The utility of formalin-fixed tissues and allozyme supernatant for population genomics and considerations for combining capture- and RADseq-based SNP datasets

Kyle O'Connell[1], Kevin Mulder[2], Addison Wynn[1], Kevin de Queiroz[3], and Rayna Bell[4]

[1]National Museum of Natural History
[2]Universidade do Porto
[3]Smithsonian Institution
[4]California Academy of Sciences

March 17, 2021

Abstract

Until recently many historical museum specimens were largely inaccessible to genomic inquiry, but high-throughput sequencing (HTS) approaches have allowed researchers to successfully sequence genomic DNA from dried and fluid-preserved museum specimens. In addition to preserved specimens, many museums contain large series of allozyme supernatant samples but the amenability of these samples to HTS has not yet been assessed. Here, we compared the performance of a target-capture approach using alternative sources of genomic DNA from ten specimens of spring salamanders (Plethodontidae: Gyrinophilus porphyriticus) collected 1985–1990: allozyme supernatants, allozyme homogenate pellets, and formalin-fixed tissues. We designed capture probes based on double-digest restriction-site associated (RADseq) sequencing derived loci from seven of the specimens and assessed the success and consistency of capture and RADseq technical replicates. This study design enabled direct comparisons of data quality and potential biases among the different datasets for phylogenomic and population genomic analyses. We found that in phylogenetic analyses, all replicates for a given specimen clustered together, but in principal component space, RADseq replicates did not cluster with corresponding capture-based replicates. SNP calls were on average 18.3% different between technical replicates, but these discrepancies were primarily due to differences in heterozygous/homozygous SNP calls. We demonstrate that both allozyme supernatant and formalin-fixed samples can be successfully used for population genomic analyses and we discuss ways to identify and reduce biases associated with combining capture and RADseq data.

INTRODUCTION

Biological collections provide an invaluable window into the past, creating an irreplaceable record of biodiversity (Suarez & Tsutsui 2004; Yeates et al. 2016). Often, samples have been collected over long time periods from the same localities, providing time series to study evolution at multiple spatial and temporal scales (Splendiani et al. 2017; Schmitt et al. 2019; Schultz et al. 2020). Temporally sampled collections have been leveraged for a wide variety of applications (Holmes et al. 2016), including studying changes in allele frequency (Bi et al. 2013; 2019), disease incidence (Avila-Arcos et al. 2012; Muletz et al. 2014), microbiome composition (Heindler et al. 2018), geographic range shifts (Tingley et al. 2009), and changes in body size (Caruso et al., 2014). Most natural history specimens were collected before the routine preparation of tissue samples for genetic analysis, and thus while DNA sequencing has revolutionized the study of evolution, a large portion of collections are inaccessible to standard methods of genomic inquiry (Lamichhaney et al. 2019).

Nonetheless, recent advances in DNA extraction and library preparation have made DNA from museum samples more accessible (Paireder et al. 2013; Sørensen et al. 2016; Toutoiu et al. 2020). In particular, target

1

capture techniques have enabled the sequencing of fragmented and low-yield DNA from vertebrate dried museum skin and bone preparations, as well as dried invertebrate, plant, and fungal specimens (Rowe et al. 2011; Avila-Arcos et al. 2012; Blaimer et al. 2016; Sánchez Barreiro et al. 2017; Schmid et al. 2018; St Laurent et al. 2018; Leavitt et al. 2019; Tsai et al. 2019; Bakker et al. 2020). Researchers have only just begun to apply this class of methods to fluid-preserved specimens that are formalin- or ethanol-fixed, often with mixed success (Hykin et al. 2015; Ruane & Austin, 2017; Hedin et al. 2018; McGuire et al. 2018; Wood et al. 2018; Derkarabetian et al. 2019; Turvey et al. 2019; Lyra et al. 2020). Formalin damages DNA in several ways (Campos & Gilbert, 2012; Cook et al. 2014), making the extraction and sequencing of genomic DNA from formalin-fixed specimens more challenging than other specimen preparation types (Stuart et al. 2006; Pierson et al. 2020).

Past work using historical DNA (hDNA) samples (including formalin-fixed samples) has mostly focused on phylogenetic applications (Wall et al. 2014; Ruane & Austin, 2017; McGuire et al. 2018; Derkarabetian et al. 2019; Turvey et al. 2019), primarily using ultraconserved elements (UCEs). A few studies have explored SNP-based population genetic analyses (Bi et al. 2013; Tin et al. 2014; Ewart et al. 2019); however, generating SNPs for hDNA samples using reduced representation approaches such as RADseq can be problematic because highly fragmented DNA may result in high rates of random allelic dropout and few overlapping loci between samples (Burrell et al. 2015). DNA fragmentation is less of a concern for capture-based approaches, but UCEs usually generate fewer SNPs at the population level than RADseq (but see McCormack, Tsai, & Faircloth, 2016). Accounting for formalin-induced damage is also important in SNP-based studies because these mutations can influence downstream analyses (Axelsson et al. 2008; Bi et al. 2013). RAD-capture approaches can bridge this divide by using target capture to generate robust SNP datasets that also have phylogenomic applications (Ali et al. 2016). This family of methods (Rapture, Ali et al. 2016; RADcap, Hoffberg et al. 2016; hyRAD, Suchan et al. 2016; hyRAD-X, Schmid et al. 2017) has been successfully applied to hDNA by capturing historical and modern samples in a single experiment (Suchan et al. 2016; Crates et al. 2018; Schmid et al. 2018; Gauthier et al. 2020) and by combining data collected from hDNA samples using target-capture with data from modern samples generated with RADseq (Bakker et al. 2020).

In contrast to other hDNA methods, only a few studies have investigated the utility of DNA from allozyme supernatant samples for genomic applications (Arbetman & Premoli, 2011; Yuan et al. 2018). Until DNA sequencing became widely adopted, allozymes were the primary marker used in studies of genomic variation (Schlötterer 2004). Large series of frozen allozyme supernatant samples are archived in some museum collections (Table 1) and likely in investigators' individual research collections as well. Two studies have shown that high molecular weight DNA can be extracted from these frozen supernatant samples and nuclear, mitochondrial, and chloroplast markers sequenced using PCR-based methods (Arbetman & Premoli, 2011; Yuan et al. 2018). Yet the amenability of these samples to target-capture protocols has not yet been assessed and it is unknown whether the resulting sequence data would vary from that obtained from frozen tissue samples and/or formalin-fixed tissues.

Table 1: Information regarding allozyme supernatant samples in select natural history collections. Museum abbreviations are as follows: AMS: Australian Museum, Sydney NSW, Australia; FMNH: Field Museum, Chicago IL, USA; MVZ: Museum of Vertebrate Zoology, Berkeley CA, USA; TNHC: The University of Texas at Austin - Texas Natural History Collections, Austin TX, USA; USNM: United States National Museum of Natural History, Washington DC, USA. The following collections were consulted and indicate they do not have allozyme supernatants: MCZ: Museum of Comparative Zoology, Cambridge MA, USA; NCSM-Herpetology: North Carolina Museum of Natural Sciences, Raleigh NC, USA; AMNH-Herpetology: American Museum of Natural History, New York NY, USA; KU-Herpetology: Kansas University Biodiversity Institute and Natural History Museum, Lawrence KS, USA; SUI-Mammalogy, Ornithology, Herpetology: University of Iowa Museum of Natural History, Iowa City IA, USA; ANSP: Academy of Natural Sciences of Drexel University, Philadelphia PA, USA; CAS Herpetology: California Academy of Sciences, San Francisco CA.

2

Institution	Collection	Homogenates w/ voucher	Homogenates no voucher	Taxonomic and/or geographic notes
AMS	Herpetology	600	N/A	Australia/Pacific skinks
FMNH	Herpetology	20–30	N/A	Inger & Voris studies from South East Asia
MVZ	Herpetology Mammalogy	4700 700	N/A N/A	Americas, Italy, mostly plethodontids Americas, Papua New Guinea
TNHC	Genetic Diversity	250	650	Amphibians, mostly *Eurycea*
USNM	Herpetology	500	12,000	Mostly plethodontid salamanders, some reptiles including phrynosomatid lizards

This study aims to quantify the utility of allozyme supernatant and formalin-fixed tissue samples for SNP-based analyses using ten specimens of spring salamanders (*Gyrinophilus porphyriticus*). We use a target capture approach (RAD-based capture) with replicated sampling (allozyme supernatant, allozyme homogenate pellet, and formalin-fixed tissue) as well as standard RADseq-derived loci from frozen blood of seven of the same ten specimens. In particular we seek to investigate: (1) what factors predict success of allozyme supernatant and formalin-fixed samples to genomic inquiry? (2) How similar are SNP-calls for differently preserved technical replicates? (3) Do replicates for a given specimen yield consistent results across phylogenomic and population genomic analyses? 4) What potential biases are introduced by pooling target-capture and RADseq datasets?

MATERIALS AND METHODS

Sample selection

We chose 10 samples from the Amphibian and Reptile collection at the Smithsonian Institution's National Museum of Natural History for which replicated sampling of frozen allozyme supernatant (supernatant), frozen allozyme homogenate tissue pellet (pellet), frozen blood cells (blood), and formalin-fixed voucher specimens (formalin-fixed) were available. Frozen blood cells were used as the DNA source to develop a set of SNPs using the standard ddRADseq approach (see below). This sampling approach allowed us to compare the performance of allozyme supernatant and formalin-fixed tissues to the frozen tissue pellet. Although the pellet is composed of tissue that has been subjected to homogenization (grinding that might result in both mechanical and heat damage) it served as a point of reference for the other hDNA preservation types with the expectation that the pellet would produce consistently higher quality and more complete data using our target-capture method.

Specimens were collected by A. Wynn and Jeremy Jacobs between 1985 and 1990 from two sites in Virginia (VA) and Ohio (OH) for allozyme-based studies (Table S1). Salamanders were euthanized and prepared as voucher specimens 4 to 25 days after capture. Blood samples were collected from euthanized specimens with heparinized capillary tubes and centrifuged to separate blood cells from the serum (to be used for

3

polyacrylamide gel electrophoresis) and were subsequently stored at -70° C. Further, mixed tissue samples (including heart, liver, intestine, spleen, and muscle from the body wall) were taken from each specimen and stored at -70° C. In the case of two of the three larval samples used in this study, the entire animal except the head and forearms were used for allozyme analysis. Tissues were homogenized in distilled water at a ratio of 1 g tissue weight to 2 mL water with a powered grinder using a conically tipped, frosted-glass grinding head and mating grinding tube, and centrifuged for 20 min at 10,000 rpm at 0 to -5° C. The supernatant was decanted from the pellet, and both supernatant and pellet were subsequently stored at -70° C. We note that many researchers did not retain the tissue pellet at this stage of their investigation and only archived the supernatants. Consequently, this particular collection provided an important opportunity to directly compare these sources of DNA.

After tissue removal, voucher specimens were fixed in 10% formalin. Fixation time in formalin was not recorded but ranged from at least one day to a month or more. Specimens were then soaked in at least one change of 60–70% non-denatured ethanol, and afterwards stored in 70% non-denatured ethanol. Larval samples were stored in buffered formalin from the time of processing (1985 or 1988) until September 1998 when they were transferred to 70% denatured ethanol. Because liver tissue was exhausted from most specimens, we harvested ˜30 μg of muscle tissue (and also ˜30 μg of liver from USNM 525133) and stored the samples in RNA-later for 2–12 hours before freezing the tissues at -80°C. We extracted DNA from ˜100 ul of allozyme supernatant, ˜15 μg of frozen tissue pellet, ˜30 μg of formalin-fixed tissue, and for blood samples we digested all blood cells from a single ˜3cm capillary tube.

DNA extraction

DNA extraction for pellet, supernatant, and blood samples followed a standard salt-extraction protocol (Sambrook & Russell 2001). DNA samples were quantified using a QUBIT® 2.0 Fluorometer HS (Life Technologies, Grand Island, NY, USA) and representative extractions of each were checked on a 1% Agarose gel using 1–5 μl of DNA.

Our formalin-fixed extraction protocol followed Bell et al. (2017) with some modifications. Extractions were conducted in a dedicated PCR-free lab within a UV irradiated fume hood; all surfaces were pre-wiped with bleach and materials (plastics, pipettes etc.) were UV irradiated before use. We soaked ˜30 μg of tissue (muscle for all specimens plus liver for USNM 525133) in GTE Buffer for 24 hours, and repeated this for three washes (3 days total). We incubated for ˜4 days at 55°C in a 1.5 ml tube containing 500 μl Cell Lysis Buffer, 100 μl Proteinase K, and 20 μl of 1 mM DTT, adding 20 μl of Proteinase K per day as needed. We placed the samples on ice for 5 min, added 200 μl of Protein Precipitate Solution, inverted 50 times, and spun samples on a centrifuge at 14,000 g for 3 min. We poured the supernatant into a new 1.5 ml tube containing 600 ul of cold 100% ETOH and 3 μl glycogen solution. We again inverted the samples 50 times, incubated samples for 48 hours at -20°C, then spun samples at 14000g for 30 min. After discarding the supernatant, we added 200 ul of 70% ETOH, inverted 50 times, spun at 14000 g for 3 min, then drained and dried the samples for ˜3–6 hours. We resuspended samples in 40 μl Tris solution. If DNA yield was expected to be low, we conducted multiple extractions of the same sample and pooled the final product. One of the larval specimens (USNM 525132) yielded only 32 ng of total DNA (based on Qubit quantification) and we did not attempt to capture this sample. The full extraction protocol is available in Appendix I. Fragment distribution of three replicates of the supernatant and formalin-fixed extractions were quantified on a TapeStation (Agilent, Santa Clara, CA; Appendix II). Specifically, we compared the liver and muscle extractions of USNM 525133 to visualize the distribution of DNA fragments extracted from both tissue types.

RADseq sample preparation and target capture probe design

Double-digest RADseq libraries for seven of our samples were prepared as part of another study (KPM unpublished) following Peterson et al. (2012) with minor modifications. Samples were digested using the enzymes*SphI* and *EcoRI* (NEB, Ipswich, MA, USA) and size selected for fragments 450–550 bps following Hime et al. (2019). Samples were sequenced with 396 other samples on a single NovaSeq 6000 S4 PE150 run, resulting in an average of 8.5 million reads per sample. We analyzed these samples using ipyrad v.0.9.12

(Eaton & Overcast, 2020) with a clustering threshold of 0.95 and left all other settings at default. We selected all loci that were present in all samples and included at least one SNP, resulting in 32,547 RADseq loci of 132 bp in length. We selected sequences from one individual at random to serve as a reference and sent these data to Arbor Biosciences who further filtered loci by soft masking for simple repeats and low-complexity regions as well as loci with GC content $> 35\%$ and$< 50\%$, resulting in 14,426 of the original 32,547 loci. Arbor Biosciences randomly sampled 10,000 target loci from this remaining set and synthesized 20,000 - 80 bp baits with ˜2x tiling of the target 132 bp loci.

Genomic library preparation

DNA extractions for supernatant, formalin-fixed and pellet samples were sent directly to Arbor Biosciences for library preparation, capture, and sequencing. Samples arrived at Arbor Biosciences fully dried and were resuspended in 50 μl of Buffer EBT. Fragment distributions of a subset of the gDNA samples were checked via Bioanalyzer (Agilent, Santa Clara, CA). Samples were quantified via an intercalating dye assay, and gDNA samples were sonicated and size-selected to a fragment length < 300 bp. Up to 200 ng of sonicated gDNA, where available, was taken into Illumina TruSeq-style library preparation. Samples received 5–7 cycles of indexing amplification to add unique dual-indexes. Capture pools of five samples were prepared with 150 ng of gDNA libraries. Captures were performed at 62°C (to account for potential divergence of hDNA samples from bait sequence) overnight according to the myBaits v4 manual. After clean-up, capture pools were amplified for 12 cycles followed by a second round of capture with 3x the recommended amount of Block C (human cot-1; following McCartney-Melstad et al. 2016) with the rest of the parameters remaining the same. The final post-capture material was amplified for 10 cycles and sequenced on a partial lane of a NovaSeq 6000 S4 PE150 run, resulting in 113.6 million PE reads (˜3.8 million reads per library).

Bioinformatic pipeline

We first cleaned and filtered raw reads using Trim Galore! v.0.6.4 (Krueger, 2015) under default settings. We screened our cleaned reads for contamination using FastQ Screen v.0.14.1 (Wingett & Andrews, 2018) by calculating the proportion of reads that mapped to either the human genome (GRCh38) or a database of representative bacterial genomes (11k+ genomes) from the NCBI RefSeq database using bowtie2 v.2.3.5 (Langmead & Salzberg, 2012). We then used FastQ Screen (no hits command) to remove reads that mapped to either the human or bacterial genomes. We conducted all further informatics using the Python pipeline SECAPR v.1.1.15 (Andermann et al. 2018), which is an extension of the Phyluce pipeline (Faircloth, 2015), as well as custom Python scripts (github.com/kyleaoconnell22/gyro_hDNA. We generated data assemblies using two different methods (Fig. 1). First, we used SECAPR reference_assembly with the cleaned reads of all samples, mapping to our target RADseq loci (probe dataset) as our reference (min coverage = 5; mean length = 135 bp). After finding significant differences in SNP calls between the RADseq and capture-based technical replicates (see Results), we took a de novo assembly-based approach to generate a SNP matrix. We hypothesized that including flanking regions (sequence adjacent to segments targeted by probes) would help filter out potential paralogs and result in more consistent SNP calls between RADseq and capture-based replicates. We assembled the 10 supernatant samples into contigs using Abyss v.1.3.7 (kmer = 90; Simpson et al. 2009). We matched assembled contigs (excluding USNM 525251 which assembled poorly) to target sequences using the 'secapr find_target_contigs' command with min-coverage = 35 (contigs were much longer than target loci), min-identity = 80 and the keep-paralogs flag. We aligned recovered markers using MAFFT v.7.130b (no-trim and ambiguous flags; Katoh et al. 2002), then created a consensus of each alignment using the cons command (plurality 0.1, setcase 0.1) from the EMBOSS program (Madeira et al. 2019) run within the secapr reference_assembly script. We mapped the cleaned reads of all samples to this new reference (contig dataset) using the 'reference_assembly' script (min coverage = 5; mean length = 454 bp).

For both the probe and contig datasets, we corrected for formalin-induced deamination using mapDamage2 v.2.0.6 (Jónsson et al. 2013), a computational framework that tracks and quantifies DNA damage patterns in ancient and historical DNA sequence. To quantify nucleotide damage we calculated the average rate of $C > T$ and $G > A$ misincorporations for each datatype across the first 25 bp of each read calculated by mapDamage2.

5

For the probe dataset, we called SNPs from our corrected, unphased bam files using BCFtools v.1.9 (Danecek & McCarthy, 2017), using the mpileup command with maximum coverage of 1000x (-A flag). We further pruned putative SNPs within 3 bp of an indel and removed clusters of indels with five or fewer bp between them. We used VCFtools v.0.1.16 (Danecek et al. 2011) to remove SNPs with quality scores < 30, non-biallelic SNPs, minor allele frequency > 0.015, minor allele count = 2, and minimum coverage < 5x. We further filtered for allele balance (AB > 0.25 & AB < 0.75) and required at least 70% of samples present at a site. This approach yielded 19,601 high-quality SNPs for the probe dataset. We chose one SNP per locus with the least missing data, leaving 2337 unlinked SNPs from 37 samples. For most analyses we removed four formalin-fixed samples (two from each geographic region) with <10% of SNPs, resulting in a matrix of 33 samples and 2337 unlinked SNPs with 9.1% missing data (calculated following de Medeiros & Farrell (2018)).

Initial SNP calling in the contig dataset followed the same pipeline as above but to ensure comparability with the probe dataset, we used BEDTOOLS (Quinlan and Hall, 2010) and VCFtools to prune SNPs to those called from within the original coordinates of the RADseq loci, leaving 24,005 SNPs. We chose one SNP per locus, retaining the SNP with the least missing data, yielding a final matrix of 3997 unlinked SNPs for 33 samples with 21% missing data.

Calculating allelic mismatch between technical replicates

We calculated the proportion of shared alleles between technical replicates for all specimens using custom scripts (see Data Accessibility) and conducted this assessment using both the probe and contig SNP datasets. We required replicates to share at least 200 biallelic SNPs (400 alleles) with no missing data and sorted allelic mismatches for a given SNP into those with different homozygous calls (two allele differences) and those related to different heterozygous/homozygous calls (one allele difference). Further, in cases of heterozygous mismatches, we identified which replicate was heterozygous and which was homozygous (e.g., supernatant vs. RADseq). The number of samples in each comparison is shown in Table 2.

Phylogenomic and population genomic analyses

Using both the probe and contig SNP datasets, we evaluated the consistency of technical replicates in a variety of applications along the phylogeography-phylogenetics continuum (Edwards et al. 2016). Using both datasets, we explored the utility of replicates in a phylogenetic context, expecting to find that technical replicates of the same specimen would cluster together in the phylogeny. We used IQ-TREE v.1.6.12 (Nguyen et al. 2015) to construct a maximum likelihood phylogeny of concatenated SNPs with 100 ultrafast bootstrap replicates with the BIC best-fit model = TVM+F+R4 (based on IQ-TREE ModelFinder analysis).

Using both datasets, we explored model-free population structure estimation using principal component analyses (PCA) with the *dudi.pca* function implemented in 'Ade4' v.1.7.15 (Dray & Dufour, 2007) for all replicates and specimens. We also explored the impact of samples with high levels of missing data on PC space by filtering the contig and probe datasets to exclude replicates with <70% of SNPs (n = 27). Finally, to explore the effect of unequal missingness between capture-based replicates on PC space, we filtered the contig dataset to exclude RADseq replicates and pruned SNPs missing from >10% of samples, leaving 713 SNPs, as well as the supernatant replicate of USNM 525151 which was an outlier in PC space (n=25).

Using both datasets, we compared estimates of nucleotide diversity between technical replicates (capture-based formalin fixed, supernatant, and frozen pellet SNPs, and RADseq SNPs) following O'Connell et al. (2020) by implementing permutation tests. We randomly subsampled the data for each replicate type to a minimum and maximum number of samples using vcftools and estimated nucleotide diversity using the populations module in STACKS v.2.54 (Catchen et al. 2013). We repeated this procedure 100 times and plotted the distribution in R (R Core Team). We tested minimum and maximum sample values of 2-10, 5-6, and 5-10 to explore the sensitivity of our estimates to sample size variation. Because we observed very little variation between subsampling regimes, we only present results for minimum samples of five (number of formalin-fixed replicates minus 1) and maximum samples of 10 individuals (number of supernatant and pellet replicates). We observed significant differences in nucleotide diversity estimates between technical replicates

6

(see Results); thus to explore the impact of biases in levels of missing data between replicates we further filtered our data to SNPs present in 95% of samples (strict filtering; 298 SNPs). Further, for both datasets, we calculated the proportion of non-missing SNPs that were heterozygous/homozygous for each replicate type.

RESULTS

Fragment distribution of hDNA extractions

We quantified the DNA fragment distribution of three supernatant extractions and three formalin-fixed extractions (Appendix II). Fragment distribution of supernatant replicates ranged from 186 bp to 12.8 kbp. Fragment distributions of formalin-fixed extractions ranged from ˜150 bp to ˜400 bp. The formalin-fixed liver replicate of USNM 525133 had higher extraction yield (1100 ng) with a fragment distribution peak at ˜400 bp, while the muscle replicate yielded 200 ng of DNA with a fragment distribution centered at ˜185 bp.

DNA assembly method

Sequence assembly methods (probe versus contig) yielded important differences across analyses. Notably, in the probe dataset, we observed systematically higher differences in heterozygous/homozygous allele calls between capture and RADseq replicates than between capture-based replicates (Table 2; Fig. 3) but we did not observe this pattern in the contig dataset. We also inferred higher rates of heterozygosity in the probe dataset (Fig. S4). Further, only by using the contig dataset pruned to SNPs within the coordinates of the original RADseq loci were we able to accurately place all four technical replicates together in phylogenetic analyses (Fig. 4; S6). As such, we present results in the main manuscript for only the contig dataset (pruned to SNPs within the coordinates of the original RADseq loci) and present results for the probe dataset in the Supporting Information. See the Discussion for more on this subject.

Predictors of hDNA sample success

Allozyme supernatant replicates performed similarly to pellet replicates across metrics including depth of coverage, on-target mapping rates, and number of loci (Fig. 2A,C,D). Both replicates also recovered similar numbers of loci to the RADseq replicates (Fig. 2A), and a higher proportion of on-target reads (Fig. 2D). Overall, we inferred low rates of on-target mapping as expected with large amphibian genomes (Fig. 2D; Table S1), with a mean mapping rate of 12.77% (SD = 7.3) for supernatant replicates, 10.24% (SD = 0.8) for pellet, and a significantly lower rate of 4.31% (SD = 0.3) for RADseq (Tukey test adjusted $p < 0.01$). In the probe dataset the RADseq data had significantly higher coverage (Fig. S5).

By contrast, the formalin-fixed replicates had fewer loci on average (Fig. 2A), and significantly lower mean on-target mapping rates than the other replicates at 3.50% (SD = 1.1). Extraction yield was lower for formalin-fixed samples (Fig. 2A) and we found that at least ˜200 ng of extracted DNA was needed to have ˜50% of SNPs in the final matrix (Fig. 2A, S5A). Life stage was also qualitatively important, in that none of the larval specimens (n=3) gave high DNA yields from the formalin-fixed samples (Table S1) and the two we sequenced were high in exogenous DNA reads. As noted in the methods, larval specimens were stored in formalin for much longer periods of time than were metamorphosed specimens.

Historical DNA damage and contamination

Nucleotide damage estimates for formalin-fixed samples estimated by mapDamage2 were 3' G > A = 0.01 (SD = 0.008), and 5' C > T = 0.02 (SD = 0.01). Although some human and/or bacterial contamination was present in all samples, supernatant, pellet, and RADseq replicates had minor levels of exogenous DNA contamination with an average of 1.2% (SD = 0.29) 1.1% (SD = 0.5), and 0.71% (SD = 0.14) of reads mapping to non-target genomes respectively. By contrast, in formalin-fixed replicates mapping rates of reads to exogenous DNA ranged from 1.4%–81.1% (mean = 25%, SD = 29.2; Fig. 2B).

Allelic mismatch

7

Pairwise comparisons of technical replicates in the contig dataset showed that SNP calls were on average 18.3% (SD = 6.0) different between technical replicates (Table 2; Fig. 3A). When looking at only homozygous call differences, however, this value fell to 0.13% (SD = 0.1; Fig. 3B), suggesting that almost all allelic mismatches were driven by differences in whether a given SNP was called as homozygous or heterozygous. Supernatant and pellet replicates had the most shared SNP calls among capture-based replicates, with heterozygous differences of 11.9% (SD = 5.2) of SNPs on average, while formalin-fixed replicates differed 22.2% (SD = 4.7) from supernatant replicates and 18.0% (SD = 4.4) from pellet replicates. Capture-based replicates exhibited fairly consistent rates of mismatch to the RADseq replicates with mean differences of 20.1% (SD = 1.6), 22.9% (SD = 3.4), and 20.4% (SD = 1.1) in heterozygous/homozygous calls for supernatant, formalin-fixed, and pellet replicates respectively. This difference dropped to < 0.26% for non-heterozygous/homozygous differences. If differences in heterozygous/homozygous calls are random we would expect a roughly 50/50 balance of heterozygous versus homozygous calls between replicates. However, we identified a systematic imbalance between heterozygous and homozygous calls, where >85% (SD = 4.98) of differences between capture and RADseq replicates were explained by a heterozygous call in the target capture replicate and a homozygous call in the RADseq replicate (Fig. S2). This consistent and significant imbalance was not observed among the capture-based technical replicates although formalin-fixed replicates exhibited greater deviation and variance from 50/50 balance than other replicate types (Fig. S2).

Table 2: Results of allelic mismatch analyses. N = number of samples, Sites = the number of SNPs shared on average between technical replicates in each comparison. Het = the average percentage of sites that exhibit different heterozygous/homozygous calls for each comparison, Hom = the percentage of sites that exhibit a homozygous call difference on average for each comparison, Allele balance = the average percentage of heterozygous/homozygous differences that are heterozygous in the first replicate listed. RADseq comparisons are bolded to highlight differences.

Dataset	Comparison	N	Sites	Het	Hom	Allele balance
Contigs	Supernatant/Formalin-fixed	4	817	22.2	0.15	64.5
	Supernatant/Pellet	10	1011	11.9	0.06	47.0
	Formalin-fixed/Pellet	4	803	18.0	0.12	34.9
	Supernatant/RADseq	5	329	20.1	0.19	**85.7**
	Formalin-fixed/RADseq	3	661	22.9	0.25	**76.5**
	Pellet/RADseq	6	783	20.4	0.15	**84.8**
Probe	Supernatant/Formalin-fixed	4	1256	24.4	0.01	63.0
	Supernatant/Pellet	10	2059	16.3	0.02	49.6
	Formalin-fixed/Pellet	4	1699	21.1	0.01	37.0
	Supernatant/RADseq	5	2244	37.4	00.5	**89.5**
	Formalin-fixed/RADseq	3	1540	36.7	00.5	**81.0**
	Pellet/RADseq	6	2006	37.0	00.5	**87.0**

Phylogenomic and population genomic analyses

Maximum Likelihood analysis based on 3997 SNPs (contig dataset) inferred a phylogenetic split between Virginia and Ohio samples (Fig. 4). In all cases, technical replicates of the same specimen were placed together, even with levels of missing data as high as 87% (USNM 525251). Principle component analyses were strongly affected by levels of missing data (Fig. 5A) and by differences between RADseq and capture-based replicates (Fig. 5B). When analyzing all samples (3997 SNPs, n=32), PC 1 (24%) separated samples by geography, but with less separation of replicates with high levels of missing data (Fig. 5A). PC 2 (17%) separated replicates with high levels of missing data, as well as the RADseq and capture-based replicates to some degree. The filtered dataset in which samples with high missingness were removed (3997 SNPs, n=27) separated replicates by geography along PC 1 (29%) and separated RADseq and capture-based replicates along PC 2 (18%) (Fig. 5B). The contig dataset filtered to only capture-based replicates and loci shared

by 90% of samples (713 SNPs, n=25) split samples by geography along PC 1 (20%), and again by level of missingness along PC 2 (12%). The most divergent samples along PC 2 were all formalin-fixed samples, which had the highest amounts of missing data (Fig. S8).

Estimates of nucleotide diversity yielded similar values for supernatant and pellet replicates: supernatant = 0.35 (SD = 0.004), pellet = 0.34 (SD = 0.008), but significantly different values for formalin-fixed and RADseq replicates: formalin-fixed = 0.25 (SD = 0.007), RADseq = 0.26 (SD = 0004; Fig. 6). The strict SNP filtering regime (95% complete, 298 SNPs) reduced differences between the formalin-fixed and other capture-based replicates, but still inferred significant differences in estimates between RADseq and all capture-based replicates (Fig. S3A). Counts of heterozygous sites inferred similar levels of heterozygosity between the three capture replicates: supernatant (31.6% SD = 5.8), pellet (30% SD = 3.3), formalin-fixed (29.3% SD = 5.6). RADseq replicates had a significant homozygote bias (12.4% SD = 3.4) compared with the capture-based replicates (Fig. S4).

DISCUSSION

Genomic utility of allozyme supernatant samples

Allozyme supernatant replicates performed well across analyses and recovered similar amounts of data compared with the pellet replicates. The one exception was USNM 525251, for which the supernatant replicate recovered only 13% of SNPs, compared to 96% of SNPs for the pellet replicate and 82% of SNPs for the formalin-fixed replicate of the same sample. This replicate also received only 2x mean sequence coverage, compared to an average of 21x across the other supernatant samples, which we hypothesize was due to poor capture efficiency as the replicate had high extraction yield and low amounts of exogenous DNA.

Allozyme supernatant and tissue pellet samples had the lowest proportion of allelic differences between technical replicates (with the exception of USNM 525251). Correspondingly, these replicates clustered together in the phylogeny and PCA and generated similar estimates of nucleotide diversity. These results suggest that allozyme supernatant samples are amenable to sequence capture approaches and that data derived from them can be combined with data derived from typical frozen tissue samples without suffering many of the biases documented for formalin-fixed samples (Williams et al. 1999; Tang, 2006). Some museum collections house large series of allozyme supernatants (Table 1) that serve as valuable records of biodiversity but have not been fully leveraged in the genomic age. These collections may serve similar purposes as other hDNA samples, such as filling sampling gaps, or evaluating genomic variation through time (Wandeler, Hoeck, & Keller, 2007; Lopez, Turner, Bellis, & Lasky, 2020).

Predictors of formalin-fixed sample success

The amount of template DNA extracted from formalin-fixed samples weakly predicted the number of high-quality SNPs (>5x coverage; adjusted R^2=0.28), suggesting that ~200 ng of extracted template was needed to recover> 25% of SNPs at > 5x coverage (Fig. 2A; S1; S5). Thus, we recommend dividing extractions into multiple replicates (from several pieces of tissue) to extract more DNA from samples that give low DNA yield. However, a greater amount of extracted endogenous DNA does not necessarily ensure downstream success as a variety of factors can degrade DNA quality in formalin-fixed samples, including specimen age, exposure to UV, temperature, and length of formalin exposure (Hykin et al. 2015; Sawyer et al. 2012). Historical samples typically contain highly fragmented DNA (Pääbo, 1989; Ewart et al. 2019), and this could affect library preparation if most fragments are too short for target probes to bind efficiently, even if relatively high amounts of DNA were extracted. The large genomes of amphibians may also require higher extraction yields (~200 ng in this study) to successfully capture genome-wide targets (McCartney-Melstad et al., 2016), whereas studies of formalin-fixed reptiles have reported successful sequence capture with as little as 1–3 ng/µl (Hykin et al. 2015; Ruane & Austin, 2017).

In addition, formalin-fixed sample extractions may contain high levels of exogenous DNA, particularly when endogenous DNA yield is low. In the four formalin-fixed samples with <10% of SNPs, levels of exogenous sequence were all >30%, and as high as 81%. The other six samples yielded > 94% endogenous sequence,

9

suggesting that the level of exogenous sequence is a strong predictor of sample success. Rates of exogenous DNA from fluid-preserved specimens have not been quantified in many studies, but Hykin et al. (2015) found low rates of exogenous sequence in a formalin-fixed lizard (only 0.27% of reads). By contrast, Lyra et al. (2020) extracted DNA from ethanol-preserved frogs and identified a high proportion of bacterial reads (based on BLAST search), and a low fraction of endogenous sequence (<0.5% mapped to closely-related reference transcriptome). Thus, it remains an open question how much endogenous DNA should be expected from formalin-fixed extractions. Two of the samples in this study with high rates of contamination were larval samples that had been stored in formalin for several years. The other two samples were adult specimens, and we are uncertain if the contamination occurred prior to or during tissue subsampling, or if the tissue subsamples had such low usable DNA that any exogenous DNA present was preferentially amplified (Pääbo 1989).

Another factor that may impact sample outcomes is the tissue type used for extractions. Studies seeking to extract DNA from formalin-fixed samples typically sample liver or muscle tissue (Hykin et al. 2015, Ruane & Austin, 2017; Pierson et al. 2020). Hykin et al. (2015) compared extraction success between these two tissue types and extracted higher yields from the liver replicates of *Anolis* lizard samples. Ruane and Austin (2017) successfully extracted DNA from snake liver tissues, while Pierson et al. (2020) were unable to extract suitable DNA for PCR or library preparation from salamander tail muscle. Here we compared success between muscle and liver replicates of specimen USNM 525133. We inferred double the rate of human contamination in the liver replicate (6.3%) than in the muscle replicate (2.9%), but by all other measures the liver replicate outperformed the muscle replicate, including total DNA extracted, fragment length, total loci, total SNPs, and average coverage. Taken together, these results suggest that DNA in formalin-fixed specimens may remain better preserved in liver than in muscle tissue, but future studies could test this hypothesis with larger sample sizes and with samples of various ages.

Utility of allozyme supernatant and formalin-fixed samples for phylogenomic and population genomic analyses

Formalin-fixed samples clustered with their corresponding specimen replicates in the phylogeny, despite having above-average levels of missing data (Fig. 4; S6). This suggests that these samples are amenable to phylogenetic analyses. By contrast, formalin-fixed samples did not cluster with corresponding specimen replicates in PC space, likely due to higher levels of missing data. Smith et al. (2020) found that missing data at informative sites biased phylogenomic estimation when using historical samples. Similar unevenness in missing data may partially explain some of the separation between the formalin-fixed and supernatant/pellet replicates in PC space in our study. In particular, when we pruned our data to only capture-based replicates and retained loci shared by 90% of samples (713 SNPs, n=25), the amount of missing data seems to load heavily on PC2 (12%), with the greatest separation between the formalin-fixed samples with the largest amounts of missing data and all other replicates. The two formalin-fixed samples with< 20% missing data clustered with their corresponding replicates, further indicating that missing data is likely driving this pattern. Rates of allelic mismatch (13–28%) could also explain some of the variation in PC space between capture-based replicates, though Ewart et al. (2019) found that simulated rates of allelic mismatch up to 10% did not impact PCA analyses in their study. Likewise, Stronen et al. (2018) report a similar pattern between historical and contemporary sample genotypes using Illumina BeadChips, where samples cluster along PC1 by sample-type (historical vs. contemporary), and along PC2 by geography, suggesting that this pattern may be common in studies using historical samples and not restricted to only target-capture approaches (but see van der Valk et al. 2019).

Finally, we inferred significantly different estimates of nucleotide diversity in formalin-fixed replicates compared with the supernatant and pellet replicates, although removing sites with high levels of missing data accounted for some, but not all, of this discrepancy. A similar pattern was reported by Ewart et al. (2019), who found a bias towards homozygosity in historical samples leading to lower estimates of genetic diversity, and a bias in missing data towards heterozygous sites, which may partially explain why removing sites with more missingness gave higher estimates of genetic diversity for formalin-fixed samples. We inferred a

10

non-significant difference in heterozygosity among the capture-based replicates, suggesting that either allelic mismatch or biases in the types of sites that are missing may best explain variation in nucleotide diversity estimates. When sample sizes are large enough, estimating genetic diversity using only loci and individuals with low missingness will likely reduce these biases, although the appropriate threshold will likely vary by study. Based on the systematic biases we report here, we recommend future studies make thoughtful choices about data assembly and SNP filtering when using formalin-fixed samples for demographic parameter estimation.

Analyzing RADseq and target capture samples together

We compared two assembly strategies for generating SNP data, mapping directly to target sequence (probe dataset) and assembling reads into contigs and generating a pseudo-reference alignment (contigs dataset). Although many target-capture studies utilize a variation of the contig approach (Faircloth et al. 2016; Andermann et al. 2018), studies using variations of the RAD-capture approach often map reads directly to the target RADseq loci (Hoffberg et al. 2016; Lang et al. 2020), or a reference genome (Ali et al. 2016), although Suchan et al. (2016) tested variations of both approaches. We found that the contig method generated almost twice as many SNPs (3997 vs. 2337), even when pruning to only the coordinates of the original RADseq loci. However, the contig method produced higher rates of missing data (21% vs. 9.1%) which may be attributable to spurious mapping in the probe dataset.

When only comparing the target capture replicates, both the probe and contig datasets performed similarly in phylogenomic and population genomic analyses. However, when combining capture and RADseq replicates, these two assembly strategies diverged significantly. Notably, phylogenomic analysis with the probe dataset separated the RADseq replicates from the capture-based replicates (Fig. S6), whereas the contig dataset placed replicate samples from the same specimen together correctly in the phylogeny (Fig. 4). Similarly, the probe dataset contained elevated heterozygous and homozygous SNP call differences between capture and RADseq replicates (Fig. 3A, B), but this pattern was not observed in the contig dataset (Fig. 3C, D). We propose that this difference between probe and contig datasets is best explained by better read mapping to the longer reference sequences (454 vs. 132 bp), which may have excluded paralogous sequences more efficiently and increased accuracy of SNP calls. With both approaches we infer strong heterozygote bias in the capture-based replicates compared to the RADseq replicates, which may be explained by differences in genomic library preparation method. Unlike the standard library preparation for RADseq, DNA for target capture was not digested with restriction enzymes or size selected, thus all genomic DNA was available for capture. This could permit more paralogous sequences in the capture-based replicates despite our data assembly method, thus introducing spurious heterozygous calls at homozygous sites. This explanation would indicate that the RADseq data have the "correct" calls for heterozygous sites, and it may explain the difference in heterozygosity for the capture-based replicates between the probe and contig datasets. Alternatively, polymorphism at restriction cut sites (allelic dropout) in the RADseq data can produce spurious homozygous calls, whereas the capture-based sequence would have a "correct" heterozygous call (Luca, Hudson, Witonsky & Di Rienzo, 2011; Gautier et al. 2013; Puritz et al. 2014). We propose that both these processes may be contributing to the heterozygous/homozygous imbalance we observed with a homozygous bias in RADseq replicates causing underestimates of genetic diversity (Luca et al. 2011; Cariou, M., Duret, L., & Charlat, 2016; Heller et al., 2021).

Although the contig dataset ameliorated some of the biases caused by mapping directly to the target loci, analyses with the contig and probe datasets still separated the RADseq and capture-based replicates in PC space and generated significantly different estimates of nucleotide diversity, regardless of SNP and individual filtering regimes. Lang et al. (2020) pooled modern RADseq and historical target-capture samples (using a hyRAD approach) and also inferred clustering differences between library types. By removing variants potentially derived from hDNA deamination (CT/TC and AG/GA), Lang et al. (2020) found that DNA damage did not explain differential clustering of sample type. Taken together with our study, this suggests that certain biases in SNP calling may persist between RADseq and RAD-capture data, but some of these biases can be reduced by mapping to either a contig-based reference (this study; Suchan et al. 2015) or a

11

reference genome (Ali et al, 2016). Another solution for studies that include both hDNA and modern samples would be to use a capture approach on all samples rather than combining RADseq for modern samples and capture for hDNA samples, although this would increase project costs substantially (or a RADseq only approach as in Ewart et al., 2019).

Conclusions and remaining questions

We show here that allozyme supernatant and formalin-fixed samples can be used for both phylogenomic and many population genomic applications. In particular, allozyme supernatant replicates performed similarly to the frozen tissue pellet replicates in all analyses. On the other hand, only six of ten formalin-fixed samples recovered >25% of SNPs and were useful for analyses. The four formalin-fixed samples that failed had lower extraction yields and high levels of exogenous DNA, potentially corresponding to the amount of time larval samples were left in formalin (10+ years). We recommend that libraries derived from formalin-fixed DNA should be sequenced at greater depths, and multiple samples could be included for lineages or populations of interest in the event that some samples fail. We also document potential biases associated with combining RADseq and capture datasets in shared analyses, including biases in heterozygous SNP calls, clustering by replicate type rather than by specimen, and systematic differences in estimates of genetic diversity. We found that mapping reads to longer reference sequences derived from assembled contigs rather than mapping directly to RAD loci addressed some of these discrepancies. However, systematic biases between RADseq and capture replicates remained in our dataset and we caution that researchers should be aware of these issues especially for studies in which such a bias could impact the interpretation of results (e.g., inferring changes in heterozygosity through time).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

KAO, RCB, KdQ, and AW conceived of the study design. KPM generated ddRADseq data. AW contributed to tissue sub-sampling. KAO extracted DNA. KAO, RCB, and KPM planned analyses, KAO conducted all analyses. KAO wrote the first draft and all authors contributed to revising the manuscript.

DATA ACCESSIBILITY

Scripts for all analyses can found at https://github.com/kyleaoconnell22/gyro_project. Genomic sequence data is deposited on the SRA database (Upon Acceptance). Input files for all analyses are deposited on FigShare (Upon Acceptance).

REFERENCES

Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016). RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics* , *202* (2), 389–400.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215* , 403–410.

Andermann, T., Cano, Á., Zizka, A., Bacon, C., & Antonelli, A. (2018). SECAPR—a bioinformatics pipeline for the rapid and user-friendly processing of targeted enriched Illumina sequences, from raw reads to alignments. *PeerJ* , *6* , e5175.

Arbetman, M. P., & Premoli, A. C. (2011). Oldies (but goldies!): extracting DNA from cryopreserved allozyme supernatants. *Journal of Heredity* , *102* , 764–769.

Ávila-Arcos, M. C., Ho, S. Y., Ishida, Y., Nikolaidis, N., Tsangaras, K., Hönig, K., . . . & Willerslev, E. (2012). One hundred twenty years of koala retrovirus evolution determined from museum skins. *Molecular Biology and Evolution, 30* (2), 299-304.

Axelsson, E., Willerslev, E., Gilbert, M. T. P., & Nielsen, R. (2008). The effect of ancient DNA damage on inferences of demographic histories.*Molecular Biology and Evolution* , *25* (10), 2181–2187.

Bakker, F. T., Bieker, V. C., & Martin, M. D. (2020). Herbarium Collection-Based Plant Evolutionary Genetics and Genomics. *Frontiers in Ecology and Evolution* , *8* , https://doi.org/10.3389/fevo.2020.603948

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

Bi, K., Linderoth, T., Singhal, S., Vanderpool, D., Patton, J. L., Nielsen, R., . . . & Good, J. M. (2019). Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. *PLoS genetics* , *15* , e1008119.

Blaimer, B. B., Lloyd, M. W., Guillory, W. X., & Brady, S. G. (2016). Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. *PLoS one* , *11* , e0161531.

Burrell, A. S., Disotell, T. R., & Bergey, C. M. (2015). The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution* , 79, 35–44.

Campos, P. F., & Gilbert, T. M. (2012). DNA extraction from formalin-fixed material. In *Ancient DNA* (pp. 81–85). Humana Press.

Card, D. C., Perry, B. W., Adams, R. H., Schield, D. R., Young, A. S., Andrew, A. L., . . . & Rochford, M. R. (2018). Novel ecological and climatic conditions drive rapid adaptation in invasive Florida Burmese pythons. *Molecular Ecology* , *27* (23), 4744–4757.

Cariou, M., Duret, L., & Charlat, S. (2016). How and how much does RAD-seq bias genetic diversity estimates? *BMC Evolutionary Biology* , *16* , 1-8.

Caruso, N. M., Sears, M. W., Adams, D. C., & Lips, K. R. (2014). Widespread rapid reductions in body size of adult salamanders in response to climate change. *Global Change Biology* , *20* (6), 1751-1759.

Cook, S., Dodge, C., Morgan, R., & Sandusky, G. E. (2014). DNA/RNA degradation rate in long term fixed museum specimens. *Forensic Medicine and Anatomy Research* , *3* , 1.

Crates, R., Olah, G., Adamski, M., Aitken, N., Banks, S., Ingwersen, D., . . . & von Takach Dukai, B. (2019). Genomic impact of severe population decline in a nomadic songbird. *PLoS one* , *14* , e0223953.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . & McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics* , *27* , 2156-2158.

Danecek, P., & McCarthy, S. A. (2017). BCFtools/csq: haplotype-aware variant consequences. *Bioinformatics* , *33* (13), 2037–2039.

de Medeiros, B. A., & Farrell, B. D. (2018). Whole-genome amplification in double-digest RADseq results in adequate libraries but fewer sequenced loci. *PeerJ* , *6* , e5089.

Derkarabetian, S., Benavides, L. R., & Giribet, G. (2019). Sequence capture phylogenomics of historical ethanol-preserved museum specimens: Unlocking the rest of the vault. *Molecular Ecology Resources* , *19* ,

1531–1544.

Dray, S., & Dufour, A. B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* , *22* (4), 1–20.

Eaton, D. A., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* , *36* (8), 2592–2594.

Edwards, S. V., Potter, S., Schmitt, C. J., Bragg, J. G., & Moritz, C. (2016). Reticulation, divergence, and the phylogeography–phylogenetics continuum. *Proceedings of the National Academy of Sciences* , *113* (29), 8025–8032.

Ewart, K. M., Johnson, R. N., Ogden, R., Joseph, L., Frankham, G. J., & Lo, N. (2019). Museum specimens provide reliable SNP data for population genomic analysis of a widely distributed but threatened cockatoo species. *Molecular Ecology Resources* , *19* , 1578–1592.

Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* , *32* (5), 786-788.

Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhue, C., Pudlo, P., ... & Estoup, A. (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology* , *22* (11), 3165-3178.

Gauthier, J., Pajkovic, M., Neuenschwander, S., Kaila, L., Schmid, S., Orlando, L., & Alvarez, N. (2020). Museomics identifies genetic erosion in two butterfly species across the 20th century in Finland. *Molecular Ecology Resources* , *20* , 1191–1205.

Heindler, F. M., Christiansen, H., Frederich, B., Dettai, A., Lepoint, G., Maes, G. E., ... & Volckaert, F. A. (2018). Historical DNA metabarcoding of the prey and microbiome of trematomid fishes using museum samples. *Frontiers in Ecology and Evolution* , *6* , 151.

Hedin, M., Derkarabetian, S., Ramirez, M. J., Vink, C., & Bond, J. E. (2018). Phylogenomic reclassification of the world's most venomous spiders (Mygalomorphae, Atracinae), with implications for venom evolution. *Scientific Reports* , *8* (1), 1-7.

Heller, R., Nursyifa, C., Erill, G. G., Salmona, J., Chikhi, J., Meisner, J., Sand Korneliussen, T., Albrechtsen, A. (2021). A reference-free approach to analyze RADseq data using standard Next Generation Sequencing toolkits. 10.1111/1755-0998.13324

Hime, P. M., Briggler, J. T., Reece, J. S., & Weisrock, D. W. (2019). Genomic data reveal conserved female heterogamety in giant salamanders with gigantic nuclear genomes. *G3: Genes, Genomes, Genetics* , *9* (10), 3467-3476.

Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R., & Glenn, T. C. (2016). RAD cap: sequence capture of dual-digest RAD seq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources* , *16* (5), 1264–1278.

Holmes, M. W., Hammond, T. T., Wogan, G. O., Walsh, R. E., LaBarbera, K., Wommack, E. A., ... & Nachman, M. W. (2016). Natural history collections as windows on evolutionary processes. *Molecular Ecology* , *25* (4), 864–881.

Hykin, S. M., Bi, K., & McGuire, J. A. (2015). Fixing formalin: a method to recover genomic-scale DNA sequence data from formalin-fixed museum specimens using high-throughput sequencing. *PloS One* , *10* , e0141579.

Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L., & Orlando, L. (2013). mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* , *29* (13), 1682–1684.

Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. In *Bioinformatics for DNA sequence analysis* (pp. 49-64). Humana Press.

Krueger, F. (2015). Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, 516, 517.

Lang, P. L., Weiss, C. L., Kersten, S., Latorre, S. M., Nagel, S., Nickel, B., . . . & Burbano, H. A. (2020). Hybridization ddRADseq-sequencing for population genomics of nonmodel plants using highly degraded historical specimen DNA. *Molecular Ecology Resources, 20* , 1228–1247.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* , *9* (4), 357.

Lamichhaney, S., Card, D. C., Grayson, P., Tonini, J. F., Bravo, G. A., Napflin, K., . . . & Sackton, T. B. (2019). Integrating natural history collections and comparative genomics to study the genetic architecture of convergent evolution. *Philosophical Transactions of the Royal Society B* , *374* (1777), 20180248.

Leavitt, S. D., Keuler, R., Newberry, C. C., Rosentreter, R., & Clair, L. L. S. (2019). Shotgun sequencing decades-old lichen specimens to resolve phylogenomic placement of type material. *Plant and Fungal Systematics* , *64* (2), 237-247.

Lopez, L., Turner, K. G., Bellis, E. S., & Lasky, J. R. (2020). Genomics of natural history collections for understanding evolution in the wild. *Molecular Ecology Resources* , *20* , 1153-1160.

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

Lyra, M. L., Lourenco, A. C. C., Pinheiro, P. D., Pezzuti, T. L., Baeta, D., Barlow, A., . . . & Faivovich, J. (2020). High-throughput DNA sequencing of museum specimens sheds light on the long-missing species of the *Bokermannohyla claresignata* group (Anura: Hylidae: Cophomantini). *Zoological Journal of the Linnean Society* ,*190* , 1–21.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., . . . & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* , *47* , W636-W641.

McCartney-Melstad, E., Mount, G. G., & Shaffer, H. B. (2016). Exon capture optimization in amphibians with large genomes. *Molecular Ecology Resources* , *16* (5), 1084-1094.

McCormack, J. E., Tsai, W. L., & Faircloth, B. C. (2016). Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources* , *16* (5), 1189–1203.

McGaughran, A. (2020). Effects of sample age on data quality from targeted sequencing of museum specimens: what are we capturing in time? *BMC Genomics, 21* , 1–10.

McGuire, J. A., Cotoras, D. D., O'Connell, B., Lawalata, S. Z., Wang-Claypool, C. Y., Stubbs, A., . . . & Bi, K. (2018). Squeezing water from a stone: high-throughput sequencing from a 145-year old holotype resolves (barely) a cryptic species problem in flying lizards. *PeerJ* , *6* , e4470.

Muletz, C., Caruso, N. M., Fleischer, R. C., McDiarmid, R. W., & Lips, K. R. (2014). Unexpected rarity of the pathogen *Batrachochytrium dendrobatidis* in Appalachian *Plethodon* salamanders: 1957–2011. *PLoS one* , *9* (8), e103728.

Muletz-Wolz, C. R., Fleischer, R. C., & Lips, K. R. (2019). Fungal disease and temperature alter skin microbiome structure in an experimental salamander system. *Molecular Ecology, 28* , 2917-2931.

Newman, C. E., & Austin, C. C. (2016). Sequence capture and next-generation sequencing of ultraconserved elements in a large-genome salamander. *Molecular Ecology* , *25* (24), 6162–6174.

15

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* , *32* (1), 268–274.

Paabo, S. (1989). Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences* , *86* (6), 1939–1943.

Paireder, S., Werner, B., Bailer, J., Werther, W., Schmid, E., Patzak, B., & Cichna-Markl, M. (2013). Comparison of protocols for DNA extraction from long-term preserved formalin fixed tissues. *Analytical Biochemistry* , *432* , 152–160.

Pierson, T. W., Kieran, T. J., Clause, A. G., & Castleberry, N. L. (2020). Preservation-Induced Morphological Change in Salamanders and Failed DNA Extraction from a Decades-Old Museum Specimen: Implications for *Plethodon ainsworthi* . *Journal of Herpetology* , *54* , 137–143.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one* , *7* (5), e37135.

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., & Bird, C. E. (2014). Demystifying the RAD fad. *Molecular Ecology* , *23* (24), 5937-5942.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* , *26* , 841-842.

Rowe, K. C., Singhal, S., Macmanes, M. D., Ayroles, J. F., Morelli, T. L., Rubidge, E. M., . . . & Moritz, C. C. (2011). Museum genomics: low-cost and high-accuracy genetic data from historical specimens. *Molecular Ecology Resources* , *11* (6), 1082–1092.

Ruane, S., & Austin, C. C. (2017). Phylogenomics using formalin-fixed and 100+ year-old intractable natural history specimens. *Molecular Ecology Resources* , *17* , 1003–1008.

Sambrook, J., Russell, D. (2001) Molecular Cloning: A Laboratory Manual, 3rd Edition. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sanchez Barreiro, F., Vieira, F. G., Martin, M. D., Haile, J., Gilbert, M. T. P., & Wales, N. (2017). Characterizing restriction enzyme-associated loci in historic ragweed (*Ambrosia artemisiifolia*) voucher specimens using custom-designed RNA probes. *Molecular Ecology Resources* , *17* (2), 209–220.

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., & Paabo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PloS one* , *7* (3), e34131.

Schlotterer, C. (2004). The evolution of molecular markers—just a matter of fashion? *Nature Reviews Genetics* , *5* (1), 63–69.

Schmid, S., Neuenschwander, S., Pitteloud, C., Heckel, G., Pajkovic, M., Arlettaz, R., & Alvarez, N. (2018). Spatial and temporal genetic dynamics of the grasshopper Oedaleus decorus revealed by museum genomics. *Ecology and Evolution* , *8* (3), 1480-1495.

Schmitt, C. J., Cook, J. A., Zamudio, K. R., & Edwards, S. V. (2019). Museum specimens of terrestrial vertebrates are sensitive indicators of environmental change in the Anthropocene. *Philosophical Transactions of the Royal Society B* , *374* , 20170387

Shultz, A. J., Adams, B. J., Bell, K. C., Ludt, W. B., Pauly, G. B., & Vendetti, J. E. (2020). Natural history collections are critical resources for contemporary and future studies of urban evolution. *Evolutionary Applications* , *2020* , 1–15.

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J., & Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Research* , *19* (6), 1117-112.

16

Smith, B. T., Mauck III, W. M., Benz, B. W., & Andersen, M. J. (2020). Uneven missing data skew phylogenomic relationships within the lories and lorikeets. *Genome Biology and Evolution* , *12* (7), 1131-1147.

Splendiani, A., Fioravanti, T., Giovannotti, M., Olivieri, L., Ruggeri, P., Nisi Cerioni, P., ... & Caputo Barucchi, V. (2017). Museum samples could help to reconstruct the original distribution of Salmo trutta complex in Italy. *Journal of Fish Biology* , *90* , 2443–2451.

Sorensen, M., Rasmussen, A. R., & Simonsen, K. P. (2016). Enzymatic detection of formalin-fixed museum specimens for DNA analysis and enzymatic maceration of formalin-fixed specimens. In *Collection Forum* (Vol. 40, No. 1, pp. 1–6). Department of Psychology, Indiana University, Bloomington, Indiana 47405: The Society for the Experimental Analysis of Behavior.

St Laurent, R. A., Hamilton, C. A., & Kawahara, A. Y. (2018). Museum specimens provide phylogenomic data to resolve relationships of sack-bearer moths (Lepidoptera, Mimallonoidea, Mimallonidae). *Systematic Entomology* , *43* (4), 729-761.

Staats, M., Erkens, R. H., van de Vossenberg, B., Wieringa, J. J., Kraaijeveld, K., Stielow, B., ... & Bakker, F. T. (2013). Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PloS one* , *8* (7), e69189.

Stronen, A. V., Iacolina, L., Pertoldi, C., Kusza, S., Hulva, P., Dykyy, I., ... & Faurby, S. (2019). The use of museum skins for genomic analyses of temporal genetic diversity in wild species. *Conservation Genetics Resources* , *11* (4), 499-503.

Stuart, B. L., Dugan, K. A., Allard, M. W., & Kearney, M. (2006). Extraction of nuclear DNA from bone of skeletonized and fluid-preserved museum specimens. *Systematics and Biodiversity* , *4* (2), 133-136.

Suchan, T., Pitteloud, C., Gerasimova, N. S., Kostikova, A., Schmid, S., Arrigo, N., ... & Alvarez, N. (2016). Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PloS one* , *11* (3), e0151651.

Tin, M. M. Y., Economo, E. P., & Mikheyev, A. S. (2014). Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. *PloS one* , 9(5), e96793.

Tingley, M. W., & Beissinger, S. R. (2009). Detecting range shifts from historical species occurrences: new perspectives on old data. *Trends in Ecology & Evolution* , *24* (11), 625–633.

Totoiu, C. A., Phillips, J. M., Reese, A. T., Majumdar, S., Girguis, P. R., Raston, C. L., & Weiss, G. A. (2020). Vortex fluidics-mediated DNA rescue from formalin-fixed museum specimens. *PloS one* , *15* , e0225807.

Tsai, W. L., Schedl, M. E., Maley, J. M., & McCormack, J. E. (2019). More than skin and bones: Comparing extraction methods and alternative sources of DNA from avian museum specimens. *Molecular Ecology Resources* , *2019* , 1–8.

Turvey, S. T., Marr, M. M., Barnes, I., Brace, S., Tapley, B., Murphy, R. W., ... & Cunningham, A. A. (2019). Historical museum collections clarify the evolutionary history of cryptic species radiation in the world's largest amphibians. *Ecology and Evolution* , *9* (18), 10070–10084.

Wall, A. R., Campo, D., & Wetzer, R. (2014). Genetic utility of natural history museum specimens: endangered fairy shrimp (Branchiopoda, Anostraca). *ZooKeys* , *457* , 1.

Wandeler, P., Hoeck, P. E., & Keller, L. F. (2007). Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution* , *22* (12), 634–642.

Wingett, S. W., & Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* , *7* .

Wood, H. M., Gonzalez, V. L., Lloyd, M., Coddington, J., & Scharff, N. (2018). Next-generation museum genomics: Phylogenetic relationships among palpimanoid spiders using sequence capture techniques (Araneae: Palpimanoidea). *Molecular Phylogenetics and Evolution* , *127* , 907-918.

van der Valk, T., Diez-del-Molino, D., Marques-Bonet, T., Guschanski, K., & Dalen, L. (2019). Historical genomes reveal the genomic consequences of recent population decline in eastern gorillas. *Current Biology* , *29* (1), 165-170.

Vershinina, A. O., Kapp, J. D., Baryshnikov, G. F., & Shapiro, B. (2019). The case of an arctic wild ass highlights the utility of ancient DNA for validating problematic identifications in museum collections. *Molecular Ecology Resources* , *2020* , 1–9.

Yeates, D. K., Zwick, A., & Mikheyev, A. S. (2016). Museums are biobanks: unlocking the genetic potential of the three billion specimens in the world's biological collections. *Current Opinion in Insect Science* , *18* , 83–88.

Yuan, M. L., Wogan, G. O., & Wang, I. J. (2018). Trehalose improves PCR amplification of vertebrate nuclear DNA from historical allozymes. *Conservation Genetics Resources* , *10* , 313–315.

FIGURE LEGENDS

Figure 1: Bioinformatic workflow used to generate SNP datasets in this study. The probe dataset included SNPs generated by mapping directly to RADseq-based target loci, while the contig dataset was generated by creating a de novo reference from assembled contigs and pruning SNPs to only those present on the original RADseq loci.

Figure 2: Summary of capture success by sample replicate for the contig dataset of 3997 unlinked SNPs. A) Relationship between extracted DNA (log scale) and number of SNPs in final matrix. B) Relationship between extracted DNA and percent exogenous sequence for formalin-fixed samples, showing that samples below ~200 ng had high levels of exogenous sequence. C) Mean coverage by replicate. Formalin-fixed replicates had significantly lower coverage. D) Percentages of reads that mapped to target sequences by replicate. Formalin-fixed and RADseq replicates were significantly different from supernatant and pellet replicates, but not from one another.

Figure 3: Results of allelic mismatch analyses for probe (A,B) versus contig (C,D) datasets. A) Percent of SNPs with heterozygous/homozygous differences between technical replicates and B) Percent of SNPs with homozygous differences in the probe dataset with 2337 SNPs. C) Percent of SNPs with heterozygous/homozygous differences and D) Percent of SNPs with homozygous differences in the contig dataset with 3997 SNPs. Note that the probe dataset (A, B) shows consistently higher allelic differences in comparisons of RADseq and capture-based replicates whereas the contig dataset (C, D) shows similar allelic differences across all comparisons. Data for individual comparisons is given in full in Table 2. Abbreviations are S = supernatant, P = pellet, F = formalin-fixed, R = RADseq.

Figure 4: Maximum likelihood phylogeny with the highest log-likelihood for 3997 unlinked concatenated SNPs from the contig dataset estimated in IQ-TREE with 100 rapid bootstrap replicates. Gray circles represent nodes with >70% bootstrap support. Tip shapes represent replicate-type and are colored by geography (blue = Virginia, red = Ohio). Bars to the right of the phylogeny show the proportion of SNPs for each sample. All replicates cluster by specimen and by geography despite high levels of missing data in some replicates.

Figure 5: Principle component analysis (PCA) for the contig dataset. Data points are colored by geography (blue = Virginia, red = Ohio), shapes correspond to replicate-type, and size corresponds to data missingness, with larger shapes missing more data. Gray arrows highlight RADseq technical replicates. A) PCA of 33 samples and 3997 SNPs. Samples cluster by geography along PC1 and missing data and replicate type along PC2. B) PCA of 27 samples with >70% of the 3997 SNPs. Even with low levels of missing data, differences in clustering between RADseq and capture-based replicates are apparent.

Figure 6: Estimates of nucleotide diversity by replicate type for 3997 SNPs from the contig dataset based on 100 estimates with sample sizes ranging from five to ten individuals per replicate. Mean estimates of nucleotide diversity for formalin-fixed and RADseq replicates were significantly lower than supernatant and pellet replicates.

Supporting Information Figures:

Figure S1: Regression of extracted DNA against proportion of SNPs in the contig dataset for formalin-fixed samples.

Figure S2: Percentage of heterozygous differences between replicate pairs for the probe (A) and contig (B) datasets. The percentage reflects the percentage of heterozygous calls in the first named replicate type for a given comparison (e.g., percentage for S-P comparison is the percentage of discordant SNP calls for which the supernatant was heterozygous and the pellet was homozygous). The neutral expectation would be close to an even 50/50 split. Abbreviations are S = supernatant, P = pellet, F = formalin-fixed, R = RADseq.

Figure S3: A) Nucleotide diversity estimates from 298 SNPs present in> 95% of individuals from the contig dataset. B) Estimates of nucleotide diversity by replicate type for the probe dataset based on 2337 SNPs. Mean estimates of nucleotide diversity were significantly different for all comparisons
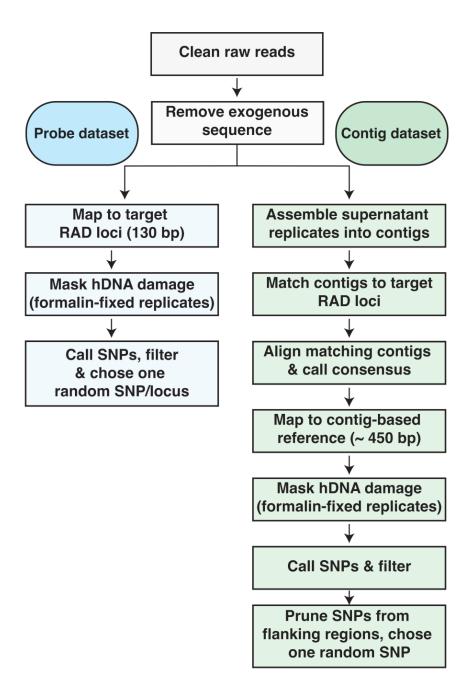
Figure S4: Percent of non-missing sites with heterozygous SNP calls for each replicate type for probe (A) and contig (B) datasets. Heterozygosity is reduced in the formalin-fixed replicates (but not significantly) compared with supernatant and pellet replicates. RADseq samples are significantly more homozygous. Note that the probe dataset has higher levels of heterozygosity in the capture-based replicates.
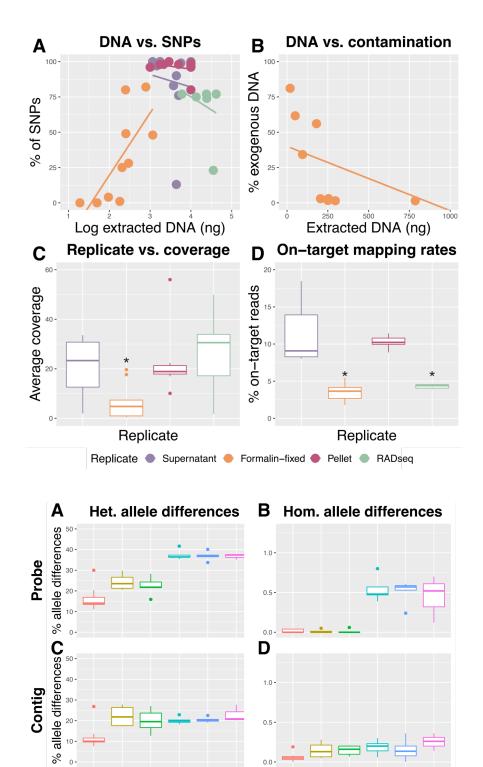
Figure S5: Summary of capture success by sample replicate for the probe dataset. A) Regression of extracted DNA (log scale) and number of SNPs in final matrix. B) Regression of extracted DNA and percent exogenous sequence for formalin-fixed samples, showing that samples above ˜200 ng had high levels of endogenous sequence. C) Mean coverage by replicate, RADseq replicates had significantly higher coverage than the capture-based replicates. D) Percentages of reads that mapped to target sequences by replicate, formalin-fixed replicates were significantly different from all other replicates.

Figure S6: Maximum likelihood phylogeny with the highest log-likelihood for 2337 unlinked concatenated SNPs from the probe dataset estimated in IQ-TREE with 100 rapid bootstrap replicates. Gray circles represent nodes with >70% bootstrap support. Tip shapes represent replicate-type and are colored by geography (blue = Virginia, red = Ohio). Bars to the right of the phylogeny show the proportion of SNPs for each sample. All capture samples cluster by replicate, except USNM 525251, which had high levels of missing data. Within geographic clade, all RADseq samples cluster together, except USNM 525139. This differs from the contig dataset show in Fig. 4. Replicates that do not cluster where expected are shown in bold.

Figure S7: Principle component analysis (PCA) for the probe dataset. Data points are colored by geography (blue = Virginia, red = Ohio), shapes correspond to replicate-type, and size corresponds to data missingness, with larger shapes missing more data. A) PCA of 33 samples and 2337 SNPs. Samples cluster by missing data and replicate type along PC1, and by geography along PC2. B) PCA of 27 samples with >70% of SNPs. Even with low levels of missing data, differences in clustering between RADseq and capture-based replicates persist.

Figure S8: Principle component analysis (PCA) of 713 SNPs and 25 samples from the contig dataset excluding RADseq replicates and pruned of SNPs missing from >10% of individuals. Samples cluster by geography on PC1 and by amounts of missing data along PC2, with formalin-fixed samples exhibiting the highest levels of missing data separated from the rest.
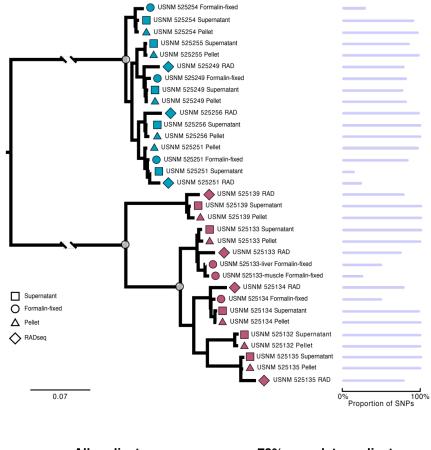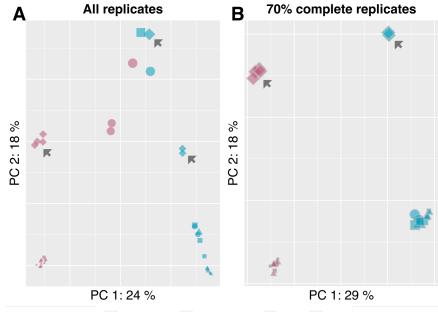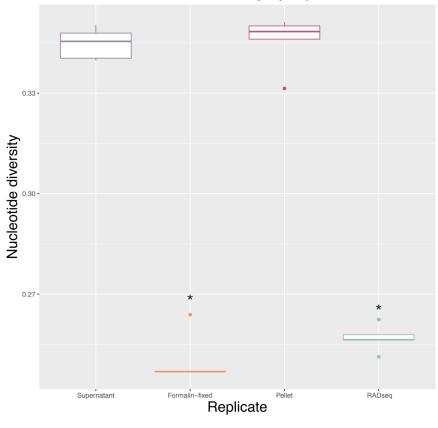
19

```
                        ┌─────────────────────┐
                        │  Clean raw reads    │
                        └─────────────────────┘
                                  │
                                  ▼
   ╭─────────────╮     ┌─────────────────────┐     ╭─────────────╮
   │ Probe       │     │ Remove exogenous    │     │ Contig      │
   │ dataset     │     │ sequence            │     │ dataset     │
   ╰─────────────╯     └─────────────────────┘     ╰─────────────╯
```

Probe dataset

- Map to target RAD loci (130 bp)
- Mask hDNA damage (formalin-fixed replicates)
- Call SNPs, filter & chose one random SNP/locus

Contig dataset

- Assemble supernatant replicates into contigs
- Match contigs to target RAD loci
- Align matching contigs & call consensus
- Map to contig-based reference (~ 450 bp)
- Mask hDNA damage (formalin-fixed replicates)
- Call SNPs & filter
- Prune SNPs from flanking regions, chose one random SNP

20

Nucleotide diversity by replicate