EFFECTIVENESS OF REDUCED REPRESENTATION SEQUENCING ON CENTURY-OLD, ETHANOL-PRESERVED MUSEUM FISHES

A Thesis

by

MARTIN G. FRENCH

BS, Texas A&M University Corpus Christi, 2016

Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in

BIOLOGY

Texas A&M University-Corpus Christi
Corpus Christi, Texas

May 2021

May 2021

EFFECTIVENESS OF REDUCED REPRESENTATION SEQUENCING ON CENTURY-OLD, ETHANOL-PRESERVED MUSEUM FISHES

A Thesis

by

MARTIN G. FRENCH

This thesis meets the standards for scope and quality of
Texas A&M University-Corpus Christi and is hereby approved.

Dr. Christopher E. Bird, PhD
Chair

Dr. David Portnoy, PhD
Committee Member

Dr. J. Derek Hogan, PhD
Committee Member

May 2021

ABSTRACT

Museum specimens have a largely underutilized potential to allow biologists to study rare, ancient, or extinct organisms using genomic methods. However, museum samples often have degraded and fragmented DNA making it more difficult to sequence. Reduced representation sequencing has proven to be affordable and effective for population genomic applications but is sensitive to the degradation inherent with museum samples. Here, sequence quality and error rates were compared between reduced representation libraries constructed from century-old, ethanol-preserved museum and contemporary samples for two fishes (*Atherinomorus duodecimalis* and *Siganus spinus*), with a focus on the barcoded adapter and SbfI restriction site expected to occur at the beginning of every sequence read due to the library preparation. Museum specimens had a larger proportion of reads filtered due to adapter dimer and low base call quality, while yielding a smaller proportion of reads with the expected adapter sequence. Elevated error rates in the adapter (synthetic sequence) and the last two positions of the restriction site (fish sequence, positions 7 & 8) of museum samples indicates that the specificity of both the DNA polymerase and restriction enzyme, respectively, was impaired by a contaminant. Errors in the last two positions of the restriction site were not independent, indicating that if the restriction enzyme misrecognized position 7, then it also misrecognized position 8. Overall, sequencing of degraded museum specimens preserved in EtOH for >100 years is possible, but all else being equal, it can result in more sequence substitution errors, unintended loci and decreased depth of coverage due to altered enzymatic activity during library preparation when compared to contemporary samples. Consequently, up to 24% more DNA per museum specimen needs to be sequenced to achieve comparable results to contemporary fish.

ACKNOWLEDGEMENTS

This work was funded by the United States National Science Foundation (NSF-OISE-PIRE-1743711), the data was processed on the TAMU-CC high performance computing cluster funded by NSF-MRI-CNS-0821475, and my salary was funded by the TAMU-CC Genomics Core Laboratory. I would like to thank the Albatross 1907-1910 Philippine expedition team including the captain and crew of R/V Albatross I, the expedition leader Hugh M. Smith; the lead scientist F.M. Chamberlain; and the scientific party H.C. Fasset, Lewis L. Radcliffe, Paul Bartch, Albert L. Barrows, Alvin Seale, and Roy Chapman Andrews. I would also like to thank the Smithsonian Institute for maintaining the 1907-1910 Philippine Albatross sample collection; Dr. Jeff Williams and the Smithsonian for providing access to the 1907-1910 Philippine Albatross sample collection. Contemporary samples were collected under Gratuitous Permit No. 0166-18 by Dr. Kent Carpenter, Dr. Jeff Williams in collaboration with Dr. Angel Alcala and Abner Bucol of Silliman University and Dr. Mudjeekewis Santos of the National Fisheries Research and Development Institute. Madeleine Kenton, John Whalen, Jem Baldisimo, and Ivan Lopez dissected the museum specimens and extracted the DNA used in this study. Sharon Magnuson, Maili Allende and the TAMUCC Genomics Core Lab processed samples and prepared libraries. Dr. Eric Garcia assisted with data filtering and Jason Selwyn assisted with error rate modeling and statistical analysis. Dr. J. Derek Hogan and Dr. David Portnoy reviewed this manuscript and provided valuable feedback that improved this thesis. Dr. Christopher E. Bird provided instruction and guidance conducting this research and assembling this thesis. Additional thanks to James. G. French, Mary. T. French, George. W. French, Margaret. E. French, Cdr. E. M. S. Windridge, and Norah. S. Windridge for their motivation and support.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES PAGE

LIST OF TABLES

TABLES PAGE

INTRODUCTION

Museum specimens are invaluable assets to the field of biology and genetics because they offer the opportunity to study organisms from distant locations and time periods (Gilbert, Moore, Melchior, & Worobey, 2007; Green et al., 2010; Meyer et al., 2012; Naumann, Krzewińska, Götherström, & Eriksson, 2014; Verdugo, Kassadjikova, Washburn, Harkins, & Fehren-Schmitz, 2016). Studies involving museum samples allow us the ability to observe how allele and haplotype frequencies change over time and test evolutionary hypotheses on rare or extinct species (Su, Wang Y., Lan, Wang W., & Zhang, 1999; Park et al., 2015). The majority of museum specimens, however, were not originally preserved with genetic analysis in mind and are almost always characterized by DNA degradation. Sample collection, preservation media, storage, age, and tissue type can all affect the quality of DNA obtained from museum specimens.

The DNA in museum and ancient specimens can be fragmented (Tin, Economo, & Mikheyev, 2014), cross-linked with the cellular matrix (Miething, Hering, Hanschke, & Dressler, 2006; Wong et al., 2014), and may experience changes in base composition (Pääbo et al., 2004), all of which make them challenging to sequence. Fragmentation of the DNA can be caused by exonuclease activity in the dying tissue prior to preservation (Lindahl, 1993), depurination and hydrolytic damage caused by the preservation media (Overballe-Petersen, Orlando, & Willerslev, 2012). While ethanol does not prevent degradation by nuclease activity, it is one of the most effective solutions for the preservation of DNA (Post, Flook, & Millest, 1993). Cross-linking is a process where DNA forms chemical bonds with other molecules in the cell, making it nearly impossible to isolate DNA from other cellular components (Tretyakova, Groehler, & Ji, 2015). Cross-linking can be caused by preservation in formalin, ionizing g radiation, UV light, platinum compounds, and exposure to various other physical and chemical agents (Wiegand,

Domhöver, & Brinkmann, 1996; Hykin, Bi, & McGuire, 2015; Tretyakova et al., 2015). Base substitutions can be caused by hydrolytic damage which occurs after the death of the tissue and while in storage (Pääbo et al., 2004; Cooper, Drummond, & Willerslev, 2004). The most common forms of hydrolytic damage are deamination (removal of an amine group) of cytosine into uracil or adenine to hypoxyanthine. When the deaminated bases are sequenced, uracil is interpreted as thymine (C → T) and the hypoxyanthine is interpreted as guanine (A → G). Hydrolytic deamination accumulates over time since the death of the tissue and is less common in contemporary specimens than ancient DNA (Pääbo, 1989; Sawyer, Krause, Guschanski, Savolainen, & Pääbo, 2012). Overall, cross-linking and fragmentation are the most prominent concerns with museum specimens, but there has been some success in sequencing these types of samples, including high-throughput methodologies (Burrell, Disotell, & Bergey, 2015).

To reduce the cost per individual in contemporary population genomic studies of both model and non-model species, researchers often sample a subset of the genome using reduced representation sequencing (Altshuler et al., 2000; Hohenlohe, Bassham, Etter, Stiffler, Johnson, & Cresko, 2010; Baxter et al., 2011; Andrews, Good, Miller, Luikart, & Hohenlohe, 2016). By reducing the amount of DNA sequenced per individual, it is possible to increase the number of individuals sequenced. Methods such as restriction-site-associated DNA sequencing (RADseq), reduce genome complexity by focusing sequencing efforts on a portion of the genome (Ali et al., 2015).

Restriction-site-associated DNA sequencing is a “short-read”, reduced-representation sequencing method which employs restriction enzymes to cut DNA and a high-throughput Illumina sequencer which produces 150 bp sequence reads. To identify and organize restriction-site digested DNA, synthesized oligonucleotides with identifying barcodes are attached via

ligation to every restriction site where a cut occurs. This method has proven effective for the simultaneous identification of polymorphic regions and for genotyping in both model and non-model organisms (Miller, Dunham, Amores, Cresko, & Johnson, 2007; Toonen et al., 2013). Although the laboratory techniques in RADseq are not a new, its potential applications have been greatly expanded by next generation sequencing. The drastic reduction in the cost brought by next generation sequencing has made research in population genetics, quantitative trait mapping, comparative genomics, and phylogeography possible using ddddRADseq (Baird et al., 2008; Etter, Bassham, Hohenlohe, Johnson, & Cresko, 2011).

One of the major limitations of RADseq is its sensitivity to degraded DNA, such as may be associated with museum specimens (Rowe, Renaut, & Guggisberg, 2011; Burrell et al., 2015). Degraded DNA is associated with inconsistent digestion by restriction enzymes, increased downstream variance in the number of sequence reads per locus and individual (coverage), and/or missing data (Zimmermann et al., 2008; Graham et al., 2015). Cross-linking in degraded samples will reduce the amount of DNA that can be extracted from the tissue and the RADseq family of techniques are dependent upon a relatively large amount of high molecular weight DNA (50-150ng / sample). Restriction enzymes recognize, bind to and cleave DNA at specific nucleotide sequences (Woodbury, Hagenbüchle, & von Hippel, 1980), called recognition sites, which can be affected by base-substitutions, causing viable restriction sites to go unrecognized and uncut by enzymatic activity. Alternatively, non-target sequences that undergo base-substitutions could be recognized and cut erroneously by restriction enzymes.

Recent studies have shown RAD-tagging and shotgun genome sequencing to be effective for desiccated museum specimens (Tin et al., 2014; Burrell et al., 2015). However, these studies had shallow depth of coverage (0.37-3 mean reads per base) and could be affected by base

substitutions. Illumina recommends a coverage depth of 30 reads per base, which it equates to a 0.999 probability of making a correct base call (Illumina, 2010). In addition to poor sequence coverage, increased base substitutions and reduced enzyme specificity were also observed in museum samples (Bi et al., 2013; Burrell et al., 2015).

While the mechanisms by which the sequencing of museum specimens can be affected by DNA alterations occurring "in the jar" have received much attention, relatively little attention has been focused on the effects of chemical and protein contaminants associated with historically preserved and maintained specimens on the enzymatic reactions employed in DNA sequencing, particularly in ethanol-preserved specimens. The purity of ethanol used in preservation may vary from source to source, especially if obtained from spirits distilleries (Riachi et al., 2014) as is likely the case for older collections from remote locations. Alcoholic beverages, and presumably higher proof distillates, can contain metals from contaminated water sources, volatile byproducts such as acetaldehyde, urethane, and methanol (Rehm et al., 2010) which may affect enzymatic activity. It is also possible for museum specimens to have been fixed in formalin without it being recorded, and formalin fixation is associated with PCR inhibition (An & Fleming, 1991). Polymerase chain reaction can be inhibited by an array of contaminants, including salts, polysaccharides, proteins, and ethanol (reviewed in Bessetti 2007). The specificity of restriction enzymes can be negatively affected by organic solvents (Malyguine et al., 1980), a high ratio of enzyme to DNA (Bitinaite & Schildkraut, 2002), a non-optimal buffer (Nasri & Thomas, 1987), and the substitution of Mg^{2+} with other divalent cations (New England Biolabs, 2021). Other enzymatic reactions in DNA sequencing that may be affected by contaminants are ligation and blunting.

Here we construct and sequence RAD libraries from century-old, ethanol-preserved fishes collected in the Philippines and contemporary re-collections. We tested for differences between the historical and contemporary specimens for sequence read yield and error rates in the barcoded adapter and SbfI restriction site expected to be at the beginning of each sequence, which inductively indicate the presence of an interfering contaminant. These tests were used to infer which steps of the library preparation protocol were negatively affected in the museum specimens (Fig. 1). Particular attention was given to the enzymatic steps of restriction digest and PCR which are known to be sensitive to contaminants and are critical to obtaining consistent, reliable data.

METHODS

Sample Collection & Preservation

Samples used in this study consisted of two species of fishes collected in the Philippines, *Atherinomorus duodecimalis* and *Siganus spinus*. Museum specimens (6 *A. duodecimalis*, 14 *S. spinus*) were collected during the 1907 to 1910 Philippine Expedition of the research vessel Albatross I (NOAA, 2021). While it is immaterial to the hypotheses tested in this paper, the *A. duodecimalis* specimens were from Tawi and Matnog Bay, and the *S. spinus* specimens were from Nato. While on board the R/V Albatross, whole fish were stored in high proof rum distillate obtained in the Philippines, before being transferred to 75% ethanol for storage at the Smithsonian Institute, Washington, D.C. There is no record of formalin fixation for the museum samples, and the preservation fluid in the museum lots had a pH of ~ 6.1-8.3, suggesting variable preservation conditions. Fin clips from contemporary specimens (6 *A. duodecimalis, 24 S. spinus*) were collected from 2017 to 2018 at Philippine fish markets and stored in laboratory.

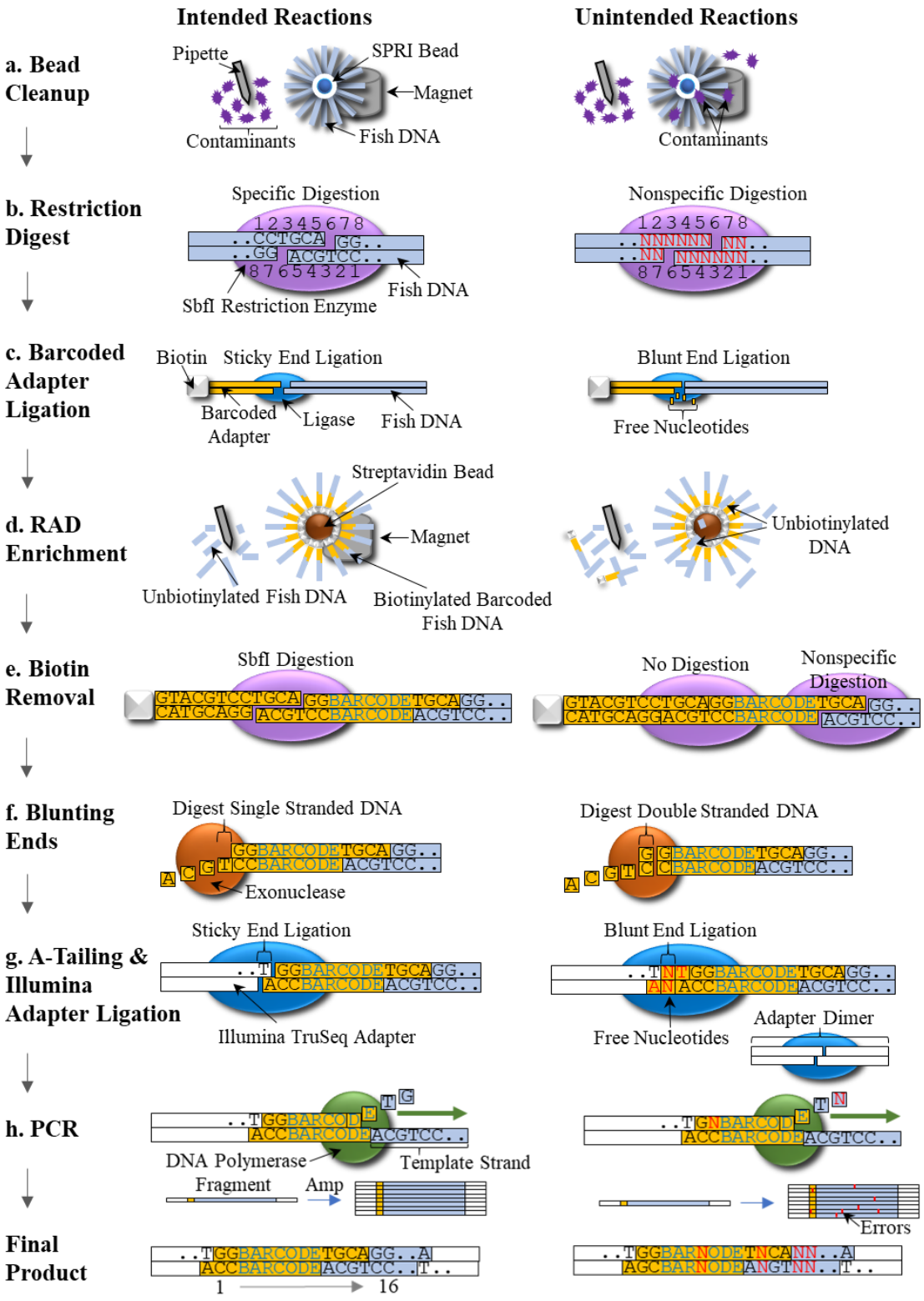

Figure 1. RAD library preparation protocol and potential unintended reactions. Fragmentation of DNA via sonication between steps c and d is not shown.

grade 95% ethanol (see Carpenter, Williams, & Santos, 2017). *Atherinomorus duodecimalis* specimens were collected from Matnog Bay and Puerto Galera, and *S. spinus* specimens were collected from Port Gubat.

Laboratory Protocol

The DNA for this study has been extracted previously by collaborators at Old Dominion University. For museum specimens, tissue was dissected from the right side of each fish, and for contemporary specimens, ~50 mg of tissue from fin clips was dissected. The Qiagen DNeasy® Blood & Tissue Kit was used to perform DNA extractions with the following alterations from the animal tissue spin column protocol. (1) DNA was extracted from ~50 mg of tissue from each fish. (2) The amount of buffer ATL/PK and buffer AL were doubled. (3) Digestion lasted for 150 mins. (4) The optional step of adding RNase was performed with double the recommended amount, 8 µl. (5) each centrifuge step was preformed twice to ensure that the silica membrane was dry with no EtOH or salt carryover. Samples were transferred to the Genomics Core Laboratory at Texas A&M University – Corpus Christi where the size distribution of the extracted DNA fragments was visualized using standard 1% gel electrophoresis in 1x TAE buffer for 40 minutes with the Bioline HyperLadder 1kb ladder. The concentration and amount of DNA extracted from each fish was determined using the AccuClear® Ultra High Sensitivity dsDNA Quantitation Kit (fluorescent quantification) with eight DNA concentration standards (0.3 - 250 ng) following the standard protocol on a SpectraMax M3 Plate Reader. Duplicate reactions were performed for each sample and the mean DNA concentration for each fish was calculated from the replicates.

We followed the *New RAD Protocol* as described in Ali et al. (2015) with the following changes. One AMPure XP Bead Cleanup was performed prior to the restriction digestion reaction (Fig. 1a). Genomic DNA (150 ng) was digested with 0.12ul NEB SbfI (20U/ul) R3642S in NEB CutSmart Buffer at 37 °C for at least 1 hour (Fig. 1b). To allow quantitation and more precise control of DNA concentrations, individual samples were not pooled after the ligation of biotinylated barcoded adapters (Fig. 1c). Samples were sonicated for 10 seconds on, 90 seconds off, for 0-8 cycles depending on their average fragment length prior to sonication. The smaller the average fragment size, the less sonication was required to reach the desired average fragment length of 500 bp. Contemporary *S. spinus* samples were sonicated for 6 cycles and museum *S. spinus* samples were sonicated for 2 cycles. Contemporary *A. duodecimalis* samples were sonicated for 5-8 cycles and museum *A. duodecimalis* samples were sonicated for 0-2 cycles. Samples were enriched for the RAD loci by capturing the biotinylated DNA fragments with 2 ul of streptavidin coated ThermoFisher Dynabeads per sample (Fig. 1d). Room temperature 1X binding and wash buffer was substituted for 1X NEBuffer4 for the two final washes. DNA was then resuspended in 9.88ul NEB CutSmart Buffer and 0.12ul NEB SbfI High Fidelity restriction enzyme (20U/ul) and incubated at 37°C for 1 hr to liberate the RAD fragments from the Dynabeads (Fig. 1e) before being precipitated with 1X Omega Biotek Mag-Bind beads. We performed dual size selection with the Mag-Bind beads. Samples were quantitated using the Biotium AccuBlue High Sensitivity dsDNA Quantitation kit, concentrations normalized, and samples with unique barcodes were pooled. Blunt end repair (Fig. 1f), A-tailing and Illumina adapter ligation (Fig. 1g) were completed using a KAPA Hyper Prep Kit at 1/3X reaction volumes. Prior to PCR, DNA was cleaned again using AMPure XP beads. DNA was PCR amplified using 2X KAPA HIFI ReadyMix with the following thermal profile (98 °C 45 sec;14

cycles of 98°C 15, 60°C 30 sec, 72°C 30 sec; 72°C 1 min; hold at 4°C) (Fig. 1h). Another AMPure XP Bead cleanup was performed post-PCR. DNA concentrations were measured using AccuBlue fluorescent dsDNA quantitation. BluePippin size selection was used to select sequences 350-700bp in length before samples were normalized using the KAPA qPCR Genomic Library Quantification Kit. Paired end 150 bp sequencing was performed by NovoGene (Sacramento, CA) on an Illumina HiSeq 4000.

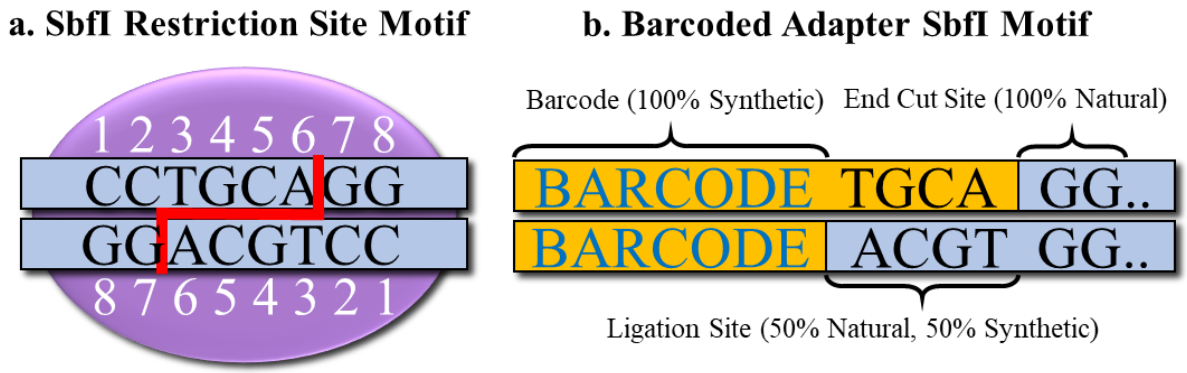

Figure 2. (a) The 8 bp SbfI recognition site that was targeted in the restriction digest of genomic DNA, and (b) the 16 bp barcoded adapter SbfI motif that is expected at the beginning of every sequence read (top strand) and its compliment (bottom strand). Purple is the restriction enzyme; orange is the synthetic oligonucleotide; and blue is the fish DNA.

Data Processing

Most of the data processing was focused on quantifying mutations associated with the barcoded adapter motif. The barcoded adapter motif is composed of both synthetic oligonucleotide barcoded adapter and restriction digested natural fish DNA (Fig. 1). Because we were interested in assessing the specificity of enzymatic reactions, during the library preparation protocol, leaving the barcoded adapter motif unmodified was essential for the accurate assessment of error rates. The barcoded adapter motif can be broken into three sections. (1) The barcode region which is 100% synthetic oligonucleotide, (2) the ligation site where the sticky end of the biotinylated barcoded adapter was ligated to the sticky end of digested DNA (Fig. 1c)

making it 50% synthetic and 50% natural, and (3) the end cut-site which comprises the last two nucleotides in the barcoded adapter motif (7-8, Fig. 2a; 15-16 Fig. 2b). The last two nucleotides in the motif are completely derived from fish; the expected sequence is known due to the nature of the RAD protocol; and they can be used to assess restriction enzyme specificity because they are part of the SbfI recognition sequence. The SbfI restriction enzyme should bind to its recognition site, the 8bp SbfI restriction enzyme motif, before making a cut (Fig. 2a). Due to the palindromic nature of the SbfI recognition site, both sticky ends created by SbfI restriction digestion (Fig. 1b) are capable of ligating to biotinylated barcoded adapters meaning that the end cut-site is representative of both the first and last two base pairs in 8bp SbfI restriction site motif.

Trimmomatic (Bolger, Lohse, & Usadel, 2014) was used to remove Illumina adapters and low-quality bases with mean phred quality scores ≤ 20 in a 20 bp sliding window, then a 1 bp sliding window, from the 3' ends of the raw sequence reads (Fig. 3a). Reads were not trimmed from the 5' end as this would likely affect the 16 bp barcoded adapter motif (Fig. 2b) which should occur at the beginning of every sequence read. Reads ≤ 75bp in length after adapter and quality trimming) were removed from further consideration because this is the length filter we typically use in a population genetic study with 150 bp paired end reads (Fig. 3a). BASH (Ramey, 1994) scripts (French, M. 2021) were used to count the number of reads which made it through quality and adapter trimming. A subset of 10,000 reads/library passing the filter were used for further data processing and analysis. Reads that had the first 14 bp of the expected barcoded adapter motif (Fig. 2b) with ≤ 1 substitution error, ≤ 2 leading deletions, and ≤ 8 leading insertions were identified using the approximate matching agrep command (Manber & Wu, 1991; Fig. 3b). The last two bases in the motif were free to vary so that error rates associated with the restriction digest were not altered. Mean substitution error rates were

calculated in the barcode, ligation site, and end cut site (Fig. 2b.) using an R (R Core Team, 2013) script. Reads with leading indels were excluded from the analysis of error rates at positions 7 & 8 of the SbfI restriction site motif (Figs. 2a, 3c) because they exhibited higher error rates, and would be deemed inviable for population genomic data analysis.

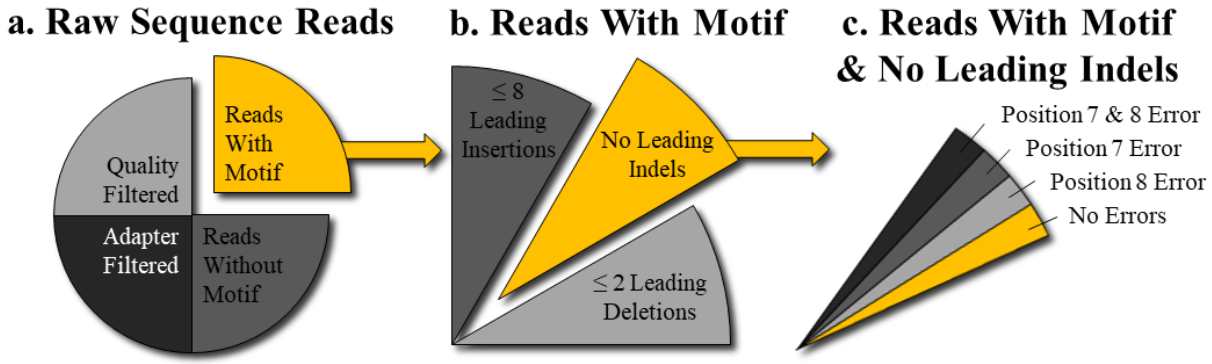

Figure 3. Conceptual diagram of the data processing pipeline that was used to classify (a) raw sequence reads into four groups based on those failing the filter and the presence or absence of the expected barcoded adapter motif at the beginning of the read, (b) the subset of reads containing the expected motif into three sub groups based upon the presence or absence of insertions or deletions at the beginning of the motif, and (c) the subgroup of reads without leading indels into four sub subgroups based upon errors in the last two positions of the SbfI recognition site. The "motif" is the first 14 bp of the barcoded adapter and SbfI recognition site (Fig. 2b). Orange identifies expected read pattern, and in (c) the reads that should be from targeted SbfI RAD loci and which are usable in downstream analyses. The proportions depicted here are not representative of the proportions observed or expected in this study, but the goal is to maximize the orange-coded categories relative to the others.

Using the methodologies in the previous paragraph, raw reads were effectively assigned to one of four categories (Fig. 3a.). (1) Quality filtered reads are indicative of poor sequencing performance. (2) Adapter filtered reads are indicative of unintended reactions in steps a-e and/or g of the laboratory protocol (Fig. 1). (3) Reads without the expected motif are indicative of unintended reactions in the RAD fragment enrichment (steps d-e). (4) Reads with the expected motif were classified into three subgroups. Reads with either (4.1) leading deletions or (4.2) insertions indicate that unintended reactions occurred in either blunting (step f) or ligation of the Illumina adapters (step g), respectively (Fig. 1). Differences in error rates in the synthetic

oligionucleotide portion (Fig. 2b) of the reads with the expected motif indicates altered specificity of the DNA polymerase during PCR (step h). Reads without leading indels were further subdivided into four sub subgroups. (4.3.1-3) Reads with errors in positions 7 and/or 8 of the SbfI recognition site are indicative of nonspecific restriction digestion in step b of the laboratory protocol.

Statistical Data Analysis

To test for the effects of species, collection-time, and their interaction on the number of read pairs filtered and those with and without the expected 16bp barcoded adapter motif (Fig. 2b), negative binomial regression was performed using glm.nb in the MASS R package (Ripley, 2021) and ANOVA from the car R package (Fox, 2021). Negative binomial regression was used due to overdispersion of the error relative to a Poisson regression. Type II likelihood ratio tests were implemented to assess model fitness if there were no significant interactions between species and collection-time. Alternatively, a type III likelihood ratio test was used if there was a significant interaction between variables.

To test for the effects of species, collection-time, and their interaction on the proportions of read pairs with and without the expected motif and those that were filtered, beta regression, from the betareg R package (Cribari-Neto, Zeileis, 2010), was chosen to allow for heteroskedasticity, which is commonly observed in proportional data. Beta regression likelihood ratio and joint tests (Abel, 2013) were used to assess model fitness. The effects of collection time within each species was assessed using joint tests.

Base call error rates in the expected 16 bp barcoded adapter motif at the beginning of each sequence read were compared among collection-times (museum and contemporary), species

(*A. duodecimalis* and *S. spinus*), and indel category (Insertions, No Indels, Deletions). Additionally, we tested for error rate differences between different positions in the motif, as well as reads with (a) no indels at the beginning of the motif (GG), (b) 1-2 deletions at the beginning of the motif (-G, G-, or --), and 1-8 insertions at the beginning of the motif ($N_{1\text{-}8}$GG). Reads were aligned by motif, and error rates were calculated for each of the 16 motif positions as the number of reads with mismatching bases divided by the total number of reads. Differences in error rates were tested using ANOVA and Pairwise Post-hoc tests

To test for the specificity of the restriction enzyme, we evaluated the error rates at positions 7 and 8 of the SbfI restriction site which are the only in our sequences for which we know the expected nucleotide sequences (GG) and that are genomic DNA sequences, (i.e., not synthesized barcode sequences). It is noteworthy that due to the palindromic nature of the SbfI restriction site and the nature of the RAD sequencing protocol which results in the removal of either the first or last two positions in the restriction site, positions 1 and 2 of the restriction site cannot be deciphered from positions 8 and 7, respectively, and we refer to them as the latter. The expectation is that positions 7 and 8 of the restriction site are derived from the fish, and error rates were estimated as the proportion of reads that have a nucleotide other than G.

To assess the effects of species and collection time on restriction enzyme specificity, we fit Bayesian generalized multinomial models (read counts ~ species * collection time + library) with 4 chains for 5000 iterations (2500 warm up) to the error rate data using the brms (Burkner, Gabry, Weber, Johnson, & Mordrak, 2021) and tidybayes (Kay, 2021) R packages to estimate the proportion of reads with (1) errors at neither position ($e_{obs_{neither}}$), (2) errors only at position 7 ($e_{obs_{only\,7}}$), (3) errors only at position 8 ($e_{obs_{only\,8}}$), and (4) errors at both positions ($e_{obs_{both}}$). For this, and all subsequent analyses, only reads without leading indels were included because of

the elevated error rates associated with sequences exhibiting leading indels (see Results). The posterior distributions were used to generate the median, 95% and 99% credible intervals for each error pattern. Contrasts were constructed to test for differences in error rates between collection times and species using the emmeans R package (Lenth, 2021). If the 99% credible interval of the difference between the estimated marginal means did not include zero, the null hypothesis of no difference was rejected.

Errors in positions 7 and 8 might be associated with altered or damaged SbfI restriction enzymes. We hypothesized, *a priori*, that the observation of an error at position 7 would likely lead to a random base in position 8 due to the chemistry and spatial mechanics involved in the interaction of a proteinaceous restriction enzyme with a DNA recognition site. This would result in an excess of reads with errors in both positions 7 and 8. As a statistical test, the credible intervals of the observed error rates from the Bayesian multinomial model fitting described above were compared with the expected read counts for each error pattern given the assumption that errors in positions 7 and 8 are independent (null model), which were calculated as follows:

$$\text{Eq. 1} \quad e_{ind_{neither}} = (1 - e_{obs_7})(1 - e_{obs_8})$$

$$\text{Eq. 2} \quad e_{ind_{only\,7}} = e_{obs_7}(1 - e_{obs_8})$$

$$\text{Eq. 3} \quad e_{ind_{only\,8}} = e_{obs_8}(1 - e_{obs_7})$$

$$\text{Eq. 4} \quad e_{ind_{both}} = e_{obs_7} e_{obs_8}$$

where *ind* is the expectation given the independence of errors in positions 7 and 8 and *obs* is observed. If the null expectation fell outside of the 99% credible interval of a point estimate, the

observed error rate was considered to represent a rejection of the null hypothesis and an indication that there was non-independence of error rates at the two positions.

To explicitly test for the dependence of the error rate at position 8 on position 7, we used the Akaike Information Criterion (AIC) and Bayesian Information criterion (BIC) (Dziak, Coffman, Lanza & Li, 2021) to determine whether the null model of error independence (Eq. 3) or an alternative model of error dependence best fit the observed error rates. In the alternative model of dependence, the overall error rate at position 8 (e_{dep_8}) is affected by errors that occur in only position 8 ($e_{obs_{only\,8}}$) and the probability of randomly drawing a nucleotide other than G (q) from the sequencing library in position 8 when there is an error at position 7 (e_{obs_7}):

$$\text{Eq. 5} \quad e_{dep_8} = e_{obs_{only\,8}} + e_{obs_7} q \qquad .$$

The probability of randomly drawing a nucleotide other than the expected G (q) was calculated independently for each library. FASTQC v0.11.9 (Babraham Bioinformatics, 2021) was used to quantify the proportion of each nucleotide at each read position in each library, and MULTIQC v1.9 (Ewels, Magnusson, Lundin, & Käller, 2016) was used to aggregate the FASTQC results from each library. The JSON file with this data was extracted from the HTML file using, GNU Bourne-Again Shell (Ramey, 2021) commands, read into R using the rjson package (Couture-Beil, 2021), and transformed into a tidy format using the tidyverse R package (Wickham, 2021). We used positions 25-50 to calculate the mean proportion of nucleotides that were A, C, or T because the first 24 nucleotides were either composed of synthetic oligonucleotides or could contain artifacts and base call error rates tend to increase towards the end of sequence reads. We additionally tested for deviations from the observed data by plotting the expectations of both models against the observed data using the core R command lm (model ~ observations) and

testing whether the slopes were significantly different than 1 and the elevations (y-intercepts) were significantly different than zero using the linear Hypothesis command in the car R package (Fox, 2021).

RESULTS

Sequence Composition & Yield

Contemporary samples performed either significantly better than or similarly to museum samples in terms of both the mean number (binomial model) and proportion (beta model) of raw reads with and without the expected motif, as well as those that were filtered (Fig. 4; Table 1; α = 0.01). There were a higher proportion (beta model) and number of read pairs (binomial model) from contemporary samples with the expected barcoded adapter motif for both species when compared with those from museum specimens (Fig. 4; Tables 1, S1; $p < 0.01$). There was a significant interaction in the number and proportion of read pairs with the expected motif between collection-time and species (Fig. 4; Tables 1, S1; $p < 0.0001$), with a larger difference between collection times in *S. spinus*.

Reads without the expected barcoded adapter motif are likely due to poor performance of the RAD enrichment and biotin removal (steps d-e, Fig. 1). In both species, the museum specimens had more and higher proportions of reads without the motif (Fig. 4; Table 1; α = 0.01). There was a significant interaction between collection-time and species in the proportion of read pairs without the expected motif (Fig. 4; Tables 1, S1; $p < 0.0001$), where the mean difference between collection times was disproportionately greater in *S. spinus* than *A. duodecimalis*.

While a greater number of contemporary reads were filtered than museum reads on average (Fig. 4; Table 1; $p < 0.01$), there was not a significant difference in the proportions of reads filtered between collection-times (Fig. 4; Table 1; $p = 0.24$) suggesting that the methodological issues causing reads to be filtered were similar between sample collection times. The overwhelming majority of filtered reads (> 99%) in both species were removed because they were too short (< 75 bp) after the trimming of adapter sequences, rather than being removed for low base call quality (Table S2), suggesting that the library preparation was negatively affected by museum specimens but not the sequencing. For most filtered reads, there were 0 bp remaining after trimming adapters, indicating that there was no DNA insert from the targeted fish species, and the filtered reads were mostly composed of adapter dimer, which can result from poor reaction performance in any or all protocol steps a-e and g (Fig. 1).

There was also a marked and statistically significant difference between species (Fig. 4; Table 1; $\alpha = 0.01$), but the purpose here was not to compare species performance, and other factors covaried with species, such as the order in which the species were processed. The *S. spinus* libraries, which were processed after *A. duodecimalis* and likely benefitted from improved protocol execution, had higher numbers and proportions of read pairs with the expected motif and lower proportions and numbers of read pairs filtered (Fig. 4; Table 1; $\alpha = 0.01$). However, the proportion of read pairs without the expected motif was greater in *S. spinus* than *A. duodecimalis* (Fig. 4; Table 1; $\alpha = 0.01$), likely due to the excessive amount of filtered reads in *A. duodecimalis*.

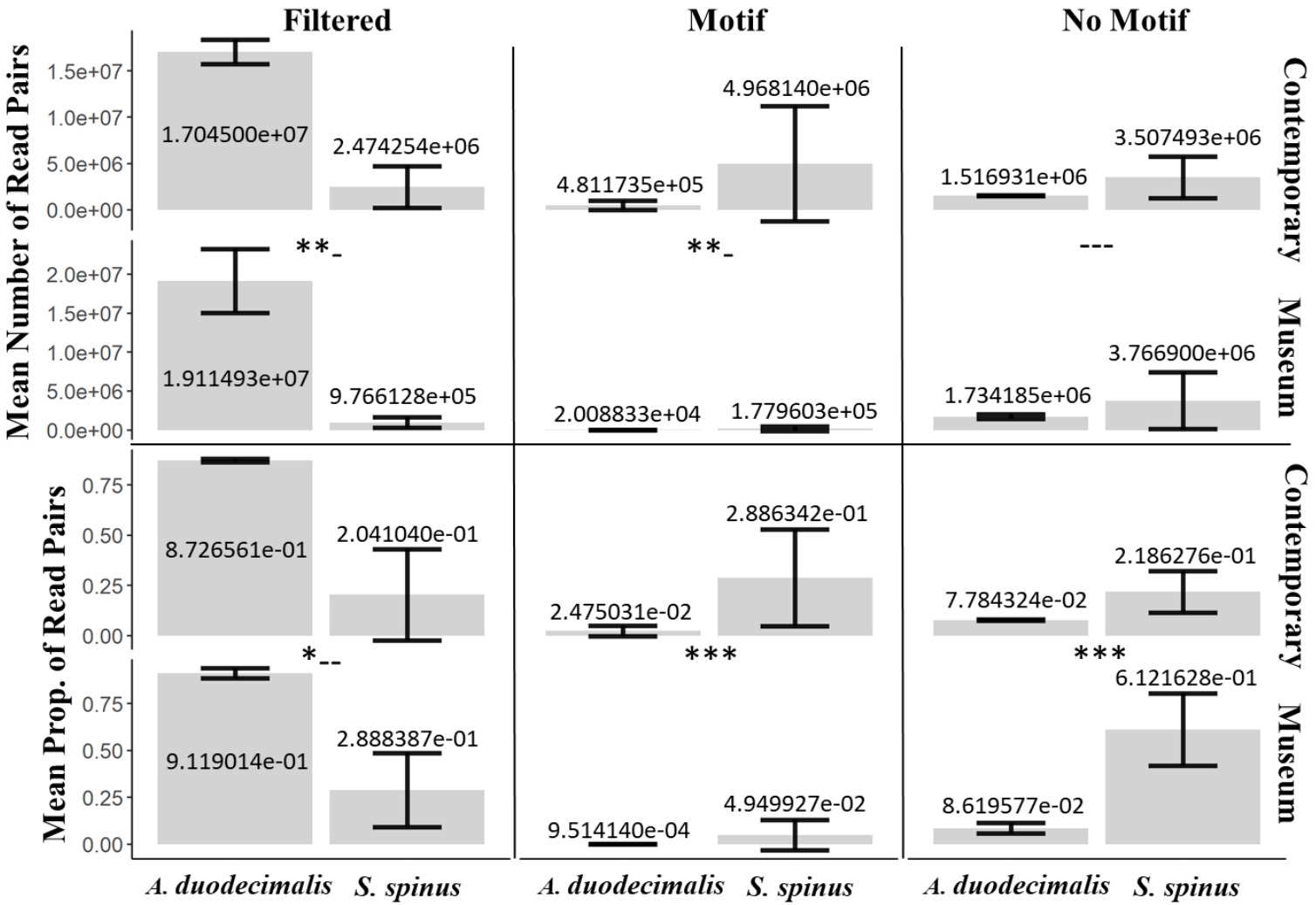

Figure 4. Mean numbers and proportions of read pairs for each combination of species and collection time. Read pairs were divided based on whether they were filtered from the data set due to excessive low quality base calls or adapter sequences (Filtered), they did (Motif) or did not have the expected sequence motif composed of the barcode and restriction site (No Motif). The * and - represent the statistical significance (p-value < .01), or lack thereof (p-value > .01) for the effects of species, collection time, and their interaction on the number and proportion of read pairs, respectively.

Table 1. Results of tests for differences in the counts and proportions of sequenced read pairs between species (*A. duodecimalis, S. spinus*), collection times (Contemporary, Museum), and their interaction using negative binomial and beta models. The read pairs were divided into categories based on whether they were filtered from the data set due to excessive low quality base calls or adapter sequences (Filtered), they had the expected sequence motif composed of the barcode and restriction site (Motif) or did not have the motif (No Motif).

Negative Binomial Model (Counts)

Category	Model Term	*df*	χ^2	*p*
Filtered	Species	1	76.3330	< **0.0001**
	CollectionTime	1	6.7510	**0.0094**
	Species:CollectionTime	1	2.7080	0.0999
Motif	Species	1	19.0000	< **0.0001**
	CollectionTime	1	62.0770	< **0.0001**
	Species:CollectionTime	1	0.0350	0.8522
No Motif	Species	1	3.4200	0.0644
	CollectionTime	1	0.0665	0.7965
	Species:CollectionTime	1	0.0060	0.9381

Beta Model (Proportions)

Category	Model Term	df_1	df_2	*F* Ratio	*p*
Filtered	Species	1	Inf	70.9040	< **0.0001**
	CollectionTime	1	Inf	1.3800	0.2402
	Species:CollectionTime	1	Inf	0.7580	0.3838
Motif	Species	1	Inf	35.9380	< **0.0001**
	CollectionTime	1	Inf	42.2460	< **0.0001**
	Species:CollectionTime	1	Inf	24.8660	< **0.0001**
No Motif	Species	1	Inf	61.0210	< **0.0001**
	CollectionTime	1	Inf	39.3450	< **0.0001**
	Species:CollectionTime	1	Inf	21.7230	< **0.0001**

Error Rates in the Barcoded Adapter & SbfI Restriction Site

Deletions and insertions at the beginning of the barcoded adapter motif are indicative of unintended reactions in the blunting and Illumina adapter ligation steps of the laboratory protocol, respectively (Figs. 1f, g). The majority of reads that made it through quality and adapter trimming with the expected barcoded adapter motif, had no leading indels for all four combinations of species and collection time (Table 2). There was a greater proportion of reads with no indels from contemporary compared to museum samples and for *A. duodecimalis* compared to *S. spinus* samples.

Table 2. Proportions of reads that had the expected sequence motif composed of the barcode and restriction site sorted by the number of leading indels (2 Deletions - 8 Insertions) for each combination of species (*A. duodecimalis*, *S. spinus*) and collection times (Museum, Contemporary).

	A. duodecimalis		***S. spinus***	
Indels	**Museum**	**Contemporary**	**Museum**	**Contemporary**
2 Deletions	0.01189	0.00308	0.00269	0.00185
1 Deletions	0.02113	0.00681	0.40997	0.25269
0 Indels	0.94850	0.98310	0.58250	0.74316
1 Insertion	0.00416	0.00204	0.00126	0.00101
2 Insertions	0.00271	0.00142	0.00047	0.00029
3 Insertions	0.00427	0.00068	0.00088	0.00029
4 Insertions	0.00150	0.00047	0.00090	0.00026
5 Insertions	0.00242	0.00068	0.00030	0.00012
6 Insertions	0.00115	0.00058	0.00054	0.00020
7 Insertions	0.00133	0.00060	0.00026	0.00006
8 Insertions	0.00092	0.00054	0.00023	0.00009

There were statistically significant differences in base substitution error rates and interactions between collection-time (museum, contemporary), barcoded adapter motif section (barcode, ligation site, end cut-site; see Fig. 2b), and indel category (deletions, no indels, insertions) for *A. duodecimalis* except for the interaction of collection-time and indel category

(Fig. 5; Table 3; $p < 0.001$). For *S. spinus*, there were statistically significant differences in base substitution error rates between motif section and indel categories, as well as interactions between collection-time, motif section and indel category (Fig. 5; Table 3; $p < 0.001$).

Table 3. Results of tests for differences in error rate between collection-time (Museum, Contemporary), target motif section (Barcode, Ligation Site, End Cut-Site), and Indels (Insertions, No Indels, Deletions) separated by species (*A. duodecimalis, S. spinus*).

Species	Model Term	*df*	χ^2	*p*
A. duo	Collection Time	1	15.42	< 0.001
	Target Motif Section	2	6844	< 0.001
	Indels	2	420.56	< 0.001
	Collection Time : Target Motif Section	2	344.6	< 0.001
	Collection Time : Indels	2	1.22	0.5440
	Target Motif Section : Indels	4	881.57	< 0.001
	CollectionTime : Target Motif Section : Indels	4	19.86	< 0.001
S. spi	Collection Time	1	0.31	0.5770
	Target Motif Section	2	5228	< 0.001
	Indels	2	1905.4	< 0.001
	Collection Time : Target Motif Section	2	23.25	< 0.001
	Collection Time : Indels	2	331.99	< 0.001
	Target Motif Section : Indels	4	4209.8	< 0.001
	CollectionTime : Target Motif Section : Indels	4	1600.3	< 0.001

Within each target motif section, reads with insertions and deletions generally had higher error rates than those with no indels (Fig. 5; Table S3.1). In the few cases where reads with no indels had a higher mean error rate, the differences were either small or not significant. There were 0.2 - 2.9% more errors in the barcodes of reads with insertions versus reads without insertions for both species and collection-times.

For both species and all indel categories, where pairwise post-hoc contrasts showed significant differences in collection-time, museum samples had higher error rates than contemporary samples of the same motif section and indel category. In the few cases where

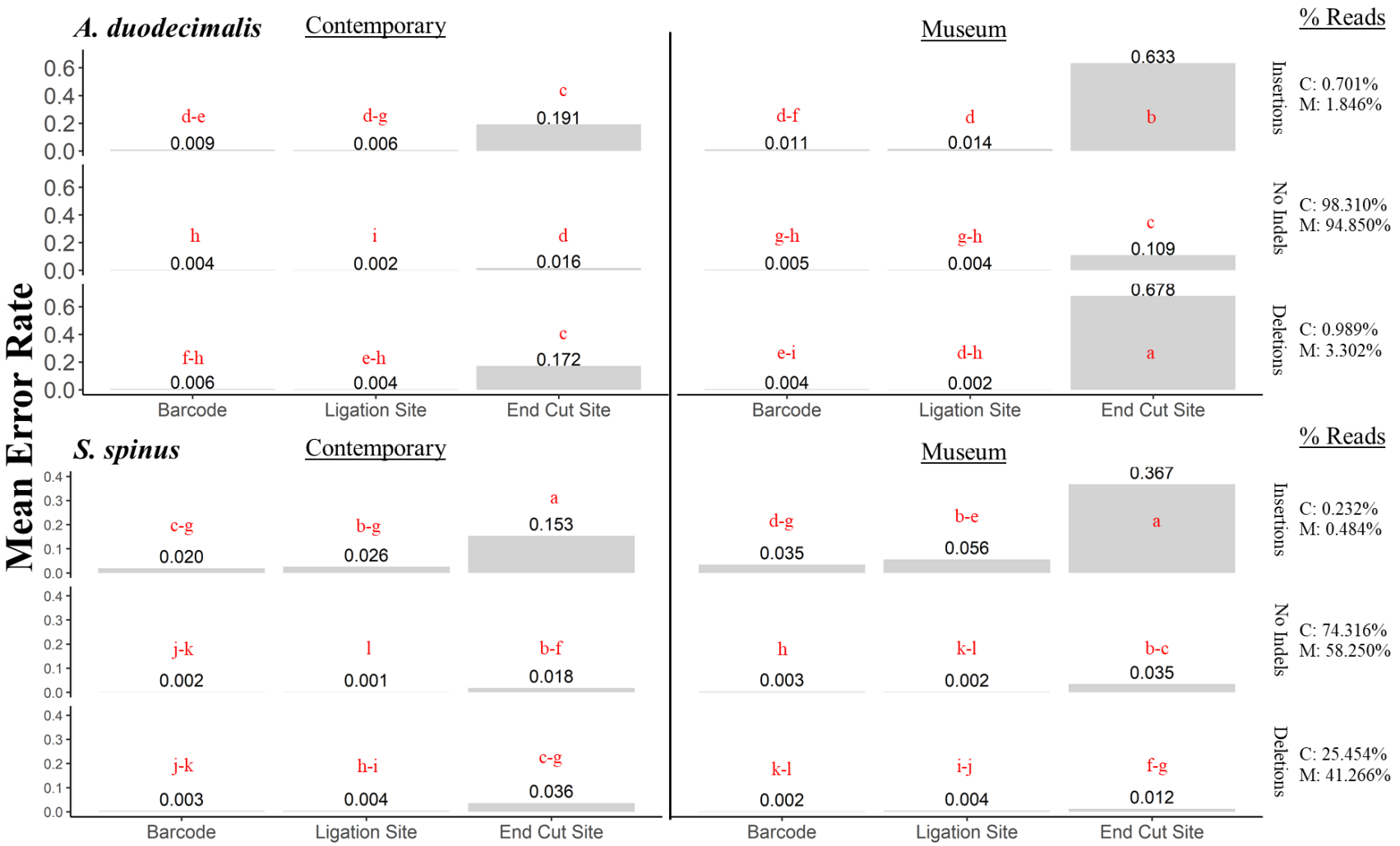

Figure 5. Mean error rates in targeted motif sections (Barcode, Ligation Site, End Cut-Site). Read pairs were separated by species, collection-time, and indel category. The mean error rates of treatment combinations with the same letter code within a species are not significantly different (species were not compared). Indel categories were determined by the presence or absence of inserted or deleted nucleotides at the start of each read. The percentage of total reads represented by each indel category is shown to the right for contemporary (C) and museum (M) samples.

contemporary samples exhibited a higher error rate than their museum counterparts the difference was not significant (Fig. 5; Table S3.2). The differences between museum and contemporary samples were the most pronounced in the end cut-site motif section and in reads with insertions.

In general, reads with leading insertions and deletions had higher base substitution error rates than those with no indels within each motif section (Fig. 2; Table S3.3). Mean error rates in the end cut-site were significantly greater than in either the ligation site or barcode for both species, collection times, and all indel categories. The differences in error rates were smaller between the ligation site and the barcode but were higher in the ligation site when significant. As reads from museum samples and those with leading indels had higher error rates than contemporary samples and those with no indels, the differences between target motif sections were the greatest for reads from museum samples with insertions of deletions.

Error Rates in Positions 7 & 8 of the SbfI Restriction Site

There was a clear difference in the observed pattern of errors in positions 7 and 8, jointly, relative to random expectation for both collection times in both species (Fig. 6; Table 4). In particular, there was an excess of sequence reads with errors in both positions, while there were fewer reads than expected with errors at only positions 7 or 8. The observed error rates at position 7 only were lower than for 8 in all four treatment combinations. Lastly, the number of reads with no errors in these positions were lower than random expectation, except for the contemporary *A. duodecimalis*.

There were more reads with errors in both positions 7 and 8, only 7, or only 8 in the museum specimens than the contemporary given the non-overlapping 99% highest posterior density (HPD) intervals (Fig. 6; Tables 4, S4). There were also more reads with errors in only

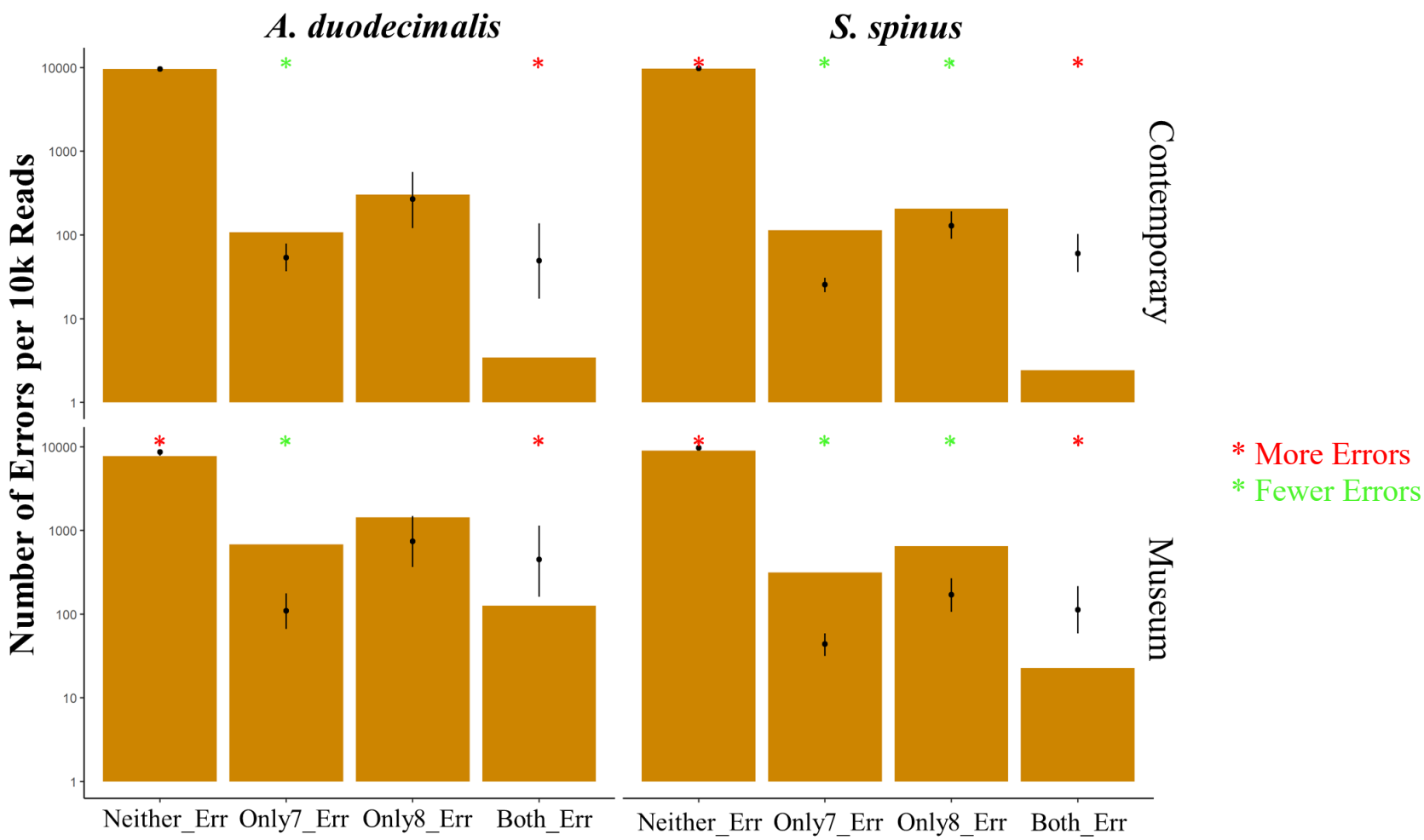

Figure 6. Comparison of observed error rates occurring in both positions 7 and 8 simultaneously, only 7, only 8, or neither 7 nor 8 to random expectation in a 10K subset of reads. The black points (best estimate) and error bars (99% credible intervals) are interval summaries of the posterior distribution of multinomial models fit to the nucleotide error data. The orange bars are the null expectation when errors in positions 7 and 8 are independent (*Eqs.* 1-4). The colored points indicate whether there were statistically significantly greater (red) or fewer (green) errors than predicted by the null model.

Table 4. Estimated marginal means pairwise comparisons for species (*A. duodecimalis, S. spinus*), collection time (Museum, Contemporary) and their interaction for errors occurring at only position 7, (Only 7), only position 8 (Only 8), and at both 7 and 8 (BothErr) using 95% and 99% highest posterior density intervals. Contrasts are significantly different when 0 is not contained within their highest posterior density intervals (shown in bold).

Error Type	Contrast	Diff in EMM	Lower 95% HPD	Upper 95% HPD	Lower 99% HPD	Upper 99% HPD
Only7	Aduo - Sspi	0.8960	**0.63**	**1.17**	**0.55**	**1.25**
Only8	Aduo - Sspi	1.1700	**0.70**	**1.63**	**0.52**	**1.78**
BothErr	Aduo - Sspi	0.6530	0.02	1.29	-0.20	1.50
Only7	Contemporary - Museum	-0.6930	**-0.96**	**-0.42**	**-1.05**	**-0.32**
Only8	Contemporary - Museum	-0.7040	**-1.17**	**-0.24**	**-1.32**	**-0.07**
BothErr	Contemporary - Museum	-1.4800	**-2.11**	**-0.86**	**-2.29**	**-0.65**
Only7	Aduo Contemporary - Sspi Contemporary	0.765	**0.43**	**1.07**	**0.32**	**1.17**
Only8	Aduo Contemporary - Sspi Contemporary	0.747	**0.11**	**1.42**	-0.20	1.59
BothErr	Aduo Contemporary - Sspi Contemporary	-0.184	-1.04	0.71	-1.36	0.97
Only7	Aduo Contemporary - Aduo Museum	-0.822	**-1.28**	**-0.35**	**-1.42**	**-0.19**
Only8	Aduo Contemporary - Aduo Museum	-1.123	**-1.93**	**-0.30**	**-2.20**	**-0.04**
BothErr	Aduo Contemporary - Aduo Museum	-2.312	**-3.38**	**-1.17**	**-3.84**	**-0.91**
Only7	Aduo Contemporary - Sspi Museum	0.201	-0.17	0.56	-0.27	0.69
Only8	Aduo Contemporary - Sspi Museum	0.466	-0.26	1.14	-0.40	1.37
BothErr	Aduo Contemporary - Sspi Museum	-0.823	-1.77	0.05	-2.02	0.45
Only7	Sspi Contemporary - Aduo Museum	-1.588	**-1.98**	**-1.18**	**-2.11**	**-1.06**
Only8	Sspi Contemporary - Aduo Museum	-1.865	**-2.51**	**-1.24**	**-2.72**	**-1.02**
BothErr	Sspi Contemporary - Aduo Museum	-2.134	**-2.98**	**-1.24**	**-3.21**	**-0.95**
Only7	Sspi Contemporary - Sspi Museum	-0.562	**-0.83**	**-0.28**	**-0.89**	**-0.16**
Only8	Sspi Contemporary - Sspi Museum	-0.284	-0.74	0.16	-0.87	0.34
BothErr	Sspi Contemporary - Sspi Museum	-0.647	**-1.23**	**-0.02**	-1.41	0.20
Only7	Aduo Museum - Sspi Museum	1.027	**0.59**	**1.46**	**0.47**	**1.61**
Only8	Aduo Museum - Sspi Museum	1.574	**0.91**	**2.24**	**0.71**	**2.49**
BothErr	Aduo Museum - Sspi Museum	1.496	**0.57**	**2.37**	**0.29**	**2.65**

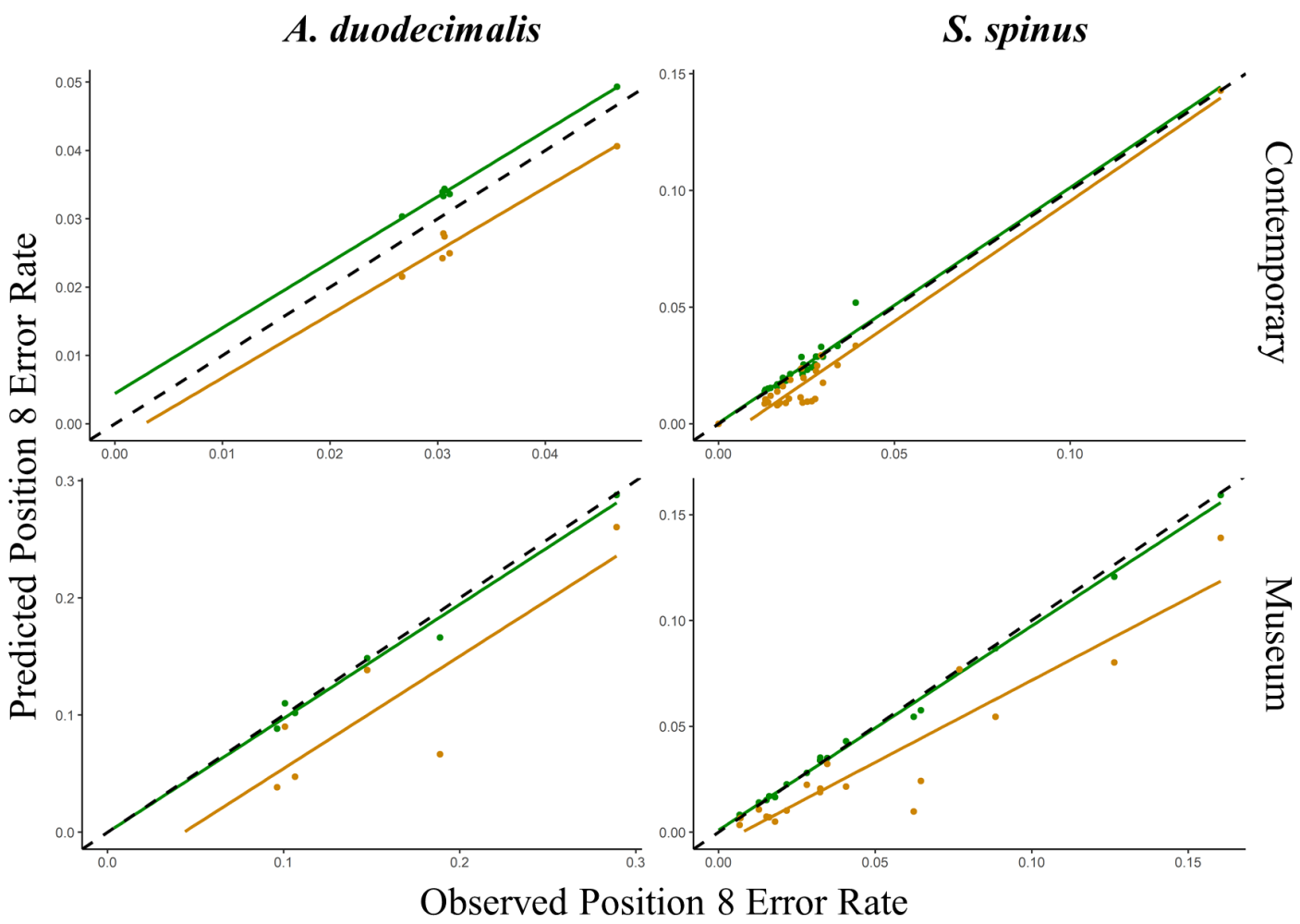

Figure 7. Model-predicted error rates at position 8 (including reads with errors in both positions 7 and 8) based upon either the independent (orange, *Eq*. 1-4) or dependent model (green, *Eq*. 5) of errors between positions 7 and 8 plotted against the observed error rates. The solid lines are best-fit least squares linear regression models, and the dashed line represents an exact fit to the observed error rate. See Table 5 for AIC and BIC estimators.

position 7 or 8 in *A. duodecimalis* than *S. spinus* given non-overlapping 99% HPD intervals. There was, however, only a small amount of overlap in 95% HPD intervals (and no overlap in the 90% HPD intervals) between the two species for reads with errors in both positions. When parsing the treatment combinations of species and collection time, there were significant differences (95% HPD) between the EMM for museum and contemporary samples for both species individually, with reads from museum samples having more errors in all of the error categories, with the one exception of position 8 errors in *S. spinus*.

When comparing the AIC and BIC for the null and alternative models, the error rates for position 8 predicted by the non-independent alternative model conformed more closely to the observed position 8 error rates than for the independent null model regardless of species and collection time (Fig. 7; Table 5).

Table 5. Akaike Information Criterion (AIC) and Bayesian Information criterion (BIC) estimators for the model of independence represented by (Eq. 1-4) and the model of dependence represented by (Eq. 5). The smaller the AIC or BIC, the better the fit (bolded).

		Dependence of Errors		**Independence of Errors**	
Species	**Collection**	**AIC**	**BIC**	**AIC**	**BIC**
Aduo	Museum	**-30.877**	**-31.501**	-16.01346	-16.63818
	Contemporary	**-146.763**	**-145.308**	-118.2916	-116.8369
Sspi	Museum	**-145.19**	**-142.52**	-99.82341	-97.1523
	Contemporary	**-251.74**	**-247.64**	-218.6214	-214.5195

DISCUSSION

The museum specimens performed significantly worse than or equal to, but not better than, the contemporary samples in terms of reads passing filters, leading indels, and base substitution error rates in the barcoded adapter and SbfI restriction site. The differences in performance between museum and contemporary sequencing libraries are consistent with more

unintended reactions occurring during the library preparation in the museum specimens (see Fig. 1). In particular, there are strong indications that the specificity of restriction digest (base substitution errors in the end cut site, positions 7 & 8 of the recognition site), ligation (base substitution errors and insertions in the ligation site), blunting (leading deletions in the barcoded adapter), and PCR (base substitution errors in the barcode) reactions are negatively affected in libraries constructed from museum specimens. These observations span both synthetic and natural DNA and, consequently, cannot all be driven by molecular changes in the historical DNA. We propose that the differences in performance between museum and contemporary samples can all be attributed to contaminants that interfere with enzymatic processes. An additional observation supporting interference with enzymatic processes in the samples collected by the Albatross Expedition was that in a separate laboratory trial, contemporary DNA spiked with museum DNA resulted in poor whole genome amplification when compared with just contemporary DNA (pers. obs.). It is possible that there have also been DNA alterations in museum specimens, but they need not be invoked to explain the results and cannot explain substitution errors and indels associated with the barcoded adapter sequence.

Identifying the contaminant(s) is beyond the scope of this effort, but they may be related to the preservation media. It is notable that the museum specimens were collected in the Philippines (1907-10) and were stored in EtOH produced by rum distilleries. Distilled EtOH, particularly from sugar cane, can be associated with a variety of volatile compounds (Riachi et al., 2014) that may interfere with the enzymatic reactions involved in the library preparation for DNA sequencing. There is no record of the museum specimens in this study being fixed in formalin, and the eye pupils of the fish were white, suggesting they were not fixed in formalin (De Bruyn et al., 2011). Formalin fixation became a common practice in museums while the

Albatross Collection was in the custody of the Smithsonian Institute, so it may be possible that the specimens were exposed to formalin after several years in EtOH. It is clear, however that the contaminant is either potent at low concentrations or not removed by silica membranes or paramagnetic beads in DNA isolation reactions.

Sequence Read Composition & Yield

The appreciable proportion of reads observed without the barcoded adapter motif (Fig. 4) is notable because these DNA fragments should have been removed prior to the ligation of Illumina adapters (step g, Fig. 1), during the enrichment of RAD loci with streptavidin Dynabeads (step d, Fig. 1). Because all of the biotinylated DNA should have the 16bp barcoded adapter motif (Fig. 2b), this cleanup should result in the vast majority of sequenced reads having that motif. The possible explanations for the surprisingly high proportion of reads without the barcoded adapter motif in museum samples are that (1) the streptavidin-biotin bead cleanup, intended to enrich for biotinylated DNA fragments, was less effective for museum samples (step d, Fig. 1), (2) the specificity of DNA polymerase was affected during PCR amplification for museum samples, resulting in the introduction of more than 1 error per read motif. Of these options, we believe that a reduction in the effectiveness of the streptavidin-biotin bead cleanup is more prevalent. If the observed pattern of errors were due to PCR, we would expect most of the "no motif" sequences to contain the barcoded adapter motif with >1 error, but this is not the case. When allowing for two errors in the pattern match to identify reads with the barcoded adapter motif, there is little impact on the number or proportion of reads without the motif. The mean number of reads with the motif (Fig. 4) increases by only 16.5% for *A. duodecimalis* museum, 1.76% for *A. duodecimalis* contemporary, 4.09% for *S. spinus* museum, and 0.47% for *S. spinus*

contemporary. The largest differences in the number and proportion read pairs with and without the barcoded adapter motif are observed in museum samples, but we are ultimately unsure why museum stored specimens would result in less efficient separation of biotinylated from non-biotinylated DNA fragments. Previous studies employing reduced representation sequencing for museum samples or samples with degraded DNA have shown similar difficulties in achieving significant per locus read depth and producing sequences of good quality or with appropriate RAD loci (Norman, Street, & Spong, 2013; Tin et al., 2014; Graham et al., 2015; Burrell et al., 2015). *Siganus spinus* had higher numbers and proportions of reads without the barcoded adapter motif, and while this was at least partially due to the unusually large amount of adapter dimer in *A. duodecimalis* libraries, it does indicate high variation in effectiveness even in contemporary samples.

A large proportion of reads in the libraries were filtered due to short sequence length (<75 bp) following adapter trimming (Fig. 4), which includes adapter dimers. This is necessarily a consequence of having too little DNA of the expected length relative to the amount of adapter in the Illumain adapter ligation reaction (step g, Fig. 1). If excessive adapter were the cause, there would be a high proportion of adapter dimer that would be identifiable as reads with no length after the adapters are removed. Indeed, almost all filtered reads were adapter dimer, indicating that excessive adapter was present relative to the DNA inserts, especially in *A. duodecimalis*. The proportion of read pairs in *A. duodecimalis* could have been especially high because they were constructed first, and it was our first attempt at performing this protocol.

Indels and Base Call Error Rates in Sequences with a Barcoded Adapter

The most likely cause for the differences in the proportions of reads with leading indels between collection times (Table 2) are errors in the blunting step (deletions, Fig. 1f) and ligation of Illumina adapters (insertions, Fig. 1g). The best explanation for leading deletions is that the blunting reaction prior to A-tailing, where exonucleases remove single-stranded DNA from the ends of the double-stranded fragments, resulted in one or two double-stranded base pairs being removed. We are unsure how contaminants associated with the museum specimens would affect the behavior of the exonucleases, but it seems that either they are more likely to remove double stranded nucleotides or the hydrogen bonds between the two nucleotide pairs at the ends of the sequences are more likely to break, presenting the exonucleases with additional single stranded DNA to remove. The insertions likely occurred during the Illumina ligation reaction by the incorporation of free nucleotides between the adapters and the target DNA (Fig. 1g). We are unsure why leading insertions would be consistently more prevalent in the museum samples for both species, but the effect size was generally small and would have little impact on efforts to sequence museum samples. In the case of *S. spinus,* where there was a relatively large proportion of reads with a single leading deletion, there was no associated increase in base substitution errors, making these reads are viable for downstream analysis (Table 2). Importantly, there was no confounding of sampling time (museum, contemporary) with the laboratory learning curve, and thus, the effect of museum is associated with the observed differences.

While the high proportion of *S. spinus* reads with leading deletions relative to *A. duodecimalis* samples would be consistent with 5’ trimming *S. spinus* reads prior to the determination of the presence of deletions (identified as a sequence missing one or both of the leading G followed by the adapter sequence), we did not 5’ trim reads from either species

meaning this explanation is not viable. This disparity is not attributable to preexisting errors in the barcoded adapter created during synthesis because the same batch of adapters was used for both species and collection times, meaning we would expect the errors to be distributed evenly across treatments. Degradation of the adapters between library preparations is also not a good explanation of the observed pattern, primarily because the expected GG at the beginning of each read is not exposed until the second restriction digest. Further, if the SbfI restriction site was mutated, then digestion and retainment in the sequencing library would be much less likely.

Reads with unexpected leading insertions and deletions were associated with higher mean error rates in the barcoded adapter motif (Fig. 5), suggesting a connection between indels and base substitution. Because reads with and without leading indels were realigned before error rate calculation, misalignment caused by the indels was not responsible for the observed pattern of elevated substitutions errors. Because most of the motif is the barcoded adapter from the first ligation and indels are caused by the second ligation reaction, the correlation is somewhat enigmatic.

Elevated base substitution error rates in museum specimens were observed across all three sections of the barcoded adapter motif (Fig. 2b), the barcode (synthesized DNA), the ligation site (synthesized & natural fish DNA), and the end of the restriction site (natural fish DNA). Elevated error rates in synthesized DNA suggests that a chemical contaminant present in the museum samples resulted in either (1) increased frequency of spontaneous base changes through hydrolytic deamination or other nucleotide alterations during library preparation, (2) elevated errors or reduced proofreading activity by DNA polymerase during PCR. We would expect errors introduced due to altered DNA polymerase activity during PCR and sequencing to be randomly distributed and therefore just as likely to occur within any section of the target

motif, but we are unable to isolate this effect from other sources of error, such as nucleotide alterations or non-specific ligation. Elevated error rates in the natural DNA of museum specimens could additionally be caused by nucleotide alterations during storage in the preservation buffer that led to misrecognition by restriction enzymes and/or polymerases and non-specific ligation. Altered restriction enzyme specificity and/or blunt-end ligation would lead to high error rates in the end of the restriction site, and this is discussed in detail in the following section.

Error Rates in Positions 7 & 8 of the SbfI Restriction Site

There are at least three explanations for the elevated error rates in positions 7 and 8 of the restriction sites of museum specimens relative to their contemporary counterparts but altered restriction enzyme specificity is the most consistent with the data. Base substitution errors could have been introduced by PCR. Base substitution errors occurring due to PCR, are most often single nucleotide substitutions (Eckert & Kunkel, 1991), and would be expected to result in the independence of errors between positions 7 and 8, but the model based on the dependence of errors between positions 7 and 8 (Eq. 5) was a better fit to the data than the independent model (Eqs. 1-4; Figures 6,7; Tables 5, S5.1).

Alternatively, errors may have been introduced during the ligation of the barcoded adapter (step c, Fig. 1) via unintended ligation of the sticky ended barcoded adapters to blunt ended DNA fragments. If the sticky end of the barcoded adapter is ligated to anything other than another sticky ended fragment, the gap of single stranded DNA could be filled by free floating nucleotides resulting in two random base pairs where positions 7 and 8 of the SbfI restriction site should occur and would result in the dependence of errors between positions 7 and 8. The museum samples were more fragmented than the contemporary, and while rare compared to

sticky-end ligation, blunt-end ligation (or partial sticky-end ligation) would be more common with more blunt ended fragments in solution. On the other hand, there are several impediments to the adapters ligating to the ends of degraded DNA fragments, and then subsequently being sequenced. First, the ends of some degraded DNA fragments would not have a 5' phosphate required for the ligation to occur. Should the non-specific ligation occur, then the overhanging nucleotides of the adapter would comprise a single stranded section of the ligated DNA molecule. Since free nucleotides will be rare due to paramagnetic, AMPure XP, bead cleanups prior to the ligation reactions, we do not expect that this section of DNA is likely to become double stranded prior to sonication when it is likely to be sheared. The non-specific ligation fragments that make it to PCR will not be amplified as efficiently as the expected ligation products. Those that do make it to sequencing would have two effectively random bases in positions 7 and 8 and would be difficult to decipher from fragments resulting from non-specific restriction enzyme activity prior to ligation. While possible, we believe the likelihood of this occurring is too low to explain the observed differences between museum and contemporary samples. Perhaps the best argument against the blunt end ligation hypothesis is that the putative contaminants associated with the museum samples would be enhancing this reaction, but the data indicates that other enzymatic processes are inhibited or degraded in some fashion.

We argue that the most likely explaination for the elevated error rates observed in positions 7 and 8 (Fig. 6; Table 4) of the SbfI restriction site (Fig. 2a) was a reduction in the specificity of the SbfI restriction enzyme resulting in more mistakes in the first and last positions of the restriction site. Previous studies have shown that specificity of restriction enzymes can be reduced in museum samples (Zimmermann et al., 2008). When restriction enzymes become less specific, alterations generally occur at the first and last recognition site positions (Polisky et al.,

1975). Given the elevated errors in only position 7 or 8, we argue that it is likely that the observed dependence of errors at position 8 on those at 7 was due to mistakes made by the restriction enzymes, and that these mistakes were more prevalent in the museum specimens. If the observed error pattern is related to decreased restriction enzyme specificity, then our results indicate that recognition errors at one position could affect an adjacent position. This seems reasonable given that the enzyme will change conformation given an interaction with a particular nucleotide in a particular site of the enzyme (Pingoud & Jeltsch, 2001).

Museum specimens accrue undesirable modifications over time, and there are several ways in which these specimens do not behave like contemporarily collected and preserved specimens. For example, while not rigorously investigated here, spectrophotometer measurements of the museum DNA exhibited lower 260/280nm and 260/230nm absorbance ratios indicating the presence of contaminants (Thermo Scientific, 2021). Non-specific digestion by the SbfI restriction enzyme could be caused by either misrecognition of chemically altered, degraded nucleotides or alterations to the enzyme itself due to interactions with molecules present in the museum DNA extracts. While we cannot propose a specific molecular model, alterations to the enzyme itself could lead to changes in the 3-dimensional conformation of the enzyme, affecting its recognition specificity. This could, in turn, affect multiple positions of its recognition sites, especially those on the ends which may be more susceptible to conformational changes. Alternatively, minor chemical changes to the museum nucleotides (Wei et al., 2008) may affect the interaction of the enzyme with the recognition site leading to misrecognition.

The elevated restriction enzyme error rate in museum specimens has practical implications because more DNA will be digested in museum samples than in contemporary due to misrecognition of non-target restriction sites, more RAD fragments will be created, especially

with single-digest methods, like that employed here. This, in turn, necessitates the need for more sequence to achieve the same depth of coverage in museum as contemporary samples. In the present study, we estimate based upon the proportion of reads with no errors at positions 7 or 8 (given the expected motif with no indels at the beginning; Table S4) that 24% more reads would be required for the *A. duodecimalis* museum specimens to overcome the additional errors and 7.3% more reads would be required for *S. spinus* museum specimens.

Conclusions

We were successful in producing genomic sequence data from 100-year-old museum samples, using reduced representation RAD sequencing, despite the complications associated with sequencing degraded DNA. While we can attribute lower base call quality scores in museum samples to the accumulation of chemical modifications and fragmentation, we do not believe this comprehensively explains observed disparities between museum and contemporary samples. Elevated base substitution error rates in the artificially synthesized oligonucleotide portions of museum reads suggests the presence of contaminants capable of altering enzymatic activity were likely to be present after DNA purification and several replicates of DNA purification that occur throughout library preparation. Elevated error rates in natural DNA (positions 7 & 8 of the SbfI restriction site) from museum specimens indicate that reduced restriction enzyme specificity is a likely cause. Further studies, including molecular, mass, or nuclear magnetic resonance spectroscopy are required to isolate and identify potential contaminants responsible for these effects. Key recommendations to improve sequencing results for museum specimens preserved in EtOH for over 100 years include filtering out read pairs with insertions and/or deletions at the beginning of the barcoded adapter sequence if they exhibit elevated base substitution error rates, and adding disproportionately more DNA from museum

libraries to increase per locus depth of coverage to account for increased error rates and reduced restriction enzyme specificity in museum samples.

REFERENCES

Abel, G. *Joint tests: What are they and why use them?*. (2013). Cambridge Centre for Health Services Research. [Online]. Retrieved from https://www.cchsr.iph.cam.ac.uk/131

Ali, O., O'Rourke, S., Amish, S., Meek, M., Luikart, G., Jeffres, C., & Miller, M. (2015). RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. Genetics, 202(2), 389-400. doi: 10.1534/genetics.115.183665

Altshuler, D., Pollara, V., Cowles, C., Van Etten, W., Baldwin, J., Linton, L., & Lander, E. (2000). An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature, 407(6803), 513-516. doi: 10.1038/35035083

Andrews, K., Good, J., Miller, M., Luikart, G., & Hohenlohe, P. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics, 17(2), 81-92. doi: 10.1038/nrg.2015.28

Baxter, S., Davey, J., Johnston, J., Shelton, A., Heckel, D., Jiggins, C., et al. (2011) Linkage Mapping and Comparative Genomics Using Next-Generation RAD Sequencing of a Non-Model Organism. PLoS ONE 6(4): e19315. https://doi.org/10.1371/journal.pone.0019315

Bioinformatics.babraham.ac.uk. (2021). Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. [Online]. Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Baird, N., Etter, P., Atwood, T., Currey, M., Shiver, A., & Lewis, Z. et al. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. Plos ONE, 3(10), e3376. doi: 10.1371/journal.pone.0003376

Bessetti, J. (2007). An introduction to PCR inhibitors. *J Microbiol Methods*, *28*, 159-67.

Bi, K., Linderoth, T., Vanderpool, D., Good, J., Nielsen, R. and Moritz, C. (2013). Unlocking the vault: next-generation museum population genomics. Molecular Ecology, 22(24), pp.6018-6032.

Bitinaite, J., & Schildkraut, I. (2002). Self-generated DNA termini relax the specificity of SgrAI restriction endonuclease. *Proceedings of the national academy of sciences*, *99*(3), 1164-1169.

Bolger, A., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114-2120. doi: 10.1093/bioinformatics/btu170

Burrell, A., Disotell, T., & Bergey, C. (2015). The use of museum specimens with highthroughput DNA sequencers. Journal of Human Evolution, 79, 35-44. doi: 10.1016/j.jhevol.2014.10.015

Burkner, P., Gabry, J., Weber, S., Johnson, A., & Mordrak, M. (2021). Bayesian Regression

Models using 'Stan'. [R package brms version 2.15.0]. Retrieved from https://cran.r-project.org/web/packages/brms/brms.pdf

Carpenter, K.E., Williams, J.T. & Santos, M.D. (2017) Acanthurus albimento, a new species of surgeonfish (Acanthuriformes: Acanthuridae) from northeastern Luzon, Philippines, with comments on zoogeography. Journal of the Ocean Science Foundation, 25, 33–46.urn:lsid:zoobank.org:pub:F6C86078-57C2-4DB3-8A67-4B70CFC7E317doi:http://dx.doi.org/10.5281/zenodo.291792

Catchen, J., Hohenlohe, P., Bassham, S., Amores, A., & Cresko, W. (2013). Stacks: an analysis tool set for population genomics. Molecular Ecology, 22(11), 3124-3140. doi: 10.1111/mec.12354

Chong, Z., Ruan, J., & Wu, C. (2012). Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. Bioinformatics, 28(21), 2732-2737. doi: 10.1093/bioinformatics/bts482

Cooper, A., Drummond, A., & Willerslev, E. (2004). Ancient DNA: Would the Real Neandertal Please Stand up?. Current Biology, 14(11), R431-R433. doi: 10.1016/j.cub.2004.05.037

Couture-Beil, A. (2021). Package 'rjson'. [R package rjson version 0.2.20]. Retrieved from https://cran.r-project.org/web/packages/rjson/rjson.pdf

Cribari-Neto F, Zeileis A (2010). "Beta Regression in R." *Journal of Statistical Software*, **34**(2), 1–24. doi: 10.18637/jss.v034.i02.

da Fonseca, R., Albrechtsen, A., Themudo, G., Ramos-Madrigal, J., Sibbesen, J., & Maretty, L. et al. (2016). Next-generation biology: Sequencing and data analysis approaches for non-model organisms. Marine Genomics, 30, 3-13. doi: 10.1016/j.margen.2016.04.012

DeChancie, J., & Houk, K. N. (2007). The origins of femtomolar protein-ligand binding: hydrogen-bond cooperativity and desolvation energetics in the biotin-(strept)avidin binding site. *Journal of the American Chemical Society, 129*(17), 5419–5429. https://doi.org/10.1021/ja066950n

Dziak, J., Coffman, D., Lanza, S. and Li, R., (2021). Sensitivity and specificity of information criteria. [Online]. Retrieved from https://www.methodology.psu.edu/files/2019/03/12-119-2e90hc6.pdf

Eckert, K. and Kunkel, T., (1991). DNA polymerase fidelity and the polymerase chain reaction. Genome Research, 1(1), pp.17-24.

Etter, P., Bassham, S., Hohenlohe, P., Johnson, E., & Cresko, W. (2011). SNP Discovery and Genotyping for Evolutionary Genetics Using RAD Sequencing. Methods In Molecular Biology, 157-178. doi: 10.1007/978-1-61779-228-1_9

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32(19), 3047-3048. doi: 10.1093/bioinformatics/btw354

Fox, J. (2021). car package | R Documentation. [R package car version 3.0-10]. Retrieved from https://rdocumentation.org/packages/car/versions/3.0-10

French, M. (2021). MartinThesis GitHub Repository. https://github.com/mfrench1/MartinThesis

Garrison E, Marth G. (2012) freebayes: Haplotype-based variant detection from short-read sequencing. [R package freebayes version 1.3-5.1]. Retrieved from https://github.com/freebayes/freebayes

Garrison E. (2012). Vcflib: A C++ library for parsing and manipulating VCF files. [R package vcflib version 1.0.2]. Retrieved from https://github. com/ekg/vcflib

Gilbert, M., Moore, W., Melchior, L. & Worobey, M., (2007). DNA Extraction from Dry Museum Beetles without Conferring External Morphological Damage. PLoS ONE, 2(3), p.e272.

Graham, C., Glenn, T., McArthur, A., Boreham, D., Kieran, T., & Lance, S. et al. (2015). Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). Molecular Ecology Resources, 15(6), 1304-1315. doi: 10.1111/1755-0998.12404

Green, R., Krause, J., Briggs, A., Maricic, T., Stenzel, U., & Kircher, M. et al. (2010). A Draft Sequence of the Neandertal Genome. Science, 328(5979), 710-722. doi: 10.1126/science.1188021

Grunwaldlab.github.io. (2019). AMOVA. [Online]. Retrieved from https://grunwaldlab.github.io/Population_Genetics_in_R/AMOVA.html

Hartl, D., & Clark, A. (2007). Principles of Population Genetics (4th ed.). Sunderland, Massachusetts: Sinauer Associates Inc. Publishers.

Hohenlohe, P., Bassham, S., Etter, P., Stiffler, N., Johnson, E., & Cresko, W. (2010). Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. Plos Genetics, 6(2), e1000862. doi: 10.1371/journal.pgen.1000862

Hykin, S., Bi, K., & McGuire, J. (2015). Fixing Formalin: A Method to Recover Genomic-Scale DNA Sequence Data from Formalin-Fixed Museum Specimens Using High-Throughput Sequencing. Plos ONE, 10(10), e0141579. doi: 10.1371/journal.pone.0141579

Illumina. (2010). Calling Sequencing SNPs. [Online]. Retrieved from https://www.illumina.com/Documents/products/technotes/technote_snp_caller_sequencing.pdf

Illumina. (2019). Sequencing Coverage for NGS Experiments. [Website]. Retrieved from https://www.illumina.com/science/education/sequencing-coverage.html

Illumina. (2021). Coverage Depth Recommendations. [Website]. Retrieved from https://www.illumina.com/science/technology/next-generation-sequencing/plan-experiments/coverage.html

Illumina. (2021). Quality Scores for Next-Generation Sequencing. [Online]. Retrieved from https://www.illumina.com/Documents/products/technotes/technote_Q-Scores.pdf

Kay, M. (2021). Tidy Data and 'Geoms' for Bayesian Models [R package tidybayes version 2.3.1]. Retrieved from https://cran.r-project.org/web/packages/tidybayes/index.html

Lenth, R. (2021). Estimated Marginal Means, aka Least-Squares Means [R package emmeans version 1.5.5-1]. Retrieved from https://cran.r-project.org/web/packages/emmeans/index.html

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., & Homer, N. et al. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078-2079. doi: 10.1093/bioinformatics/btp352

Lindahl, T., (1993). Instability and decay of the primary structure of DNA. Nature, 362(6422), pp.709-715.

Liu, X., Fu, Y., Maxwell, T., & Boerwinkle, E. (2009). Estimating population genetic parameters and comparing model goodness-of-fit using DNA sequences with error. Genome Research, 20(1), 101-109. doi: 10.1101/gr.097543.109

L.G. Riachi, Â. Santos, R.F.A. Moreira, C.A.B. De Maria, (2014). A review of ethyl carbamate and polycyclic aromatic hydrocarbon contamination risk in cachaça and other Brazilian sugarcane spirits. Food Chemistry, 149, 159-169, doi.org/10.1016/j.foodchem.2013.10.088.

Luca, F., Hudson, R., Witonsky, D., & Di Rienzo, A. (2011). A reduced representation approach to population genetic analyses and applications to human evolution. Genome Research, 21(7), 1087-1098. doi: 10.1101/gr.119792.110

Malyguine, E., Vannier, P., & Yot, P. (1980). Alteration of the specificity of restriction endonucle, ases in the presence of organic solvents. *Gene*, *8*(2), 163-177.

Mangiafico, S. (2021). R Handbook: Regression for Count Data. [Online]. Retrieved from https://rcompanion.org/handbook/J_01.html

Meyer, M., Kircher, M., Gansauge, M., Li, H., Racimo, F., & Mallick, S. et al. (2012). A HighCoverage Genome Sequence from an Archaic Denisovan Individual. Science, 338(6104), 222-226. doi: 10.1126/science.1224344

Miething, F., Hering, S., Hanschke, B., & Dressler, J. (2006). Effect of Fixation to the

Degradation of Nuclear and Mitochondrial DNA in Different Tissues. Journal Of Histochemistry & Cytochemistry, 54(3), 371-374. doi: 10.1369/jhc.5b6726.2005

Miller, M., Dunham, J., Amores, A., Cresko, W., & Johnson, E. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research, 17(2), 240-248. doi: 10.1101/gr.5681207

Nasri, M., Thomas, D., Alteration of the specificity of PvuII restriction endonuclease, *Nucleic Acids Research*, Volume 15, Issue 19, 12 October 1987, Pages 7677–7687, https://doi.org/10.1093/nar/15.19.7677

Naumann, E., Krzewińska, M., Götherström, A., & Eriksson, G. (2014). Slaves as burial gifts in Viking Age Norway? Evidence from stable isotope and ancient DNA analyses. Journal Of Archaeological Science, 41, 533-540. doi: 10.1016/j.jas.2013.08.022

NOAA. (2021). R/V Albatross I, 1882-1921. [website]. Retrieved from https://www.fisheries.noaa.gov/new-england-mid-atlantic/r-v-albatross-i-1882-1921

Overballe-Petersen, S., Orlando, L., & Willerslev, E. (2012). Next-generation sequencing offers new insights into DNA degradation. Trends In Biotechnology, 30(7), 364-368. doi: 10.1016/j.tibtech.2012.03.007

Pääbo, S. (1989). Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proceedings Of The National Academy Of Sciences, 86(6), 1939-1943. doi: 10.1073/pnas.86.6.1939

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., & Rohland, N. et al. (2004). Genetic Analyses from Ancient DNA. Annual Review Of Genetics, 38(1), 645-679. doi: 10.1146/annurev.genet.37.110801.143214

Park, S., Magee, D., McGettigan, P., Teasdale, M., Edwards, C., & Lohan, A. et al. (2015). Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. Genome Biology, 16(1). doi: 10.1186/s13059-015-0790-2

Pingoud, A. and Jeltsch, A., (2001). Structure and function of type II restriction endonucleases. Nucleic Acids Research, 29(18), pp.3705-3727. [Online]. Retrieved from https://pubmed.ncbi.nlm.nih.gov/11557805/

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2021). Linear and Nonlinear Mixed Effects Models [R package nlme version 3.1-152]. Retrieved from https://cran.r-project.org/web/packages/nlme/index.html

Polisky, B., Greene, P., Garfin, D., McCarthy, B., Goodman, H. and Boyer, H., (1975). Specificity of substrate recognition by the EcoRI restriction endonuclease. Proceedings of the National Academy of Sciences, 72(9), pp.3310-3314.

Post, R., Flook, P., & Millest, A. (1993). Methods for the preservation of insects for DNA studies. Biochemical Systematics And Ecology, 21(1), 85-92. doi: 10.1016/0305-1978(93)90012-g

Puritz, J., Hollenbeck, C., & Gold, J. (2014). dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. Peerj, 2, e431. doi: 10.7717/peerj.431

Ramey, C., (2021). The GNU Bourne-Again Shell. Tiscase.edu. [Online]. Retrieved from https://tiswww.case.edu/php/chet/bash/bashtop.html

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rehm, J., Baliunas, D., Borges, G., Graham, K., Irving, H., & Kehoe, T. et al. (2010). The relation between different dimensions of alcohol consumption and burden of disease: an overview. Addiction, 105(5), 817-843. doi: 10.1111/j.1360-0443.2010.02899.x

Ripley, B. (2021). MASS package | [R package MASS versions 7.3-53.1]. Retrieved from https://www.rdocumentation.org/packages/MASS/versions/7.3-53.1

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Research, 22(5), 939-946. doi: 10.1101/gr.128124.111

Rowe, H., Renaut, S., & Guggisberg, A. (2011). RAD in the realm of next-generation sequencing technologies. Molecular Ecology, doi: 10.1111/j.1365-294x.2011.05197.x

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., & Pääbo, S. (2012). Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. Plos ONE, 7(3), e34131. doi: 10.1371/journal.pone.0034131

Staden, R. (1979). A strategy of DNA sequencing employing computer programs. Nucleic Acids Research, 6(7), 2601-2610. doi: 10.1093/nar/6.7.2601

Stiller, M., Knapp, M., Stenzel, U., Hofreiter, M., & Meyer, M. (2009). Direct multiplex sequencing (DMPS)--a novel method for targeted high-throughput sequencing of ancient and highly degraded DNA. Genome Research, 19(10), 1843-1848. doi: 10.1101/gr.095760.109

Su, B., Wang, Y., Lan, H., Wang, W., & Zhang, Y. (1999). Phylogenetic Study of Complete Cytochrome b Genes in Musk Deer (Genus Moschus) Using Museum Samples. Molecular Phylogenetics And Evolution, 12(3), 241-249. doi: 10.1006/mpev.1999.0616

Thermofisher. 2021. Biotinylation | Thermo Fisher Scientific - UK. [Online]. Retrieved from https://www.thermofisher.com/us/en/home/life-science/protein-biology/protein-biology-

learning-center/protein-biology-resource-library/pierce-protein-methods/biotinylation.html

Tin, M., Economo, E., & Mikheyev, A. (2014). Sequencing Degraded DNA from NonDestructively Sampled Museum Specimens for RAD-Tagging and Low-Coverage Shotgun Phylogenetics. Plos ONE, 9(5), e96793. doi: 10.1371/journal.pone.0096793

Toonen, R., Puritz, J., Forsman, Z., Whitney, J., Fernandez-Silva, I., Andrews, K., & Bird, C. (2013). ezRAD: a simplified method for genomic genotyping in non-model organisms. Peerj, 1, e203. doi: 10.7717/peerj.203

Tretyakova, N., Groehler, A., & Ji, S. (2015). DNA–Protein Cross-Links: Formation, Structural Identities, and Biological Outcomes. Accounts Of Chemical Research, 48(6), 1631-1644. doi: 10.1021/acs.accounts.5b00056

Thermo Scientific. 2021. 260/280 and 260/230 Ratios. [Online]. Retrieved from https://www.uvm.edu/~vgn/microarray/documents/T042-NanoDrop-Spectrophotometers-Nucleic-Acid-Purity-Ratios.pdf

Verdugo, C., Kassadjikova, K., Washburn, E., Harkins, K., & Fehren-Schmitz, L. (2016). Ancient DNA Clarifies Osteological Analyses of Commingled Remains from Midnight Terror Cave, Belize. International Journal Of Osteoarchaeology, 27(3), 495-499. doi: 10.1002/oa.2550

Wei, H., Therrien, C., Blanchard, A., Guan, S., & Zhu, Z. (2008). The Fidelity Index provides a systematic quantitation of star activity of DNA restriction endonucleases. Nucleic Acids Research, 36(9), e50-e50. doi: 10.1093/nar/gkn182

Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). [Online] Retrieved from www.genome.gov/sequencingcostsdata

Wickham, H. (2021). Easily Install and Load the Tidyverse. Retrieved 23 March 2021, from https://tidyverse.tidyverse.org/

Wiegand, P., Domhöver, J., & Brinkmann, B. (1996). DNA-Degradation in formalinfixierten Geweben. Der Pathologe, 17(6), 451-454. doi: 10.1007/s002920050185

Wong, S., Li, J., Tan, A., Vedururu, R., Pang, J., & Do, H. et al. (2014). Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. BMC Medical Genomics, 7(1). doi: 10.1186/1755-8794-7-23

Woodbury, C., Jr, Hagenbüchle, O., & von Hippel, P. (1980). DNA site recognition and reduced specificity of the Eco RI endonuclease. The Journal of biological chemistry, 255(23), 11534–11548.

Zeileis, A., Cribari-Neto, F., Gruen, B., Kosmidis, I., Simas, A., & Rocha, A. (2021). betareg: Beta Regression. [R package betareg version 3.1-4]. Retrieved from https://cran.r-

project.org/web/packages/betareg/index.html

Zimmermann, J., Hajibabaei, M., Blackburn, D., Hanken, J., Cantin, E., Posfai, J., & Evans, T. (2008). DNA damage in preserved specimens and tissue samples: a molecular assessment. Frontiers In Zoology, 5(1), 18. doi: 10.1186/1742-9994-5-18

LIST OF APPENDICES

APPENDIX PAGE

APPENDIX

Supplementary Tables

Table S1. Results of tests for differences in the proportions of sequenced read pairs between collection times within each species. The read pairs were divided into categories based on whether they were filtered from the data set due to excessive low quality base calls or adapter sequences (Filtered), they did (Motif) or did not have the expected sequence motif composed

Beta Model by Species (Proportions)

Category	Species	Model Term	df_1	df_2	*F* Ratio	p.value
Filtered	*A. duo*	CollectionTime	1	Inf	1.3250	0.2497
	S. spi	CollectionTime	1	Inf	1.3660	0.2425
Motif	*A. duo*	CollectionTime	1	Inf	7.3110	**0.0069**
	S. spi	CollectionTime	1	Inf	57.8710	**< 0.0001**
No Motif	*A. duo*	CollectionTime	1	Inf	11.4250	**0.0007**
	S. spi	CollectionTime	1	Inf	45.6120	**< 0.0001**

Table S2. The number and proportions of read pairs for each combination of species (*A. duodecimalis*, *S. spinus*) and collection time (Museum, Contemporary) which were filtered from the data set due to inadequate length (< 75 bp) after trimming adapters and low quality bases.

	Museum		Contemporary		Total	
	Aduo	**Sspi**	**Aduo**	**Sspi**	**Aduo**	**Sspi**
Total # Read Pairs	6.27E+07	6.03E+07	5.86E+07	2.39E+08	1.21E+08	2.99E+08
# Read Pairs After Adapter Trim	5.32E+06	5.01E+07	7.44E+06	2.02E+08	1.28E+07	2.52E+08
# Read Pairs After Qual. & Adapter Trim	5.32E+06	5.01E+07	7.44E+06	2.02E+08	1.28E+07	2.52E+08
# Qual & Adapter Trimmed & Filtered	5.73E+07	1.02E+07	5.11E+07	3.66E+07	1.08E+08	4.68E+07
# Quality Trimmed & Filtered	9.17E+02	1.07E+04	7.59E+02	9.52E+03	1.68E+03	2.03E+04
# Adapter Trimmed & Filtered	5.73E+07	1.02E+07	5.11E+07	3.66E+07	1.08E+08	4.68E+07
% Remaining After Filtering	8.494%	83.078%	12.698%	84.657%	10.525%	84.339%
% Removed by Adapter Trim	91.504%	16.904%	87.300%	15.339%	89.473%	15.655%
% Removed by Quality Trim	0.001%	0.018%	0.001%	0.004%	0.001%	0.007%
% of Filtered Removed by Adapter Trim	99.998%	99.895%	99.999%	99.974%	99.998%	99.957%
% of Filtered Removed by Quality Trim	0.00%	0.11%	0.00%	0.03%	0.00%	0.04%

Table S3.1 Pairwise post-hoc contrasts between each indel category (Insertion, No Indels, Deletions) within each species (*A. duodecimalis, S. spinus*), collection time (Museum, Contemporary) and section of the target motif (Barcode, Ligation Site, End Cut Site). Bolding indicates statistical significance at a = 0.05.

Species	Contrast	Collection-time	Grouping	Estimate	SE	*df*	z.ratio	p.value
A. duo	Insertions - No Indels	Contemporary	Barcode	-0.7370	0.1232	Inf	-5.9830	**< 0.0001**
A. duo	Insertions - Deletions	Contemporary	Barcode	-0.5328	0.1746	Inf	-3.0520	**0.0068**
A. duo	No Indels - Deletions	Contemporary	Barcode	0.2041	0.1254	Inf	1.6280	0.3108
A. duo	Insertions - No Indels	Museum	Barcode	-0.5908	0.1948	Inf	-3.0330	**0.0073**
A. duo	Insertions - Deletions	Museum	Barcode	-0.8102	0.2883	Inf	-2.8100	**0.0149**
A. duo	No Indels - Deletions	Museum	Barcode	-0.2194	0.2252	Inf	-0.9740	0.9899
A. duo	Insertions - No Indels	Contemporary	Ligation Site	-1.1433	0.1921	Inf	-5.9500	**< 0.0001**
A. duo	Insertions - Deletions	Contemporary	Ligation Site	-0.2891	0.2566	Inf	-1.1270	0.7793
A. duo	No Indels - Deletions	Contemporary	Ligation Site	0.8542	0.1745	Inf	4.8960	**< 0.0001**
A. duo	Insertions - No Indels	Museum	Ligation Site	-1.4607	0.2168	Inf	-6.7380	**< 0.0001**
A. duo	Insertions - Deletions	Museum	Ligation Site	-1.0443	0.3274	Inf	-3.1900	**0.0043**
A. duo	No Indels - Deletions	Museum	Ligation Site	0.4163	0.2731	Inf	1.5250	0.3821
A. duo	Insertions - No Indels	Contemporary	End Cut Site	-2.3755	0.0712	Inf	-33.3730	**< 0.0001**
A. duo	Insertions - Deletions	Contemporary	End Cut Site	0.0093	0.0888	Inf	0.1050	1.0000
A. duo	No Indels - Deletions	Contemporary	End Cut Site	2.3848	0.0576	Inf	41.3930	**< 0.0001**
A. duo	Insertions - No Indels	Museum	End Cut Site	-2.4502	0.0991	Inf	-24.7300	**< 0.0001**
A. duo	Insertions - Deletions	Museum	End Cut Site	0.4335	0.1222	Inf	3.5460	**0.0012**
A. duo	No Indels - Deletions	Museum	End Cut Site	2.8836	0.0835	Inf	34.5340	**< 0.0001**
S. spi	Insertions - No Indels	Contemporary	Barcode	-1.9577	0.0647	Inf	-30.2480	**< 0.0001**
S. spi	Insertions - Deletions	Contemporary	Barcode	-1.9472	0.0662	Inf	-29.4320	**< 0.0001**
S. spi	No Indels - Deletions	Contemporary	Barcode	0.0106	0.0202	Inf	0.5240	1.0000
S. spi	Insertions - No Indels	Museum	Barcode	-1.0092	0.0914	Inf	-11.0400	**< 0.0001**
S. spi	Insertions - Deletions	Museum	Barcode	-2.7977	0.0989	Inf	-28.2860	**< 0.0001**
S. spi	No Indels - Deletions	Museum	Barcode	-1.7886	0.0429	Inf	-41.7000	**< 0.0001**
S. spi	Insertions - No Indels	Contemporary	Ligation Site	-2.2454	0.0903	Inf	-24.8650	**< 0.0001**
S. spi	Insertions - Deletions	Contemporary	Ligation Site	-1.5472	0.0910	Inf	-16.9940	**< 0.0001**
S. spi	No Indels - Deletions	Contemporary	Ligation Site	0.6983	0.0252	Inf	27.7480	**< 0.0001**
S. spi	Insertions - No Indels	Museum	Ligation Site	-3.2522	0.1150	Inf	-28.2710	**< 0.0001**
S. spi	Insertions - Deletions	Museum	Ligation Site	-1.5958	0.1071	Inf	-14.9030	**< 0.0001**
S. spi	No Indels - Deletions	Museum	Ligation Site	1.6564	0.0553	Inf	29.9250	**< 0.0001**
S. spi	Insertions - No Indels	Contemporary	End Cut Site	-1.5725	0.0551	Inf	-28.5580	**< 0.0001**
S. spi	Insertions - Deletions	Contemporary	End Cut Site	-1.8879	0.0562	Inf	-33.5760	**< 0.0001**
S. spi	No Indels - Deletions	Contemporary	End Cut Site	-0.3154	0.0159	Inf	-19.8010	**< 0.0001**
S. spi	Insertions - No Indels	Museum	End Cut Site	-1.7933	0.0748	Inf	-23.9700	**< 0.0001**
S. spi	Insertions - Deletions	Museum	End Cut Site	-2.7896	0.0775	Inf	-35.9870	**< 0.0001**
S. spi	No Indels - Deletions	Museum	End Cut Site	-0.9963	0.0275	Inf	-36.2370	**< 0.0001**

Table S3.2 Pairwise post-hoc contrasts between collection times (Museum, Contemporary) within each species (*A. duodecimalis*, *S. spinus*), indel category (Insertions, No Indels, Deletions) and each section of the target motif (Barcode, Ligation Site, End Cut Site). Bolding indicates statistical significance at a = 0.05.

Species	Contrast	Target Motif Section	Indels	Estimate	SE	*df*	z.ratio	p.value
A. duo	Museum - Contemporary	Barcode	Insertions	0.0936	0.2660	Inf	0.3530	0.7243
A. duo	Museum - Contemporary	Ligation Site	Insertions	0.8818	0.3100	Inf	2.8410	**0.0045**
A. duo	Museum - Contemporary	End Cut-Site	Insertions	1.9076	0.1840	Inf	10.3580	**< 0.0001**
A. duo	Museum - Contemporary	Barcode	No Indels	0.2398	0.1530	Inf	1.5660	0.1174
A. duo	Museum - Contemporary	Ligation Site	No Indels	0.5644	0.1690	Inf	3.3470	**0.0008**
A. duo	Museum - Contemporary	End Cut-Site	No Indels	1.8330	0.1470	Inf	12.4820	**< 0.0001**
A. duo	Museum - Contemporary	Barcode	Deletions	-0.1837	0.2900	Inf	-0.6340	0.5264
A. duo	Museum - Contemporary	Ligation Site	Deletions	0.1266	0.3430	Inf	0.3690	0.7121
A. duo	Museum - Contemporary	End Cut-Site	Deletions	2.3318	0.1720	Inf	13.5440	**< 0.0001**
S. spi	Museum - Contemporary	Barcode	Insertions	0.0214	0.2760	Inf	0.0780	0.9381
S. spi	Museum - Contemporary	Ligation Site	Insertions	0.4422	0.2870	Inf	1.5400	0.1237
S. spi	Museum - Contemporary	End Cut-Site	Insertions	0.6563	0.2690	Inf	2.4410	**0.0146**
S. spi	Museum - Contemporary	Barcode	No Indels	0.9700	0.2540	Inf	3.8180	**0.0001**
S. spi	Museum - Contemporary	Ligation Site	No Indels	-0.5646	0.2590	Inf	-2.1830	**0.0291**
S. spi	Museum - Contemporary	End Cut-Site	No Indels	0.4355	0.2540	Inf	1.7150	0.0863
S. spi	Museum - Contemporary	Barcode	Deletions	-0.8291	0.2570	Inf	-3.2230	**0.0013**
S. spi	Museum - Contemporary	Ligation Site	Deletions	0.3935	0.2550	Inf	1.5400	0.1235
S. spi	Museum - Contemporary	End Cut-Site	Deletions	-0.2454	0.2550	Inf	-0.9620	0.3359

Table S3.3 Pairwise post-hoc contrasts between each section of the target motif (Barcode, Ligation Site, End Cut Site) within each species (*A. duodecimalis*, *S. spinus*), collection time (Museum, Contemporary) and indel category (Insertions, No Indels, Deletions). Bolding indicates statistical significance at a = 0.05.

Species	Contrast	Collection-time	Indels	Estimate	SE	*df*	z.ratio	p.value
A. duo	Barcode - Ligation Site	Contemporary	Insertions	-0.2008	0.2258	Inf	-0.8890	1.0000
A. duo	Barcode - End Cut Site	Contemporary	Insertions	2.9261	0.1402	Inf	20.8720	**< 0.0001**
A. duo	Ligation Site - End Cut Site	Contemporary	Insertions	3.1269	0.2021	Inf	15.4710	**< 0.0001**
A. duo	Barcode - Ligation Site	Museum	Insertions	0.5874	0.2735	Inf	2.1480	0.0953
A. duo	Barcode - End Cut Site	Museum	Insertions	4.7401	0.2096	Inf	22.6100	**< 0.0001**
A. duo	Ligation Site - End Cut Site	Museum	Insertions	4.1527	0.2204	Inf	18.8450	**< 0.0001**
A. duo	Barcode - Ligation Site	Contemporary	No Indels	-0.6071	0.0310	Inf	-19.5670	**< 0.0001**
A. duo	Barcode - End Cut Site	Contemporary	No Indels	1.2876	0.0211	Inf	61.1110	**< 0.0001**
A. duo	Ligation Site - End Cut Site	Contemporary	No Indels	1.8947	0.0315	Inf	60.1460	**< 0.0001**
A. duo	Barcode - Ligation Site	Museum	No Indels	-0.2825	0.0992	Inf	-2.8480	**0.0132**
A. duo	Barcode - End Cut Site	Museum	No Indels	2.8807	0.0593	Inf	48.5990	**< 0.0001**
A. duo	Ligation Site - End Cut Site	Museum	No Indels	3.1632	0.0891	Inf	35.4840	**< 0.0001**
A. duo	Barcode - Ligation Site	Contemporary	Deletions	0.0429	0.2126	Inf	0.2020	1.0000
A. duo	Barcode - End Cut Site	Contemporary	Deletions	3.4683	0.1364	Inf	25.4330	**< 0.0001**
A. duo	Ligation Site - End Cut Site	Contemporary	Deletions	3.4253	0.1810	Inf	18.9260	**< 0.0001**
A. duo	Barcode - Ligation Site	Museum	Deletions	0.3532	0.3397	Inf	1.0400	0.8950
A. duo	Barcode - End Cut Site	Museum	Deletions	5.9838	0.2328	Inf	25.7080	**< 0.0001**
A. duo	Ligation Site - End Cut Site	Museum	Deletions	5.6305	0.2713	Inf	20.7550	**< 0.0001**
S. spi	Barcode - Ligation Site	Contemporary	Insertions	0.0227	0.1093	Inf	0.2080	1.0000
S. spi	Barcode - End Cut Site	Contemporary	Insertions	1.8348	0.0839	Inf	21.8710	**< 0.0001**
S. spi	Ligation Site - End Cut Site	Contemporary	Insertions	1.8121	0.1042	Inf	17.3880	**< 0.0001**
S. spi	Barcode - Ligation Site	Museum	Insertions	0.4435	0.1370	Inf	3.2360	**0.0036**
S. spi	Barcode - End Cut Site	Museum	Insertions	2.4697	0.1160	Inf	21.2830	**< 0.0001**
S. spi	Ligation Site - End Cut Site	Museum	Insertions	2.0262	0.1270	Inf	15.9520	**< 0.0001**
S. spi	Barcode - Ligation Site	Contemporary	No Indels	-0.2650	0.0178	Inf	-14.9280	**< 0.0001**
S. spi	Barcode - End Cut Site	Contemporary	No Indels	2.2200	0.0113	Inf	197.2100	**< 0.0001**
S. spi	Ligation Site - End Cut Site	Contemporary	No Indels	2.4850	0.0163	Inf	152.0850	**< 0.0001**
S. spi	Barcode - Ligation Site	Museum	No Indels	-1.7996	0.0512	Inf	-35.1660	**< 0.0001**
S. spi	Barcode - End Cut Site	Museum	No Indels	1.6855	0.0194	Inf	86.9680	**< 0.0001**
S. spi	Ligation Site - End Cut Site	Museum	No Indels	3.4851	0.0508	Inf	68.5410	**< 0.0001**
S. spi	Barcode - Ligation Site	Contemporary	Deletions	0.4227	0.0252	Inf	16.8040	**< 0.0001**
S. spi	Barcode - End Cut Site	Contemporary	Deletions	1.8941	0.0210	Inf	90.1170	**< 0.0001**
S. spi	Ligation Site - End Cut Site	Contemporary	Deletions	1.4713	0.0230	Inf	64.0690	**< 0.0001**
S. spi	Barcode - Ligation Site	Museum	Deletions	1.6454	0.0476	Inf	34.5520	**< 0.0001**
S. spi	Barcode - End Cut Site	Museum	Deletions	2.4778	0.0469	Inf	52.7940	**< 0.0001**
S. spi	Ligation Site - End Cut Site	Museum	Deletions	0.8325	0.0349	Inf	23.8630	**< 0.0001**

Table S4. Observed and expected error rates with .99 highest posterior density intervals for each species (*A. duodecimalis, S. spinus*) and collection time (Museum, Contemporary) for errors occurring at only position 7, (Only 7), only position 8 (Only 8), and at both 7 and 8 (BothErr). Bolding indicates expected error rates not within the expected .99 highest posterior density upper and lower limits for error rates.

Species	Time	Error Type	Observed	Upper	Lower	Expected
Aduo	Contemporary	neither_err	0.9620	0.9750	0.9420	0.9580
Aduo	Contemporary	only7	0.0054	0.0071	0.0040	**0.0108**
Aduo	Contemporary	only8	0.0270	0.0466	0.0151	**0.0305**
Aduo	Contemporary	both_err	0.0049	0.0108	0.0023	**0.0003**
Aduo	Museum	neither_err	0.8660	0.9070	0.8070	**0.7760**
Aduo	Museum	only7	0.0110	0.0158	0.0075	**0.0683**
Aduo	Museum	only8	0.0743	0.1260	0.0435	**0.1430**
Aduo	Museum	both_err	0.0449	0.0922	0.0207	**0.0126**
Sspi	Contemporary	neither_err	0.9780	0.9820	0.9730	**0.9680**
Sspi	Contemporary	only7	0.0025	0.0029	0.0022	**0.0114**
Sspi	Contemporary	only8	0.0129	0.0172	0.0098	**0.0206**
Sspi	Contemporary	both_err	0.0060	0.0089	0.0041	**0.0002**
Sspi	Museum	neither_err	0.9670	0.9740	0.9570	**0.9010**
Sspi	Museum	only7	0.0044	0.0055	0.0034	**0.0315**
Sspi	Museum	only8	0.0171	0.0242	0.0120	**0.0649**
Sspi	Museum	both_err	0.0113	0.0183	0.0069	**0.0023**

Table S5. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) out of sample estimators for the model of independence represented by (Eqs. 1-4) and the model of dependence represented by (Eq. 5) for positions 7 & 8 of the barcoded adapter-SbfI motif. The smaller the AIC or BIC, the better the fit. Bolding indicates models with better fit.

			Model of Dependence Between Positions 7 & 8								**Model of Independence Between Positions 7 & 8**					
Species	**Collection-time**	**Test**	**Est**	**Res.Df**	**RSS**	**Df**	**Sum of Sq**	**F**	**Pr(>F)**	**AIC|BIC**	**Est**	**RSS**	**Df**	**Sum of Sq**	**F**	**Pr(>F)**	**AIC|BIC**
Aduo	Museum	slope = 1	7.77E-04	4	7.52E-04	1	2.48E-05	1.32E-01	0.7348	**-30.877**	9.01E-03	8.96E-03	1	4.71E-05	0.021	0.8917	-16.01346
		intercept = 0	7.57E-04	4	7.52E-04	1	4.84E-06	2.57E-02	0.8804	**-31.501**	1.06E-02	8.96E-03	1	0.0016875	0.7534	0.4344	-16.63818
	Contemporary	slope = 1	2.93E-06	10	2.08E-06	1	8.50E-07	4.09E+00	0.0708	**-146.763**	2.49E-05	2.23E-05	1	2.59E-06	1.16E+00	0.3065	-118.2916
		intercept = 0	1.02E-05	10	2.08E-06	1	8.11E-06	3.90E+01	0.0001	**-145.308**	2.52E-05	2.23E-05	1	2.90E-06	1.2979	0.2812	-116.8369
Sspi	Museum	slope = 1	3.13E-04	16	2.37E-04	1	7.56E-05	5.10E+00	0.0382	**-145.19**	4.51E-03	2.95E-03	1	0.0015638	8.487	0.0102	-99.82341
		intercept = 0	2.37E-04	16	2.37E-04	1	6.58E-09	4.00E-04	0.9835	**-142.52**	3.22E-03	2.95E-03	1	0.00026711	1.4497	0.2461	-97.1523
	Contemporary	slope = 1	2.36E-04	27	2.35E-04	1	1.28E-06	1.48E-01	0.7038	**-251.74**	7.50E-04	7.35E-04	1	1.51E-05	0.5547	0.4628	-218.6214
		intercept = 0	2.36E-04	27	2.35E-04	1	1.57E-06	1.80E-01	0.6745	**-247.64**	1.50E-03	7.35E-04	1	0.00076654	28.173	< 0.0001	-214.5195