally, in order to increase the final concentration of the pooled libraries and increase sequencing efficiency, we then followed the standard Ampure double-sided size selection protocol, using a 0.63:1 bead to DNA ratio to remove large fragments and a 0.73:1 bead to DNA ratio to remove small fragments, and eluted into 30 µL of 10mM Tris-HCl (Figure 1, step 7). After the pooled library has been concentrated and double size selected (either with or without the optional reconditioning PCR), we perform final quality control (QC) with Qubit quantification and Tapestation 2200 fragment distribution analysis (Agilent) (Figure 1, step 8).

To address issues of overamplification, also called 'PCR bubble,' we encountered while using Therkildsen and Palumbi's original method (2017) with our low input DNA, we added an optional 'reconditioning PCR' step which provides additional reagents, especially primers, so that the PCR product does not anneal to itself (Thompson et al 2002). To recondition a final pooled library with overamplification, we added 12µL of pooled, size selected library and added 7.6 µL of KMM, 1.6 µL of 10uM P1 (AATGATACGGCGACCACCGA), 1.6 µL of 10uM P2 (CAAGCAGAAGACGGCATACGA), and 4.4 µL of ultrapure water. We then ran it in a thermocycler as follows: held 95° for 5 minutes, cycled once through 98° for 20 seconds, then 62° for 20 seconds, then 72° for 2 minutes, held at 72° for 2 minutes, and then held at 4° until removed from the thermocycler. Next, we used Ampure beads to clean up and performed an additional

181 double size selection and bead cleanup before quantifying and running the library through a
182 Tapestation 2200 fragment distribution analyzer (Agilent) for final quality control. This
183 reconditioning PCR is optional and may not be required for all library preparations.
184
185 Using the above method, we prepared two WGS libraries from American Redstart (*Setophaga*
186 *ruticilla*) samples for low (2x) coverage sequencing. One library was prepared with 50 unique
187 blood samples of normalized DNA concentrations between 1.18 - 4.78 ng/uL. The other library
188 was prepared with 50 unique feather samples, extracted from a single feather calamus, with
189 starting DNA concentrations of 0.48 - 5.7 ng/µL. Both libraries had one cycle of reconditioning
190 PCR performed on the final library. Libraries were each sequenced on one full 2 x 150 bp PE
191 (paired end) HiSeq 4000 lane (Illumina).
192
193 Bioinformatic Analysis and Quality Checking
194
195 We trimmed the sequence data to remove potential PCR artifacts using the program FastUniq
196 version 0.11.9 (Xu et al 2012). PCR duplicates need to be removed in order to ensure high-
197 quality sequence data in downstream processes such as creating scaffolds in whole-genome
198 sequencing. We mapped reads to an assembled genome of the yellow warbler (*Setophaga*
199 *petechia*; Bay et al 2018), using the Burrows-Wheeler Aligner software version 0.7.17 (Li and
200 Durbin 2010). The resulting SAM files were sorted, converted to BAM files, and then indexed
201 using samtools version 1.9 (Li et al 2009). Depth of sequencing coverage at every read position

was calculated using the depth function in samtools (Li et al 2009). The quality of the BAM files for the two different libraries was assessed by comparing the average read depth by individual as well as the average read depth by scaffold. We quantified the GC content of 100 base pair windows in the BAM files from the two libraries using CollectGcBiasMetrics function in Picard version 2.23.1 (Broad Institute 2019). We determined the proportion of reads that passed quality filters for the two libraries using CollectWgsMetrics in Picard (Broad Institute 2019). Two-tailed t-tests were used to compare the quality diagnostics for the different libraries and were implemented in R version 3.6.2 (R Core Team 2019).

Results

Library Preparation for Sequencing

A comparison between libraries prepared using the method of Therkildsen and Palumbi (2017) and our modified method revealed that doubling the ratio of tagmentation enzyme to DNA, increasing the tagmentation time to 20 minutes, and decreasing the indexing PCR elongation time to 30 seconds resulted in maximization of fragments in the target distribution (Table 1; Figure 2).

Table 1. Matrix of the original (Therkildsen) protocol parameters and final LI-Seq parameters, tested with 3 high input samples (blood).

Test Name	total DNA (ng)	enzyme (ul)	buffer (ul)	total rxn vol (ul)	tagmentation time (min)	Index Elongation Time (sec)	Sample Number	Final Conc (ng/ul)	Avg fragment length (bp)
Original	2.5	0.25	1.25	2.5	5	180	A	8	663

							B	8.54	438
							C	8.3	439
LI-Seq	2.5	0.5	2.5	4	20	30	A	2.79	386
							B	2.69	309
							C	2.39	397

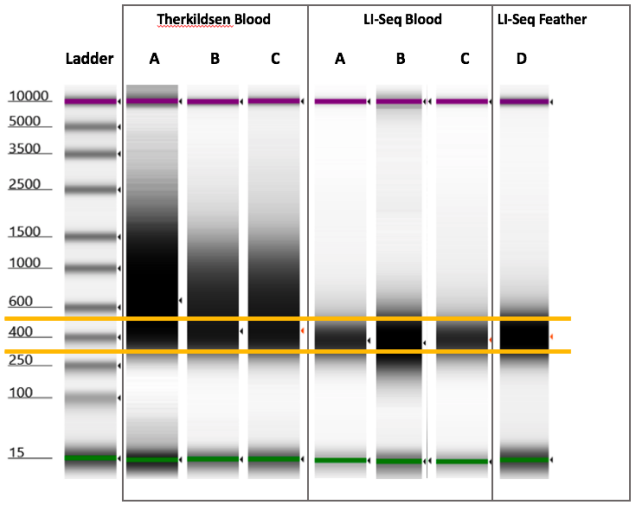

Figure 2. Tapestation 2200 gel of Therkildsen protocol conditions and final LI-Seq conditions from Table 1. Yellow bands show preferred fragment range (320-500bp) for HiSeq 4000 as recommended by Novogene. Individuals A, B and C were duplicated between the Therkildsen Blood and LI-Seq Blood, with the conditions described in Table 1. Individual D was feather DNA prepared using only the LI-Seq method.

Using these modified methods, we successfully prepared two libraries from American Redstart DNA. The library from 50 high input samples (blood) had a final concentration of 12.1 ng/µL, a molarity of 39.3 nM, and an average library size of 466 bp (Figure 3d). The library from 50 low input samples (feathers) had a final concentration of 2.56 ng/µL, a molarity of 8.56 nM, and an average library size of 453 bp (Figure 3f).

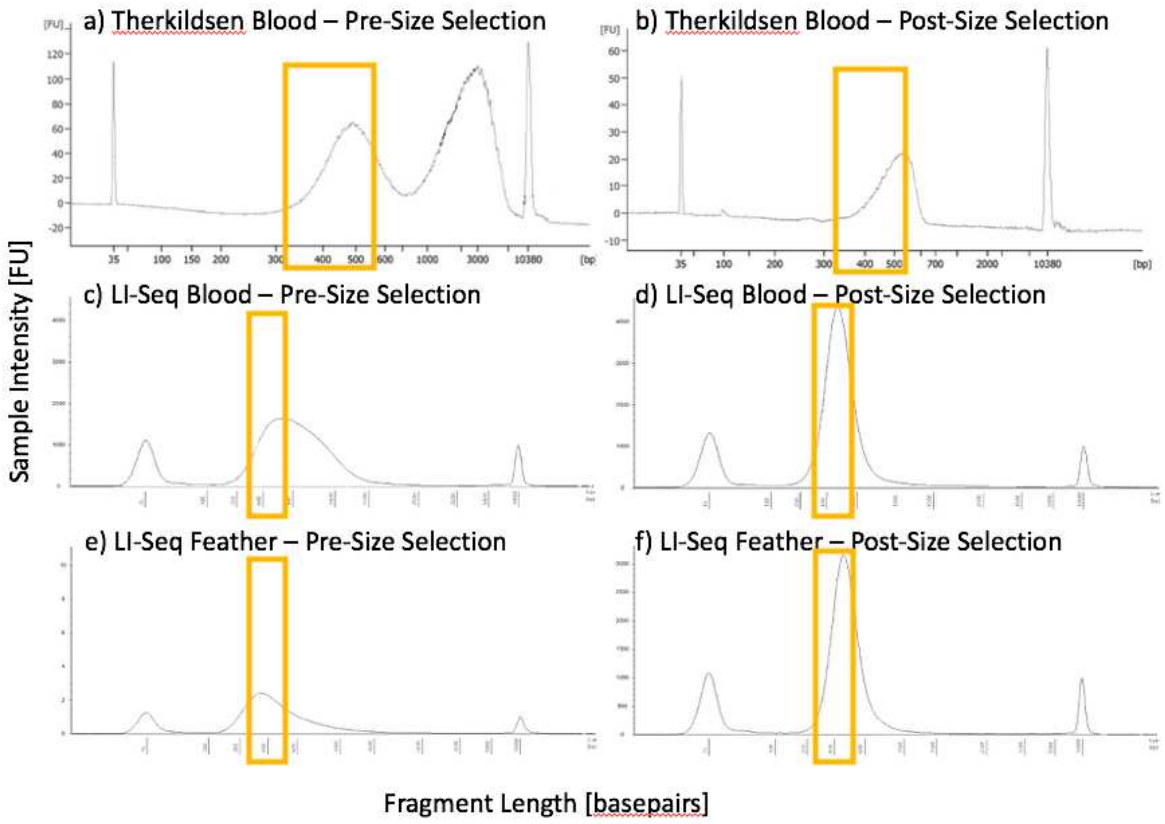

Figure 3. Final library fragment analysis of a library prepared using the original Therkildsen and Palumbi protocol (2017), before (a) and after (b) size selection, run on a Bioanalyzer (Agilent). The second peak of (a) indicated overamplification. The final libraries of the high input (blood) and low input (feather) samples were run on a Tapestation 2200 (Agilent) prior to size selection (high input (c) and low input (e)) and after size selection (high input (d), low input (f)). The x-axis is the size of the DNA fragment in basepairs, and the y-axis is the sample intensity which is correlated with sample concentration. Note the different x- and y-axis scales. Yellow boxes indicate the desired fragment size of ~320 – 500 bps.

Bioinformatic Analysis and Quality Checking

The depths of coverage of the sequence data from the two libraries were not significantly different (t = 1.06, df = 98, p-value = 0.29), but the high input library had a slightly higher depth (mean 1.79, standard deviation 0.52; Figure 3a) than the low input library (mean 1.70, standard deviation 0.38; Figure 3a). The high input library had individuals that had a slightly higher

proportion of the genome with sequence data (mean 0.66, standard deviation 0.06; Figure 3b) than the low input library (mean 0.64, standard deviation 0.08; Figure 3b), but this difference was not statistically significant (t = 1.85, df = 98, p-value = 0.07). The difference in GC distribution between the two libraries was non-significant (42.9%; t = 1.94e-14, df = 100, p-value = 1), with the mean GC content of the high input library being 42.7% and 42.9% for the low input library (Figure 4C). The sequence data from these two libraries also had very similar patterns of coverage across the scaffolds in the genome (Figure 5). The similarity in coverage across scaffolds shows that certain genomic regions are not being over- or under-amplified in either of the libraries due to a difference in quality and quantity of input DNA. All BAM quality metrics for the two libraries produced comparable results (Table 2).

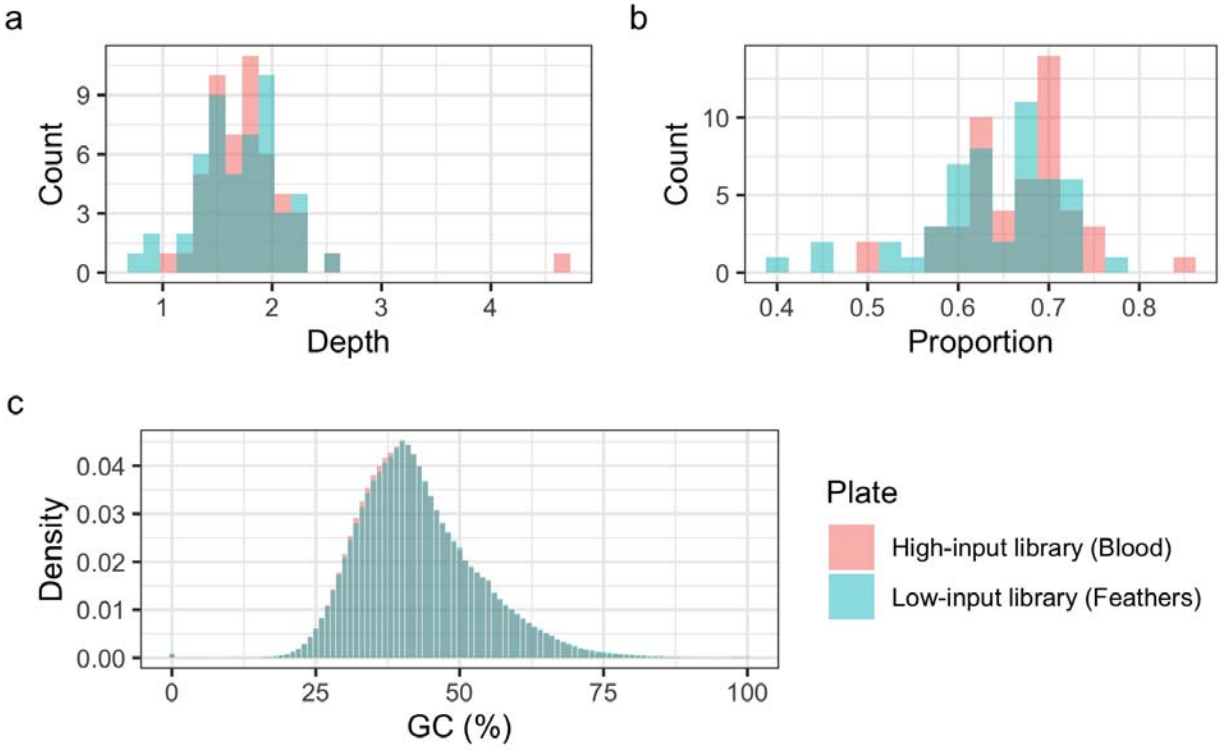

Figure 4. Both histograms of (a) average depth across the genome by individual and (b) proportion of genome with sequence data by individual overlapped considerably between the two libraries and neither measures were significantly different (depth: t = 1.06, df = 98, p-value = 0.29; proportion: t = 1.85, df = 98, p-value = 0.07). GC content was determined for 100 base pair region windows in each individual and then averaged across individuals for each of the two plates. (c) The bins of GC content (ranging from 0 to 100%) were nearly identical for the two libraries (t = 1.94e-14, df = 100, p-value = 1). Data from the high input library from blood samples are shown in red, the low input feather library in light blue, and the overlapping data in darker blue.

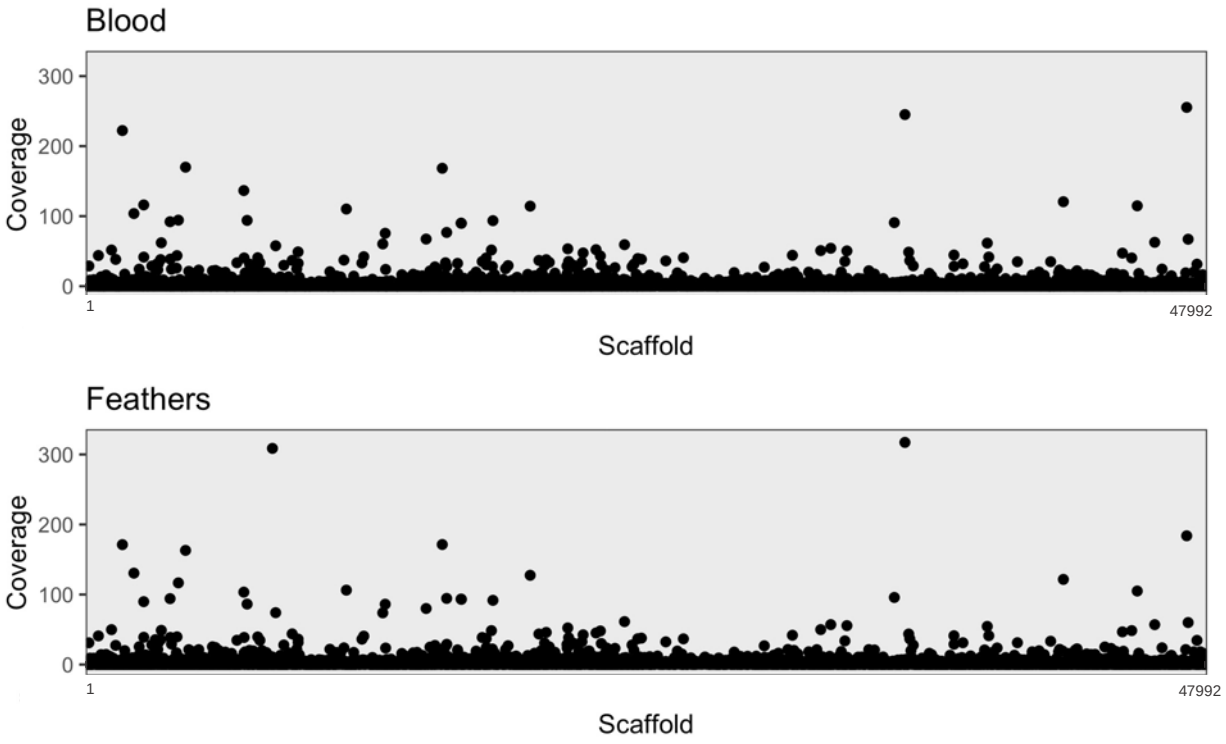

Figure 5. Mean coverage by scaffold averaged across all individuals for the two libraries reveal similar patterns in coverage across the genome between the two libraries.

Table 2. BAM quality statistics: the average proportion of reads with adaptors (Adaptors), mapping quality less than 20 (MapQ), marked as duplicates (Duplicate), without a mapped mate pair (Unpaired), quality score less than 20 (BaseQ), and at least 1x (1x) and 5x (5x) coverage after removing low quality reads.

Library	Adaptors	MapQ	Duplicate	Unpaired	BaseQ	1x	5x

High input (Blood)	0	0.110	0	0.001	0.017	0.685	0.049
Low input (Feathers)	0	0.117	0	0.001	0.015	0.657	0.039

Discussion

In conservation genetics, researchers must often work from low quality and low input DNA samples, especially in the case of noninvasive or minimally invasive sampling. These types of samples can be a challenge to use with many genomic techniques, such as WGS, which typically require high input and quality input DNA. Yet, collecting high input samples (like blood and tissue) can present ethical, logistical, and financial roadblocks.

The potential knowledge gained from utilizing WGS and similar genomic tools can provide insights not yet achievable from other methods. WGS can be a powerful tool for management in conservation efforts but, until now, has been challenging and or cost-prohibitive when working with low input, low quality DNA. Here we have described LI-Seq, a method for cost-effective WGS library preparation that can be used for both high quality, high input and low quality, low input samples. Results suggest that, with the modifications described above, one can successfully produce high quality sequence data from DNA with input as low as 0.48 ng, for a fraction of the cost of traditional library preparation methods. More specifically with this method, we were able to prepare approximately 12 libraries for the same price as a single

293 library with a more traditional WGS library preparation kit, thus allowing us to sequence more
294 than an order of magnitude more samples. Overall, the increase in efficiency and cost-
295 effectiveness provided by our method will allow conservation biologists to more broadly apply
296 WGS methods to samples collected with non-invasive or minimally invasive methods.
297
298 Another potential challenge of using low input DNA with WGS methods is biased or incomplete
299 amplification of the genome due to the low number of copies of the entire genome present
300 (Meynert et al 2014). To assess this, we prepared a library from high quality, high input DNA
301 (extracted from blood) in addition to preparing a library from low quality, low input DNA
302 (extracted from feathers) using the same protocol. The final library molarity and fragment
303 distribution was similar with both DNA sources, suggesting that the lower input DNA yielded
304 equally high-quality libraries as high input DNA. In addition to impacting the quality of the final
305 library, low input DNA could result in biased amplification leading to preferential sequencing of
306 certain regions of the genome and result in less equal coverage across the genome than with
307 high input DNA. The observed correspondence in patterns of genome-wide coverage between
308 the libraries is to be expected in the absence of external influences causing biases related to the
309 quality of input DNA (Ekblom, Smeds, and Ellegren 2014). Additionally, we checked for
310 differential GC bias between the two libraries because it can indicate that biases were
311 introduced during library preparation (Sims et al 2014). Our breadth of genome coverage (i.e.
312 proportion of the genome sequenced) of 64% and 66% for the two libraries (high input and low
313 input respectively) is comparable to that of mammalian genomes sequenced at similar depth

(Green 2007). This breadth of coverage is suitable for numerous conservation genomic applications (e.g. identifying inbreeding across the genome) and resource efficient considering that approximately 30x coverage depth is required to achieve 95% coverage breadth (Sims et al 2014). Overall, our results suggest that when our method is employed, libraries prepared from low input and high input DNA both produce high quality sequence data.

When applying this protocol to other species and sample types, researchers may want to think about a couple important considerations. First, the genome size of the study organism should be considered. The protocol presented herein is an excellent option for use with organisms with a small genome, as the protocol was optimized using DNA from birds which on average have a genome size of 1.1 Gb. For species with larger genomes, more sequencing is required to achieve the same level of coverage, and therefore may require optimized methods and will have a relatively higher cost as well. Additionally, the ratio of tagmentation enzyme to input DNA may need to be adjusted in order to maintain a similar average fragment size. The tagmentation enzyme is one of the most expensive components of this method, so increasing the amount of enzyme per sample could also increase the per sample cost. Second, when considering combining libraries from low quality, low input and high quality, high input samples into a single sequencing run, researchers may want to take additional steps to ensure equal representation of the libraries. Specifically, performing a double size selection on the individual libraries prior to quantification and pooling can help ensure that the most accurate concentrations are used when pooling individuals for sequencing and, therefore, the most equal sequencing effort per

sample will occur (Zamudio et al unpublished). Overall, with the aforementioned modifications taken into account, the method presented here could be applied more broadly to increase the efficiency and cost-effectiveness of WGS across a multitude of taxa.

Here we present a cost-effective method for producing WGS libraries using low input DNA from minimally-invasively collected samples. LI-seq provides a much-needed tool to bridge the gap between the conservation management applications of WGS data and frequently collected sample types, such as feathers and other non-invasively collected samples. Although a recent method was published that also provided a method for preparing non-invasively collected samples for WGS (Taylor et al 2020), the per sample cost may be prohibitive for use with population-scale studies for conservation efforts. By providing an efficient, cost-effective WGS method for low quantity and quality DNA samples, we hope conservation management efforts will be able to better take advantage of the applications WGS can provide for enhancing management efforts.

Acknowledgements

Thank you to Dr. Nina Therkildsen fo sharing her WGS method, Harmony B. Borchardt-Wier for early guidance on trouble-shooting Dr. Therkildsen's protocol, and Dr. Cristian Gruppi for his technical advising. We thank the UC Davis Genome Center for their help with the sequencing. This work was made possible by an NSF CAREER award (008933-00002), an NSF grant Rules of Life grant (007604-00002), and a National Geographic grant (WW-202R-170) to K. Ruegg as well as the Extreme Science and Engineering Discovery Environment (XSEDE) computer cluster, which is supported by National Science Foundation grant ACI-1548562.

References

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature reviews genetics*, *11*(10), 697-709.

Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., & Kishony, R. (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. *PloS one*, *10*(5), e0128036.

Bay, R. A., Harrigan, R. J., Le Underwood, V., Gibbs, H. L., Smith, T. B., & Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, *359*(6371), 83-86.

Carroll, E.L., Bruford, M.W., DeWoody, J.A., Leroy, G., Strand, A., Waits, L., & Wang, J. (2018). Genetic and genomic monitoring with minimally invasive sampling methods. *Evolutionary applications*, *11*(7), pp.1094-1119.

Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, *107*(1), 1-15.

Ekblom, R., Smeds, L., & Ellegren, H. (2014). Patterns of sequencing coverage bias revealed by ultra-deep sequencing of vertebrate mitochondria. *BMC genomics*, *15*(1), 467.

Faircloth, B. C., & Glenn, T. C. (2014). Protocol: Preparation of an AMPure XP substitute (AKA Serapure). doi: 10.6079. *J9MW2F26*.

Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular ecology*, *26*(20), 5369-5406.

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in ecology & evolution*, *27*(9), 489-496.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*(6), 333.

Green, P. (2007). 2× genomes—Does depth matter?. *Genome research*, *17*(11), 1547-1549.

Kohn, M. H., Murphy, W. J., Ostrander, E. A., & Wayne, R. K. (2006). Genomics and conservation genetics. *Trends in ecology & evolution*, *21*(11), 629-637.

Kryazhimskiy, S., Rice, D. P., Jerison, E. R., & Desai, M. M. (2014). Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science*, *344*(6191), 1519-1522.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*(5), 589-595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

Lukacs, P. M., & Burnham, K. P. (2005). Review of capture–recapture methods applicable to noninvasive genetic sampling. *Molecular ecology*, *14*(13), 3909-3919.

Meynert, A.M., Ansari, M., FitzPatrick, D.R., & Taylor, M.S. (2014). Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC bioinformatics*, *15*(1), pp.1-11.

Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R. K., & Hedrick, P. W. (2010). Conservation genetics in transition to conservation genomics. *Trends in genetics*, *26*(4), 177-187.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Regnaut, S., Lucas, F. S., & Fumagalli, L. (2006). DNA degradation in avian faecal samples and feasibility of non-invasive genetic studies of threatened capercaillie populations. *Conservation Genetics*, *7*(3), 449-453.

Roques, S., Furtado, M., Jácomo, A. T., Silveira, L., Sollmann, R., Torres, N. M., ... & Palomares, F. (2014). Monitoring jaguar populations Panthera onca with non-invasive genetics: a pilot study in Brazilian ecosystems. *Oryx*, *48*(3), 361-369.

Ruegg, K., Bay, R. A., Anderson, E. C., Saracco, J. F., Harrigan, R. J., Whitfield, M., ... & Smith, T. B. (2018). Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. *Ecology letters*, *21*(7), 1085-1096.

Russello, M. A., Waterhouse, M. D., Etter, P. D., & Johnson, E. A. (2015). From promise to practice: pairing non-invasive sampling with genomics in conservation. *PeerJ*, *3*, e1106.

Ryder, O. A. (2005). Conservation genomics: applying whole genome studies to species conservation efforts. *Cytogenetic and genome research*, *108*(1-3), 6-15.

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, *15*(2), 121-132.

Stenglein, J. L., Waits, L. P., Ausband, D. E., Zager, P., & Mack, C. M. (2010). Efficient, noninvasive genetic sampling for monitoring reintroduced wolves. *The Journal of Wildlife Management*, *74*(5), 1050-1058.

Taberlet, P., & Bouvet, J. (1992). Bear conservation genetics. *Nature*, *358*(6383), 197-197.

Taylor, R., Manseau, M., Request, B., & Wilson, P. (2020). Whole genome sequences from non-invasively collected samples. *Molecular ecology resources,* https://doi.org/10.22541/au.158809437.78730399

Therkildsen, N. O., & Palumbi, S. R. (2017). Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. *Molecular ecology resources*, *17*(2), 194-208.

Thompson, J. R., Marcelino, L. A., & Polz, M. F. (2002). Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Research*, *30*(9), 2083-2088.

Valière, N., Fumagalli, L., Gielly, L., Miquel, C., Lequette, B., Poulle, M. L., ... & Taberlet, P. (2003). Long-distance wolf recolonization of France and Switzerland inferred from non-invasive genetic sampling over a period of 10 years. *Animal Conservation*, *6*(1), 83-92.

Waits, L. P., & Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *The Journal of Wildlife Management*, *69*(4), 1419-1433.

Weissenbach, J. (2016). The rise of genomics. *Comptes rendus biologies*, *339*(7-8), 231-239.

Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., ... & Chen, S. (2012). FastUniq: a fast de novo duplicates removal tool for paired short reads. *PloS one*, *7*(12), e52249.

Author Contributions

TMS and KCR designed the experiments. TMS carried out the experiments. MGD and KCR analyzed data. TMS, MGD, and KCR wrote the manuscript.