

Benchmarking ultra-high molecular weight DNA preservation methods for long-read and long-range sequencing

Hollis A. Dahn^{1,†}, Jacquelyn Mountcastle^{2,†}, Jennifer Balacco², Sylke Winkler³, Iliana Bista^{4,5}, Anthony D. Schmitt⁶, Olga Vinnere Pettersson⁷, Giulio Formenti², Karen Oliver⁴, Michelle Smith⁴, Wenhua Tan³, Anne Kraus³, Stephen Mac⁶, Lisa M. Komoroske⁸, Tanya Lama⁸, Andrew J. Crawford⁹, Robert W. Murphy¹, Samara Brown², Alan F. Scott¹⁰, Phillip A. Morin¹¹, Erich D. Jarvis^{12,12} and Olivier Fedrigo^{2,*}

¹Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario M5S 3B2, Canada

²The Rockefeller University, New York, NY 10065, USA

³Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Saxony 01307, Germany

⁴Tree of Life Program, Wellcome Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK

⁵Department of Genetics, University of Cambridge, Cambridge, Cambridgeshire CB2 3EH, UK

⁶Arima Genomics, Inc., San Diego, CA 92121, USA

⁷National Genomics Infrastructure, SciLifeLab, Uppsala University, Uppsala 75108, Sweden

⁸Department of Environmental Conservation, University of Massachusetts Amherst, Amherst, MA 01003-9285, USA

⁹Department of Biological Sciences, Universidad de los Andes, Bogotá 111711, Colombia

¹⁰Department of Medicine, Johns Hopkins University, Baltimore, MD 21287, USA

¹¹Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, La Jolla, CA 92037, USA

¹²Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

*Correspondence address. Olivier Fedrigo, Vertebrate Genome Laboratory, The Rockefeller University, 1230 York Avenue, Box 366, New York, NY 10065, USA. E-mail: ofedrigo@rockefeller.edu

†Co-first authors.

Abstract

Background: Studies in vertebrate genomics require sampling from a broad range of tissue types, taxa, and localities. Recent advancements in long-read and long-range genome sequencing have made it possible to produce high-quality chromosome-level genome assemblies for almost any organism. However, adequate tissue preservation for the requisite ultra-high molecular weight DNA (uHMW DNA) remains a major challenge. Here we present a comparative study of preservation methods for field and laboratory tissue sampling, across vertebrate classes and different tissue types.

Results: We find that storage temperature was the strongest predictor of uHMW fragment lengths. While immediate flash-freezing remains the sample preservation gold standard, samples preserved in 95% EtOH or 20–25% DMSO-EDTA showed little fragment length degradation when stored at 4°C for 6 hours. Samples in 95% EtOH or 20–25% DMSO-EDTA kept at 4°C for 1 week after dissection still yielded adequate amounts of uHMW DNA for most applications. Tissue type was a significant predictor of total DNA yield but not fragment length. Preservation solution had a smaller but significant influence on both fragment length and DNA yield.

Conclusion: We provide sample preservation guidelines that ensure sufficient DNA integrity and amount required for use with long-read and long-range sequencing technologies across vertebrates. Our best practices generated the uHMW DNA needed for the high-quality reference genomes for phase 1 of the Vertebrate Genomes Project, whose ultimate mission is to generate chromosome-level reference genome assemblies of all ~70,000 extant vertebrate species.

Keywords: long-read sequencing, genome assembly, tissue preservation, HMW DNA extraction

Introduction

The past 2 decades have seen genome sequencing become increasingly easy and affordable, driven by advancements in sequencing and computing technologies. Growing accessibility spurred the formation of large-scale consortia, such as the Genome 10 K project (G10K), with the goal of generating genome assemblies for many species to enable new scientific discoveries and aid in conservation efforts [1]. However, initial efforts used short-read sequencing (<200 bp), such as Illumina technology, which were later found to often result in genome assemblies that were highly fragmented, incomplete, and plagued with structural

inaccuracies [1–3]. Subsequently, G10K initiated the Vertebrate Genomes Project (VGP), with the mission of producing high-quality, near-complete, and error-free genome assemblies of all ~70,000 extant vertebrate species [4]. By comparing sequencing data types and assembly algorithms, the VGP consortium determined that it was not possible to obtain high-quality reference assemblies at the chromosomal level without the complementary use of multiple long-read sequencing technologies. Long reads (generally >10 kb; e.g., Pacific Biosciences, Menlo Park, CA, USA and Oxford Nanopore, Oxford Science Park, UK), long-range molecules (generally >50 kb; e.g., linked reads from 10X

Received: September 6, 2021. Revised: January 26, 2022. Accepted: June 16, 2022

© The Author(s) 2022. Published by Oxford University Press GigaScience. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Genomics, Pleasanton, CA, USA), or optical mapping (>150 kb; e.g., Bionano Genomics, San Diego, CA, USA) and Hi-C proximity ligation (>1 Mb; e.g., Arima Genomics, San Diego, CA, USA) can span repeats thousands of base pairs in length [4], greatly improving assembly outcomes. To take full advantage of these new sequencing and assembly methods, molecules of DNA need to be as long as possible.

While long-read and long-range (LR) data simplify and accelerate the assembly, they come with a major challenge: they require large amounts of very high-quality DNA. For short-read technologies, many nucleic acid isolation methods developed over the years, including the standard phenol–chloroform method [5], had been sufficient. LR technologies require relatively pure DNA in the 10 to 300 kb range. Additionally, the Hi-C method requires physical cross-linking of contacting DNA regions within the same chromosomes, thus requiring cell nuclei to be intact before processing and isolation of cross-linked DNA [4]. With Hi-C, 3-dimensional (3D) interactions within chromosomes serve to assemble contigs or short scaffolds into chromosomal-scale scaffolds. For LR technologies, only a few extraction methods are currently able to produce high-molecular weight (HMW) DNA ranging from 45 to 150 kb or ultra-high molecular weight (uHMW) DNA, which is over 150 kb long. These include bead-based (MagAttract HMW DNA Kit; Qiagen, Hilden, Germany), high-salt [6], and agarose plug methods (Bionano Prep Soft/Fibrous Tissue Protocol; Bionano Genomics) [7]. More recently, a less laborious thermoplastic magnetic disk (Nanobinds) method was developed by Circulomics (Baltimore, MD, USA) [8]. Regardless of their capabilities, the performance of HMW and uHMW DNA extraction methods primarily depends on the type of sample and how it was collected, handled, and preserved.

The long-held “gold standard” in tissue preservation for high-quality DNA isolation has been flash-freezing tissues in liquid nitrogen directly after collection, followed by ultra-cold -80°C long-term storage [9–14]. While liquid nitrogen is readily available in most laboratory setups, its limited availability in many fieldwork conditions can be an insurmountable hurdle. Indeed, a large portion of global biodiversity is located far from labs, and sampling such species will require long expeditions under rustic field conditions. Thus, transporting sufficient amounts of liquid nitrogen from the point of collection to the laboratory is often infeasible, and the applicability of flash-freezing outside the lab environment is greatly limited [10, 13, 15]. Additional considerations specific to the studied species exacerbate the challenge of sample collection and preservation. DNA degradation is promoted by enzymes whose concentrations are likely to be tissue specific and possibly species specific. Small organisms provide little tissue, and preferred tissue types may be unavailable. Permitting restrictions also vary widely among species and among countries. Yet, methods for field sampling in nonmodel species for the purposes of LR sequencing remain anecdotal or unsubstantiated, as failed attempts are not published and very few preservation experiments have measured fragment sizes relevant to LR technologies [16, 17]. Thus, methods that bridge the gaps between uHMW DNA, the lab, and field conditions still require benchmarking.

Here, we perform a series of benchmarking experiments to assess sample preservation methods under laboratory and simulated field conditions and compare the quality of uHMW DNA obtained. Specifically, we extract uHMW DNA from multiple tissue types of representative vertebrate species, which were collected under various preservation and temperature conditions. For each experimental sample, we evaluate the fragment length, yield, and purity of the uHMW DNA extracted. Based on our findings, we pro-

pose a new set of guidelines for tissue preservation, ranging from best to minimally adequate practices for acquiring uHMW DNA from both laboratory and field collected samples, necessary for producing high-quality reference genome assemblies.

Results

In this study, we used the agarose plug method optimized by Bionano Genomics [7] across all species and preservation methods, albeit with small protocol variations for fibrous tissues, soft tissues, and blood. We tested 6 preservation methods (Fig. 1): (i) flash frozen in liquid nitrogen, which served as the “gold standard” and our point of reference; (ii) 95% ethanol (EtOH), a long preferred method of field preservation of tissues [10, 15, 18]; (iii) 20–25% dimethyl sulfoxide (DMSO) buffer (see Methods), which has been shown to be very effective at permeating tissues and preserving HMW DNA after long-term storage at ambient temperature [19, 20]; (iv) RNAlater Stabilization Solution (RNAlater; Invitrogen, Waltham, MA, USA), a commonly used preservative that also facilitates transcriptomics; (v) DNAgard tissue and cells (DNAgard; Biomatrix, San Diego, CA, USA), a commercial preservative designed for stabilizing DNA in tissues at room temperature; and (vi) Allprotect Tissue Reagent (Allprotect; Qiagen), another commercial preservative targeting stable room-temperature tissue preservation. We exposed preserved samples to different temperatures (4°C , room temperature, and 37°C) for various durations of time (6 hours to 5 months). We did so with up to 6 tissue types (muscle, blood, ovary, spleen, isolated red blood cells [RBCs], and whole body) from 6 species representing 5 vertebrate lineages (a mammal, a bird, 2 turtles, an amphibian, and a bony fish; Fig. 1), for a total of 140 samples (Supplementary Table S1). We assessed the fragment length distribution and DNA yield for each DNA sample. Statistical analyses were performed using linear models that included type of preservative, temperature/time treatment, vertebrate group, and tissue type as variables.

Fragment length distribution analysis

For extractions that yielded a detectable amount of DNA, we measured their fragment length distributions using at least 1 of 2 available techniques: pulsed-field gel electrophoresis (PFGE) and the Agilent Femto Pulse system (Agilent, Santa Clara, CA, USA; FEMTO). PFGE was more informative for analyzing uHMW DNA molecules above 200 kb, due to greater dynamic range in molecular weight separation (Supplementary Fig. S1A), whereas FEMTO was more useful for separating molecules within the 50–165 kb range (Supplementary Fig. S1B). Overall, the agarose plug method yielded high-quality DNA concentrated in the 300–400 kb range (Fig. 2).

Temperature

From the linear modeling of both PFGE (Fig. 2, Supplementary Table S2) and FEMTO results (Supplementary Figs. S2 and S3), we found that temperature treatment was the predictor with the strongest evidence of an effect on the proportion of DNA fragments above 145 kb for PFGE ($df = 6$, LR $\chi^2 = 36.62$, $P = 2.09\text{e-}06$; Fig. 3a) and above 45 kb for FEMTO ($df = 8$, LR $\chi^2 = 44.80$, $P = 4.01\text{e-}07$; Supplementary Fig. S4A). Samples held at higher temperatures yielded a lower proportion of uHMW DNA, with flash-freezing performing best (Fig. 3A). However, samples refrigerated at 4°C for 6 hours following collection were statistically indistinguishable from flash-frozen samples (PFGE: $z = 0.56$, $P = 1.00$; FEMTO: $z = 2.03$, $P = 0.48$). Samples refrigerated at 4°C for longer periods of

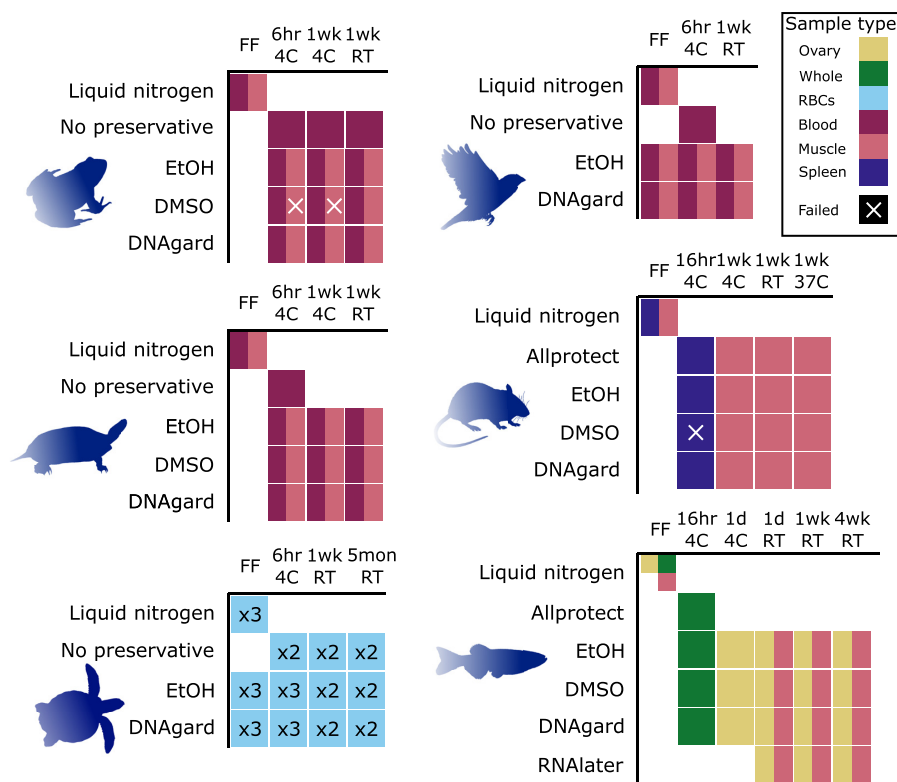


Figure 1: Experimental design for benchmarking tissue preservation. Graphical visualization of samples and treatments used in this study. Rows denote preservative treatments and columns temperature treatments. Colors indicate different types of tissue samples (see legend at top right). For the sea turtle samples, cells with numbers (x2 or x3) indicate conditions where samples from more than 1 individual were processed for comparison. All samples were transferred to -80°C after the specified temperature treatment (e.g., “6hr 4C” means stored at 4°C for 6 hours before transfer to -80°C). Abbreviations are as follows: RBCs, isolated red blood cells; EtOH, 95% ethanol; DMSO, a mix of 20–25% dimethyl sulfoxide, 25% 0.5 M EDTA, and 50–55% H_2O , saturated with NaCl. DNAGard, DNAGard tissue and cells (cat. no. #62001–046; Biomatrix); Allprotect, Allprotect Tissue Reagent (cat. no. 76405; Qiagen); RNAlater, RNAlater Stabilization Solution (cat. no. AM7021; Invitrogen); FF, flash-frozen in liquid nitrogen immediately upon dissection; 6hr, 6 hours; 1d, 1 day; 1wk, 1 week; 5mon, 5 months; RT, room temperature ($20\text{--}25^{\circ}\text{C}$). Samples were collected from these species: house mouse (*Mus musculus*), zebra finch (*Taeniopygia guttata*), Kemp’s Ridley sea turtle (*Lepidochelys kempii*), painted turtle (*Chrysemys picta*), American bullfrog (*Rana catesbeiana*), and zebrafish (*Danio rerio*).

up to 1 week showed some signs of degradation, albeit not consistently across tissue types and species (Fig. 2, Supplementary Figs. S2 and S3).

Preservation method

The predictor with the second strongest evidence of an effect on the proportion of DNA fragments above 145 kb or 45 kb was preservative treatment (PFGE: $df = 5$, $\text{LR } \chi^2 = 24.43$, $P = 0.0002$, Fig. 3B; or FEMTO: $df = 6$, $\text{LR } \chi^2 = 25.01$, $P = 0.0003$, Supplementary Fig. S4B, respectively). In PFGE measurements, significant differences were found between DNAGard and EtOH preservation ($z = 4.24$, $P = 0.001$, Fig. 3B), with DNAGard generally performing poorer. Flash-freezing and EtOH performed better than the other preservation methods in PFGE, and albeit not statistically significant, they had the lowest standard deviation (Fig. 3B). Based on PFGE, EtOH was slightly better than DMSO (Fig. 3B). Based on FEMTO, DMSO was slightly better than EtOH (Supplementary Fig. S4B). Neither relationship showed significant differences in preservation. In FEMTO measurements, flash-frozen and DMSO-preserved samples showed significantly better preservation efficiency than RNAlater samples (vs. DMSO: $z = 3.42$, $P = 0.009$; vs. flash-frozen: $z = 3.50$, $P = 0.007$), tested on fish samples. Allprotect outperformed EtOH in room temperature mouse samples but underperformed in the refrigerated fish body set (Fig. 2 and Supplementary Fig. S3).

Tissue type

Tissue type did not have a significant effect on fragment length overall (Fig. 3C, Supplementary Fig. S4c, Supplementary Table S2). However, muscle showed more variability than blood samples in uHMW DNA yield (>145 kb). The RBC samples showed the smallest proportion of degradation, while some muscle samples showed the highest degradation (Fig. 3C). In terms of variation between species, the mouse and fish samples showed a higher degree of degradation with respect to temperature treatment than the other species (Fig. 2, Supplementary Figs. S2 and S3). It is unclear if this can be explained by a species-specific temperature sensitivity or if it is caused by technical variation.

Interactions among variables

In terms of qualitatively assessing combinations of variables, storage in EtOH appeared to perform best at preserving uHMW DNA for all 4°C refrigerated samples (Fig. 2). Notably, nucleated blood samples refrigerated with no added preservatives were stable for up to 1 week with no substantial signs of degradation (Fig. 2). An increased proportion of smaller DNA fragments was evident in refrigerated samples preserved using DNAGard, with the exception of turtle RBCs and muscle samples, for which DNAGard results were equivalent to other preservation methods (Fig. 2). Fish body samples stored for 16 hours at 4°C showed notable degradation, but mouse spleen samples under the same treatment did not

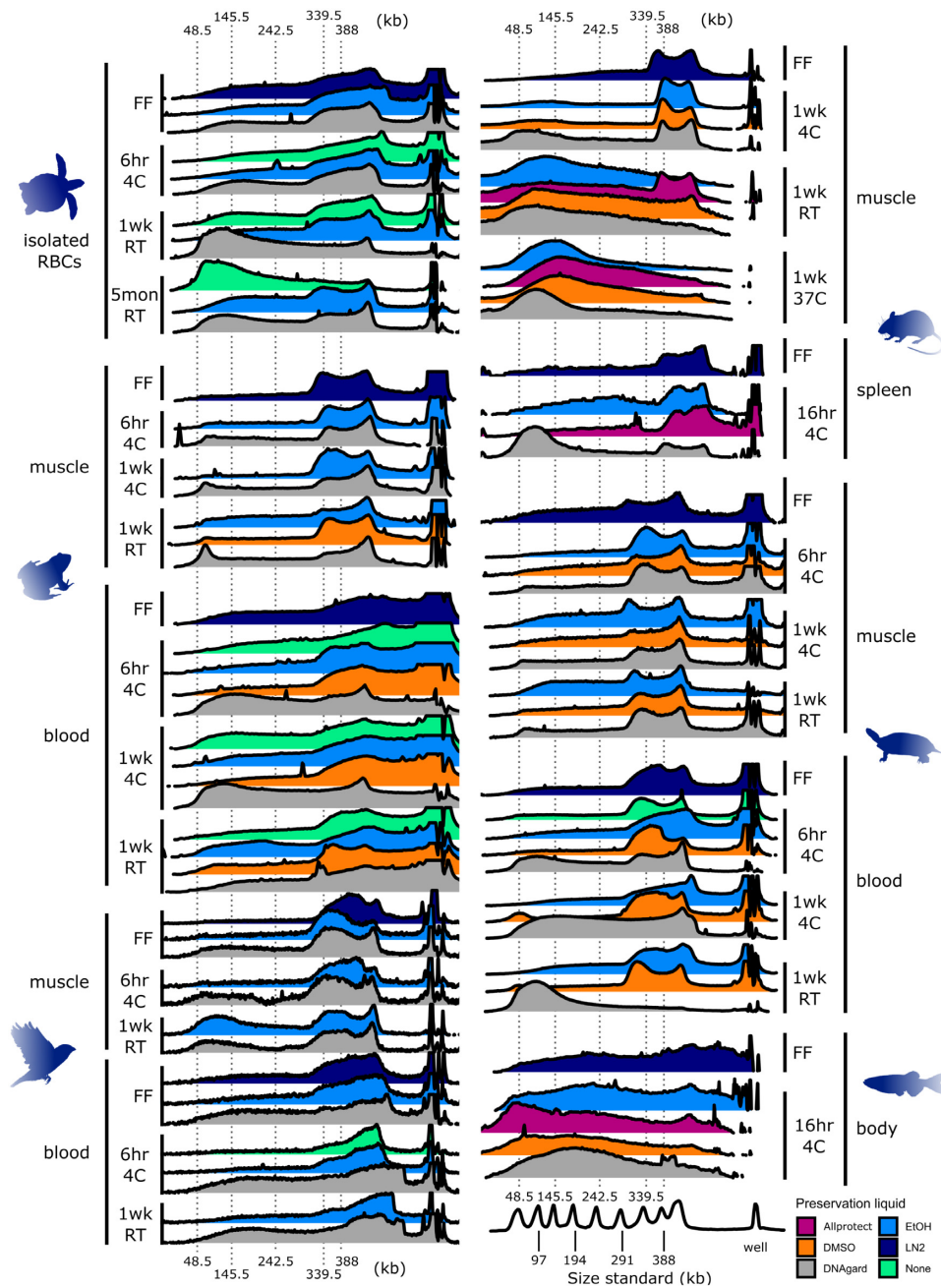


Figure 2: Pulsed-field gel electrophoresis (PFGE) measurements of uHMW DNA comparing different sample temperature and storage times. PFGE traces are visualized as overlapping ridgeline plots. Fluorescent-stained DNA fragments are drawn with an electric current from the well at the right toward the left. Smaller fragments generally travel farther than larger fragments. The fragments that greatly exceed the targeted size range remain in the well and cannot be reliably interpreted. Each ridgeline plot corresponds to a gel lane and a single DNA extract with brightness converted to a plot profile. The x-axis denotes molecule length scaled via piecewise linear scaling to match across gels of different lengths with a common size standard (Lambda PFG Ladder, New England Biolabs). The x-axis is the same in both columns. The y-axis of each plot is brightness scaled proportionally in each gel lane from just below the well to just beyond the 48.5-kb ladder peak such that the relatively intense brightness of the well itself is excluded from scaling. The well brightness is cropped where it exceeds the brightness of the rest of the gel lane. DNA fragments with lengths longer or shorter than peaks of the size standard cannot be reliably interpreted due to lack of size reference and artifacts of gel electrophoresis as well as limitations of any type of gel electrophoresis to correctly size megabase-length fragments. Colors represent different sample preservation methods, as indicated in the legend at bottom right. All samples were transferred to -80°C after the specified temperature treatment (e.g., "6hr 4C" means stored at 4°C for 6 hours before transfer to -80°C). Abbreviations are as follows: RBCs, isolated red blood cells; EtOH, 95% ethanol; DMSO, a mix of 20–25% dimethyl sulfoxide, 25% 0.5 M EDTA, and 50–55% H_2O , saturated with NaCl. DNAgard, DNAgard tissue and cells (cat. no. #62001–046; Biomatrix); Allprotect Tissue Reagent (cat. no. 76405; Qiagen); FF, flash-frozen in liquid nitrogen immediately upon dissection; 6hr, 6 hours; 1d, 1 day; 1wk, 1 week; 5mon, 5 months; RT, room temperature (20 – 25°C). Three additional samples were tested but produced insufficient DNA for fragment length analysis: frog muscle in DMSO for 1 week at 4°C and 6 hours at 4°C and mouse spleen in DMSO for 16 hours at 4°C . For measurements based on the FEMTO pulse instrument and additional tissue types, see Supplementary Figs. S2 and S3.

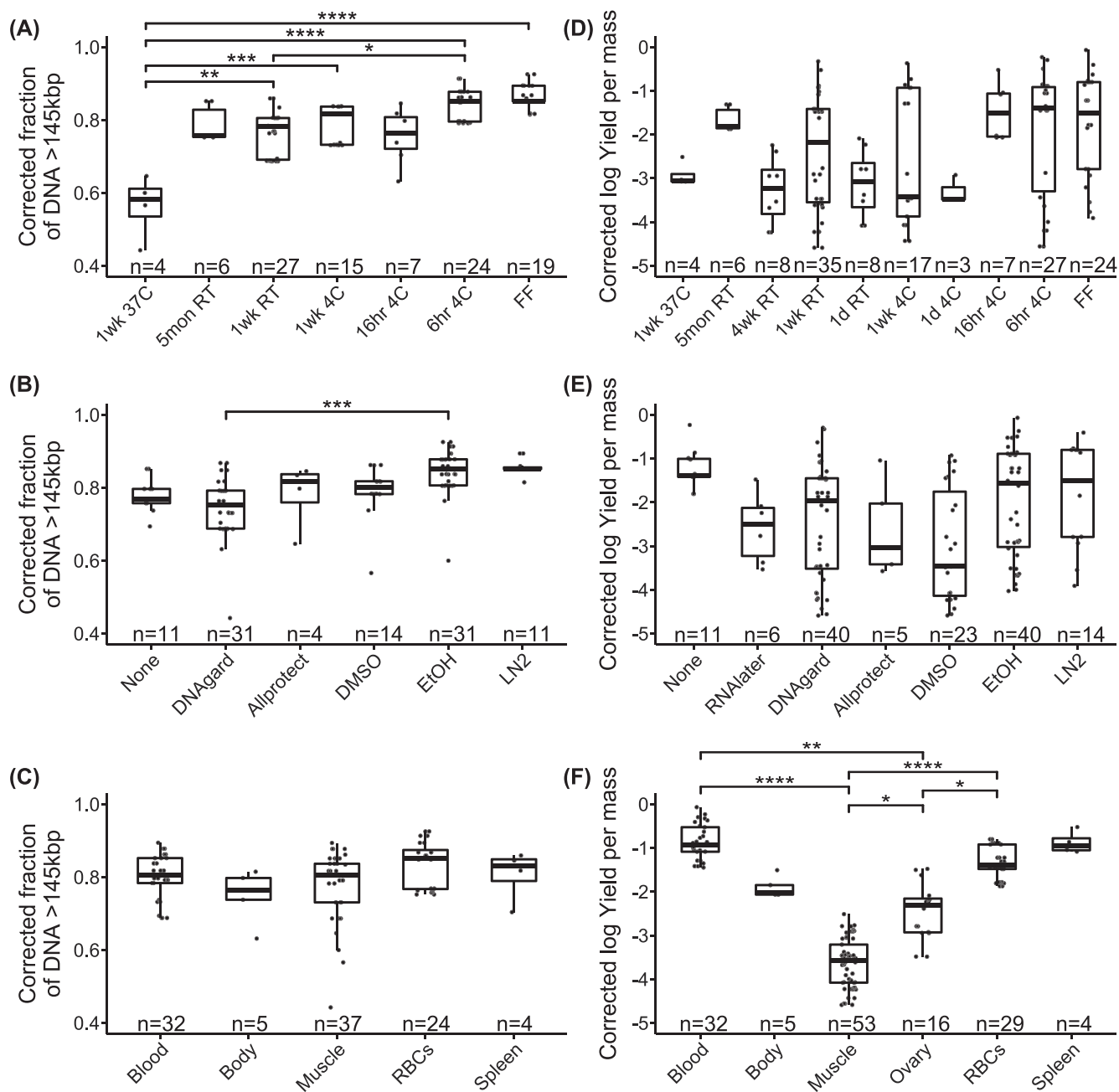


Figure 3: Testing the effect on two measures of uHMW DNA quality. Distributions of sample groups are overlaid with results of linear modeling of fragment length ($n = 102$, A–C) and DNA yield ($n = 139$, D–F). Shown are univariate scatterplots overlaid with box plots indicating the median, quartiles, and full range of individual observations. Fragment length was quantified here as the proportion of pulsed-field gel electrophoresis (PFGE) signal above 145 kb and was modeled in a generalized linear model with temperature (A), preservative (B), and tissue type (C) as predictors. DNA yield per input mass was log-transformed and modeled with temperature (D), preservative (E), tissue type (F), and vertebrate group as predictors. Significant relationships from post hoc comparisons are shown as connecting bars with significance levels: **** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$. Sample sizes for each factor are given along the x-axis.

vary substantially from samples stored at 4°C for 6 hours (Fig. 2). Replicate sea turtle RBC samples showed less variation within treatments for fragment size than for DNA yield (Supplementary Fig. S5A,B).

Mouse muscle, fish muscle, and fish ovary samples showed considerable accumulation of smaller fragment sizes after 1 week at room temperature, where blood or muscle samples from other species did not show as dramatic an impact (Fig. 2, Supplementary Figs. S2 and S3). However, fish muscle and ovary samples stored at room temperature for just 1 day still retained high proportions of

uHMW DNA with marginal degradation (Supplementary Fig. S2). For mouse muscle, DMSO, EtOH, or DNAgard did not seem to provide any added DNA protection against room temperature conditions (Fig. 2, Supplementary Fig. S3). At the same temperature conditions, mouse samples in Allprotect retained a nonnegligible fraction of uHMW DNA, although with some degradation (Fig. 2, Supplementary Fig. S3). Overall, similar to the 4°C exposure, room temperature DMSO and EtOH samples performed relatively well, albeit showing some signs of degradation. Surprisingly, 2 samples left at room temperature for 1 week without any preservative (sea

turtle RBCs and frog blood) were quite stable and yielded an appreciable fraction of uHMW DNA (Fig. 2). Additionally, sea turtle RBC samples, when preserved with EtOH or even DNAGard and stored at room temperature for 5 months, yielded a large fraction of workable uHMW DNA (Fig. 2). This suggested that turtle RBCs may be viable for longer durations at room temperature. Additional replicates and further experimentation will be necessary to determine if the isolated RBC tissue type or some biological difference in turtles is the key to this stability.

DNA yield

When the variables were tested individually, vertebrate group explained the least variance in DNA yield (3.69%, $df = 4$, $F = 3.25$, $P = 0.01$; Fig. 3D), temperature treatment explained a similarly small proportion (7.35%, $df = 9$, $F = 2.88$, $P = 4.25e-3$), preservative explained slightly more of the total variance (10.24%, $df = 6$, $F = 6.01$, $P = 1.73e-5$; Fig. 3E), and tissue type explained the largest amount of variance (46.35%, $df = 5$, $F = 32.65$, $P = 2.20e-16$; Fig. 3F). Both preservative and tissue type together explained 56.59% of the total variance (Supplementary Table S2). Specifically, whole blood tended to generate the highest DNA yields, followed by spleen, RBCs, whole body, and ovary, while muscle generated relatively lower yield (Fig. 3F). In post hoc tests, whole blood, RBCs, and ovary significantly outperformed muscle (vs. whole blood: $t = 11.75$, $P = 0.002$; vs. RBCs: $t = 8.36$, $P < 0.001$; vs. ovary: $t = 3.28$, $P = 0.01$), while the differences between muscle and whole body or spleen were not significant. Whole blood and RBCs also showed significantly higher yields than ovary samples (vs. whole blood: $t = 3.89$, $P = 0.002$; vs. RBCs: $t = 3.36$, $P = 0.01$). Post hoc comparisons of different temperature treatments or preservation reagents were not significant, possibly due to the higher variance influenced by the other variables of tissue type and species (Fig. 3D–F). Birds tended to have slightly better yields, with a marginally significant effect over nonavian reptiles ($t = 3.04$, $P = 0.02$).

Hi-C sequencing

The VGP is currently using Hi-C reads as a standard tool to generate chromosomal scale assemblies [4, 21], as well as to phase haplotypes in some cases [22]. These chromosome interactions are captured *in situ* in the tissue before DNA is isolated and sequencing libraries made. To enable appropriate collection recommendations for use in this technology, we also explore the effect of tissue preservation on the quality of the Hi-C library preparation. Using a single species (zebra finch), we test a subset of tissue preservation methods (flash-frozen, 6 hours at 4°C, 1 week at room temperature) and tissue types (muscle, blood), with 2 replicates per treatment combination. These were processed to generate *in situ* Hi-C chromatin interactions maps against the VGP male reference genome [23, 24].

We found that blood samples flash-frozen in EtOH yielded similar results compared to our flash-frozen positive control with no added preservative: 75–80% of all read-pairs were derived from *cis* interactions within the same chromosomes (Fig. 4A), and among them, ~55–60% were derived from long-range (> 15 kb) *cis* interactions. This indicates a high degree of useful long-range intrachromosomal signal necessary for genome assembly. However, storage of blood in DNAGard resulted in the elimination of almost all *cis* interactions, down to ~10% total, across temperature treatments (Fig. 4A–C), indicating largely random ligations and the loss of useful signal. Blood refrigerated for 6 hours maintained a high yield of long *cis* interactions, both when stored in EtOH and with no preservative. Blood samples stored at 1 week at room tempera-

ture in EtOH also yielded mostly long *cis* interactions similar to the flash-frozen treatments.

Overall, muscle and blood samples performed similarly across all treatments measured using Hi-C reads. They both yielded large amounts of long *cis* interactions (> 15 kb) when flash-frozen or refrigerated at 4°C with no preservative or with EtOH (Fig. 4A–E). Muscle and blood samples also responded similarly to preservative treatments, with EtOH samples performing well across treatments and DNAGard samples underperforming across treatments (Fig. 4).

Discussion

During development of the assembly pipeline for the first set of VGP genomes [4], we tested various HMW and uHMW DNA extraction protocols compatible with several LR technologies, including the Qiagen MagAttract HMW DNA, the phenol–chloroform method [5], and the agarose plug protocol. The agarose plug method, optimized by Bionano Genomics [7], was the most consistent method for producing a high yield of uHMW DNA suitable across all the LR technologies in the VGP pipeline. This method used agarose as a protective matrix to minimize DNA shearing during the extraction process and had long been shown to be an effective method for isolating megabase-size DNA from organisms, including plants, animals, algae, and microbes [7]. In this study, we use only the agarose plug DNA extraction method.

Our study explored the effects of three variables—preservation method, tissue type, and storage temperature—in preserving the high-quality DNA required for generating chromosome-scale genome assemblies in 6 species representing 5 major vertebrate lineages. The results identified promising alternatives to the standard flash-freezing method that is not easily performed in the field, particularly the preservation of samples in 95% ethanol (EtOH) or 20–25% DMSO-EDTA (DMSO) at 4°C.

We did not test all possible combinations of variables, which would require over 252 tests per species, but focused instead on the salient combinations of tissue types, reagents, and protocols that reflect real-world applications. There are also likely intervening stages of exposure to different temperatures, such as immediately postmortem, that may have a considerable effect in hotter climates and are not simulated here. Additionally, we are only able to visualize DNA fragment size distributions within a certain range of sizes (approx. 40–400 kb for PFGE, 1.3–165 kb for FEMTO). Although we have targeted a size range that includes both ideal fragment sizes for long-read sequencing and fragments of lower molecular weight that may indicate degradation, fragments outside this range are not measured here. Despite these limitations, our results are consistent with samples from the over 136 species we have processed for the VGP to date (NCBI Bioproject PRJNA489243 as of 13 July 2021). We believe that the results presented here can inform the many logistical decisions of field researchers collecting samples from wild populations (Fig. 5).

Temperature exposure was the strongest predictor of fragment length distribution for these data. The potential of increased temperatures to destabilize DNA is well known, and samples exposed to higher temperatures for a longer period will allow for enzymatic activity that degrades DNA [25]. However, under certain conditions, some samples stored at 4°C or even at room temperature show surprising viability. For example, samples preserved in EtOH and refrigerated for up to 1 week were nearly as good as flash-frozen samples. This is evident through high proportions of uHMW DNA molecules, although with some signs of degradation and variability across species and tissue types.

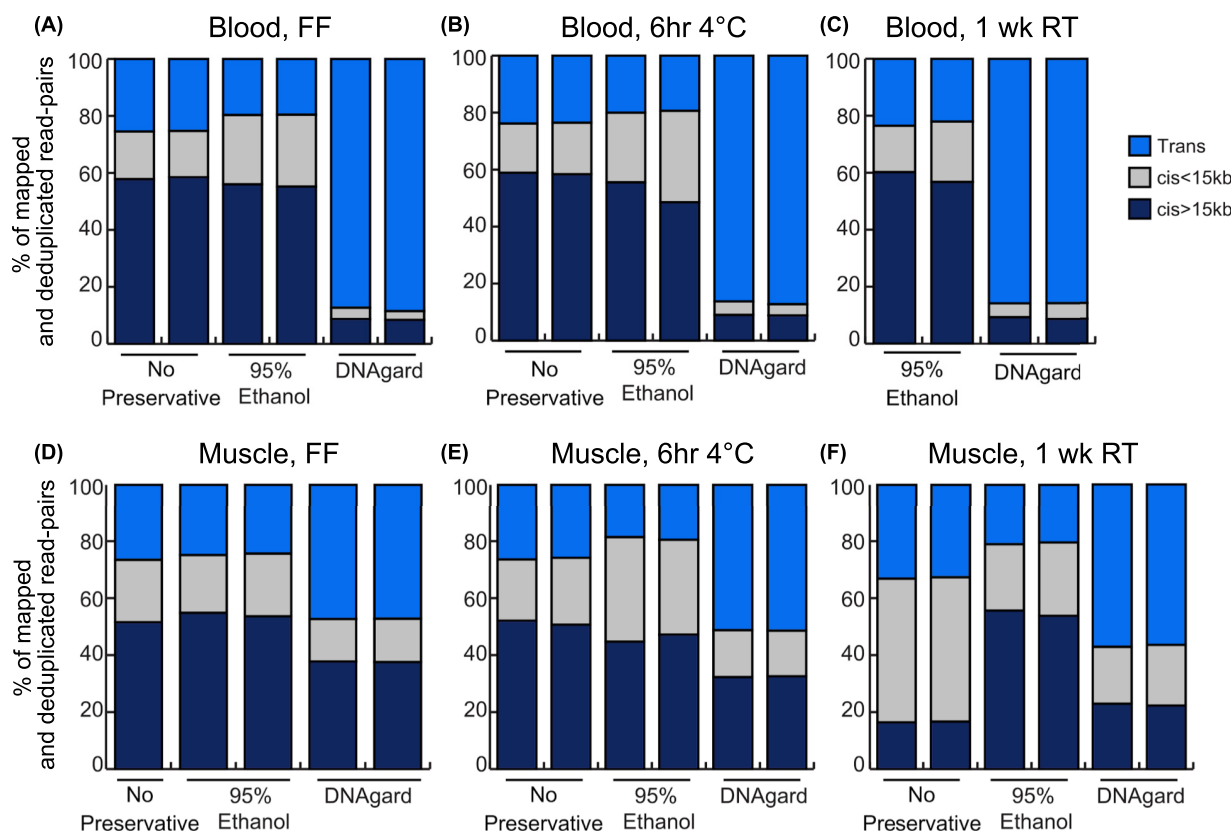


Figure 4: Hi-C platform benchmarking of bird samples. Stacked bar plots denoting proportions of Hi-C reads mapped to the zebra finch reference genome involving different chromosomes (*trans*), on the same chromosome but less than 15 kb apart (*cis* < 15 kb), and on the same chromosome and greater than 15 kb apart (*cis* > 15 kb). Tested samples include blood samples (A–C) and muscle samples (D–F). The desirable outcome is to have much greater proportions of Hi-C reads being long-range *cis* pairs, which reflects an efficient capture of long-range interactions needed for genome scaffolding and haplotype phasing. Hi-C data were generated by Arima Genomics following their standard protocol.

The ambient temperature of the intended collecting locality should be a major consideration in planning field collections for high-quality samples. Here we test a limited number of samples at 37°C to resemble fieldwork conditions in warmer climates, resulting in no retention of workable amounts of uHMW DNA in any of these samples (4 mouse muscle samples; Fig. 2). Thus, in hotter climates, sample cooling or exploring alternative preservatives is critical. Options such as insulated boxes, ice packs, wet ice, dry ice, and electronic coolers should be considered for maintaining samples at low temperatures in the field. To minimize the time before storing in ultra-cold freezers, investigators might also choose to ship samples from the field to the lab before the conclusion of fieldwork. Further experimentation in conditions resembling warmer climates can more precisely define tolerable exposure intervals for sampling targeting uHMW DNA.

The “gold standard” for preserving samples for uHMW DNA extraction remains flash-freezing in liquid nitrogen before ultra-cold storage [9–14]. Our results highlight alternative preservation methods that are more readily available in the field. Liquid nitrogen can be challenging to acquire, contain, and transport in many fieldwork settings. Fortunately, samples preserved in EtOH or DMSO perform well with simple refrigeration, although a small portion of DMSO samples fail (near-zero DNA extracted) for unclear reasons. In addition, these solutions consistently outperform the commercial preservatives RNAlater and DNAgard. Further, DNAgard is not suitable for maintaining long interac-

tion distances for Hi-C library preparation. While these commercial reagents rely on mechanisms that were likely optimized for preserving lower molecular weight nucleic acids, they appear to be harmful to uHMW DNA and chromosomal 3D interactions. Preservatives that promote cell lysis may undermine the stability of DNA if they cannot adequately counter the increased exposure to sources of chemical degradation [14, 25, 26]. Although our washing protocol should minimize its effect, it is also possible that some unknown aspect of the DNAgard treatment of cells inhibited the cross-linking reaction, and Hi-C of unfixed cells would be expected to have low signal and high noise similar to degraded DNA. Of the 3 commercial reagents tested, Allprotect shows the most promising results for preserving uHMW DNA, but more testing is necessary to better evaluate its performance relative to other preservatives and assess its compatibility with LR technologies.

In addition to popular commercial reagents, we evaluate some of the more commonly applied preservation methods today. EtOH has long been used for preserving samples for DNA analysis, and its proficiency at stabilizing specimens continues to be validated [12, 18, 27, 28]. For example, Mulcahy et al. [16] studied preservative effects on DNA integrity in white perch and blue crab muscle samples, using only a maximum of 45 kb DNA size resolution. Nevertheless, their finding that EtOH generally performs well as a DNA preservative agent is consistent with our results at this DNA size range. While EtOH is a compelling option, it comes with its

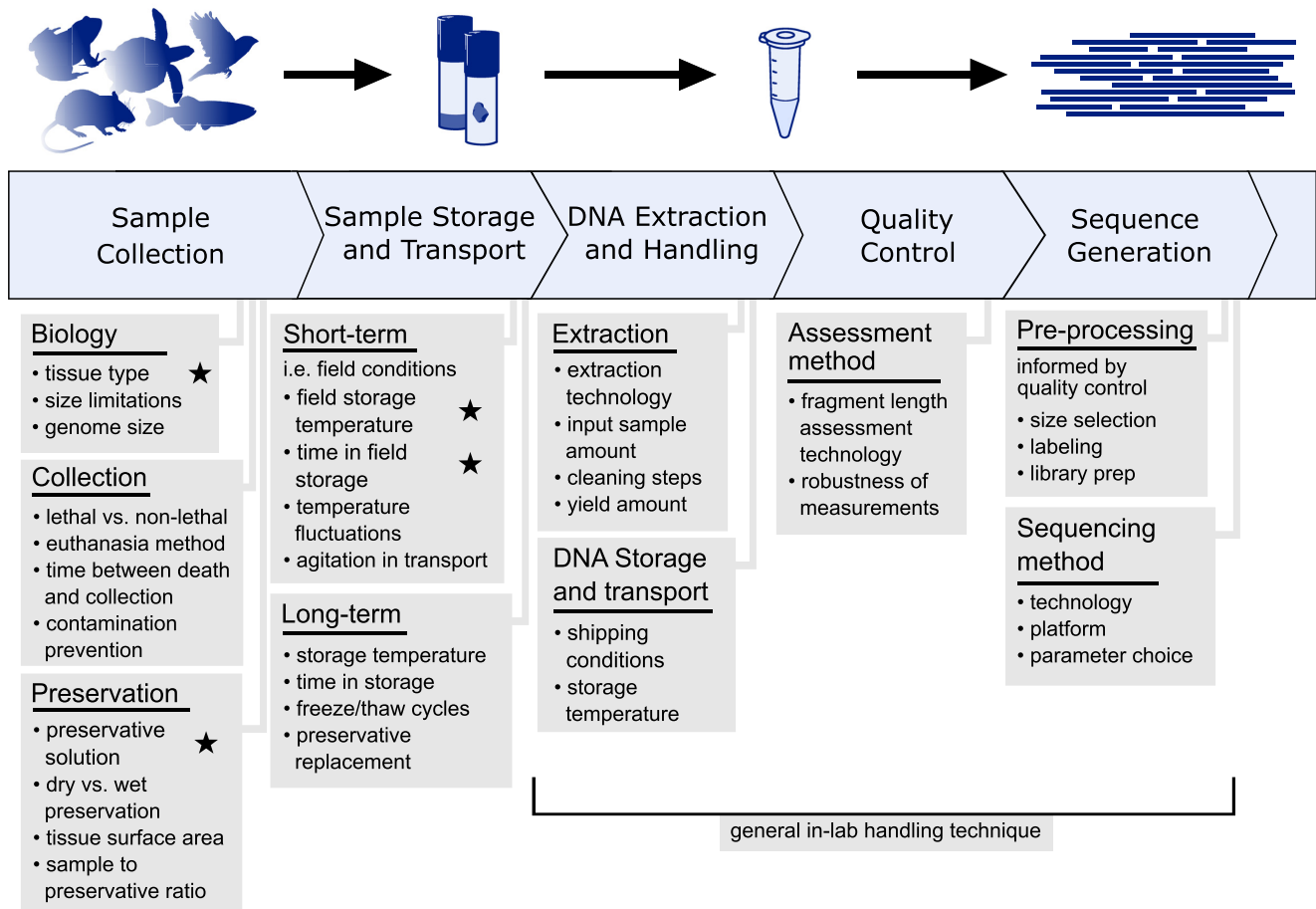


Figure 5: Considerations for collection of tissues for long-read sequencing of nonmodel organisms. General representation of a sequencing pipeline and considerations that may directly or indirectly affect the quality of sequencing output. Stars indicate particular sources of variation manipulated in this study. Several logistical aspects need to be considered prior to sample collection for uHWM DNA isolation with the goal of producing reference-quality genomes. The collector needs to identify what tissue types can be collected from the target species, what preservation methods and cold storage are available, and how quickly samples can be transported to a -80°C ultra-cold freezer.

own logistical considerations. EtOH can be problematic to transport on commercial flights or trains, or to ship in large quantities. Alternatively, DMSO benefits from fewer transport restrictions but requires laboratory preparation prior to fieldwork and can be hazardous to handle. Commercial preservation reagents are usually more costly than EtOH or DMSO solutions but are also under less restricted transport regulations.

The negative impact of DNAgard on Hi-C long-distance *cis* interactions is striking. This solution likely permeates the cell to inhibit nuclease activity, potentially affecting other protein integrity and impeding cross-linking. The increased fraction of interchromosomal interactions and decreased fraction of *cis* interactions (>15 kb) together are evidence of DNA degradation. These interchromosomal interactions are counterproductive noise with regard to chromosome-level scaffolding in that they erroneously provide scaffolding links between contigs derived from 2 different chromosomes. Our Hi-C data analysis also indicates, at least for birds, that EtOH storage of blood at 4°C or room temperature for 1 week or less tends to yield high-quality Hi-C chromatin interaction maps. Excluding samples in DNAgard, blood seems to be slightly more resistant to reducing chromosome interactions than muscle when stored at 4°C or room temperature for 1 week, which would be a valuable feature for field collection.

Contrary to the differences in Hi-C performance, we did not find notable differences in DNA fragment length distributions between most tissue types. The exception is whole-body fish samples that were all significantly degraded, regardless of treatment. Potentially, this could owe to the larger mass of tissue taking longer to freeze through or infuse with preservative, hence allowing more time for degradation. However, we did observe substantial differences in total DNA yield, where blood and spleen samples tend to yield a larger amount of DNA while muscle samples produce the least. The comparatively lower DNA yield makes muscle samples a less practical choice in species where nucleated blood is available. Lower yield could also be costlier and more time-consuming in the long run, as more DNA extractions would be required to achieve the necessary input amount. For species without nucleated blood (mammals), soft tissue samples such as the spleen outperform muscle in terms of yield. Note that low yield does not necessarily preclude muscle samples from usefulness, especially given they still perform well in terms of fragment length if appropriately collected and stored. We note that, as we demonstrated in a related study [29], blood is often not suitable for uHWM mitochondrial DNA extraction, while muscle tends to yield abundant mitochondrial DNA. This is an important consideration if the goal of collection is to sequence the mitochondrial genome.

Our study considers today's LR sequencing technologies and current DNA isolation protocols. Time will likely continue to yield new methods for preventing, assessing, and mitigating DNA degradation. Even since the outset of this study, promising new extraction methods have become available for uHMW DNA, such as Nanobind DNA extraction (Circulomics). Our comparisons focus on maximizing the quality of field-collected input material, and we expect this to be largely independent of downstream extraction methods. Our results and experience acquired with uHMW DNA and Hi-C data for more than 136 VGP genomes produced, yield guidelines for tissue type, preservatives, temperature, and other treatments necessary for generating high-quality genome assemblies from several vertebrate lineages, for laboratory and field collected samples (Table 1).

In planning biobanking for genomic purposes, another important strategy is to avoid or reduce the need for field-preserved samples. Seeking out animals already in captive collections and salvaging material reduces the methodological difficulty of preserving samples. Delaying blood collection, biopsy, or euthanasia of wild-caught specimens can also buy researchers time to move into more amenable preservation conditions such as a field station. However, this poses ethical challenges in the care of animals being held for days or weeks, and it is not feasible for larger animals.

Few studies have explored the effects of preservation methods on uHMW DNA integrity [17], but none that we are aware of have done so in as broad a set of field-relevant conditions as in the present study. Being able to collect samples well suited for producing high-quality genome assemblies is a major undertaking. Our recommendations will enable many new high-quality sample collections and contribute to establishing a greater and more diverse array of vertebrate genomes from around the world.

Methods

Sample collection

We collected samples from species representing major taxonomic classes of vertebrates, that is, house mouse (*Mus musculus*), zebra finch (*Taeniopygia guttata*), Kemp's Ridley sea turtle (*Lepidochelys kempii*), painted turtle (*Chrysemys picta*), American bullfrog (*Rana catesbeiana*), and zebrafish (*Danio rerio*). All animal handling and euthanasia protocols were approved by the Institutional Animal Care and Use Committees or equivalent regulatory bodies at the respective facilities: the Rockefeller University for the frog and bird samples, the Max Planck Institute for the mouse samples, the University of Toronto for the painted turtle samples, the Wellcome Sanger Institute for the fish samples, and the New England Aquarium rehabilitation facility for the sea turtle samples (Supplementary Table S1).

For this experiment, tissue samples were collected as available at facilities already handling the target species (Fig. 1). The tissue types collected per species are as follows: mouse, spleen and muscle; zebra finch, whole blood and muscle; sea turtle, isolated RBCs; painted turtle, whole blood and muscle; bullfrog, whole blood and muscle; and zebrafish, whole body, ovary, and muscle. For all species except the sea turtle and the fish, samples originate from a single individual. In the sea turtle set, duplicate samples were obtained from 3 individuals. In the fish set tissue samples in some cases originated from different individuals, as their small body size does not allow for sufficient amounts of tissue from a single specimen.

Each taxon required a slightly different handling procedure. All samples except for those from sea turtles were sourced from captive individuals humanely euthanized in a laboratory setting with approved protocols cited below. All soft or fibrous tissue samples were collected in small 20- to 30-mg pieces until each 2-ml tube had roughly 50–100 mg total to allow for full penetration of the preservative. Laboratory-raised mice were euthanized by CO₂ treatment in a GasDocUnit (Medres Medical Research GmbH, Cologne, Germany) following the instructions of the manufacturer (DD24.1–5131/451/8, Landesdirektion Sachsen). Skeletal muscle and spleen samples were then dissected and placed in standard cryotubes. Birds were euthanized via isoflurane overdose, and whole blood was collected into chilled sodium heparin-treated 1.5-ml microfuge tubes (IACUC #19101-H). Then, 25–50 µl was immediately aliquoted into cryotubes. Sea turtle RBC samples were collected from wild individuals undergoing medical treatment by drawing whole blood into 2-ml sodium heparin-treated collection tubes and then spinning down to separate RBCs from plasma. RBCs were then aliquoted into sodium heparin-treated tubes. Painted turtle samples were collected from 1 individual euthanized via decapitation as part of another study (AUP 20 012 070). Painted turtle muscle samples were immediately taken from the pectoral girdle and whole blood was drawn from the heart before placement in standard cryotubes. Frog samples were sourced from 1 captive adult purchased from Rana Ranch in Twin Falls, Idaho, USA. The frog was euthanized using an intracoelomic injection with Euthasol™ or Fatal-Plus™ (pentobarbital and phenytoin) at a dosage of 100 mg/kg. After confirming that a deep plane of anesthesia was reached, the frog was rapidly and doubly pithed cranially and spinally, then decapitated (19085-USDA). Frog muscle tissue samples were immediately taken from the rear legs, and blood was drawn from internal veins before placement in standard cryotubes. We extracted fish samples from multiple lab-raised individuals. To euthanize the fish, we used tricaine and then the brain was destroyed with a scalpel (PPL No.70/7606). We collected white muscle and ovary samples, which were dissected out and placed into 2-ml cryotubes immediately after euthanasia. Fish whole-body samples were taken by removing the head, intestines, and swim bladder of individual fish and placing the remaining tissue into a cryotube.

Preservation treatments

A total of 140 freshly collected samples were subjected to different preservation and temperature treatments to test common preservation methods under lab or simulated field conditions (Fig. 1), with flash-frozen samples being used as baseline controls. Preservation method treatments refer to the preservative agent applied directly to the sample before ultra-cold (–80°C) storage; temperature treatments refer to the temperature exposed and the amount of time the sample remained at that temperature before ultra-cold storage.

All temperature treatments were applied immediately upon dissection of the material and placement into specimen tubes. Samples were exposed to temperature treatments of varying lengths of time in refrigeration (4°C), room temperature (20–25°C), and elevated temperature in an incubator to simulate field conditions in a tropical climate (~37°C). All temperature conditions tested and the samples to which they were applied are as follows: control condition submerged in liquid nitrogen from dissection to ultra-cold storage (all tissue types and species), 6 hours at 4°C (frog blood and muscle, bird blood and muscle, painted turtle blood and muscle, sea turtle RBCs), 16 hours at 4°C (mouse spleen,

Table 1: Sample: collection guidelines for generating high-quality genomes. Compiled here are guidelines based on the best-performing protocols tested in this study and broadly in the phase 1 VGP genomes

Tissue selection	Tissues listed in decreasing preference. Multiple tissue types should be collected when possible.	
Preservation	Fish	Soft tissues; muscle; body with head, digestive tract, and swim bladder removed
	Amphibians	Blood, muscle
	Birds	Blood, muscle
	Nonavian reptiles	Blood/isolated red blood cells, muscle
	Mammals	Soft tissues like spleen, muscle
	Ideal	Flash-freezing or short-term refrigeration before deep freeze Blood or tissue specimens in 95% EtOH or 20–25% DMSO-EDTA can be stored at 4°C or on ice for up to 6 hours after dissection with little to no decrease in sample quality relative to immediate flash-freezing.
	Good	Midterm refrigeration before deep freeze Samples in 95% EtOH or 20–25% DMSO-EDTA can be stored for longer periods on ice/4°C of up to 1 week with minimal potential decrease in sample quality.
	Acceptable	Midterm room temperature storage before deep freeze Blood in 95% EtOH can be stored at room temperature (20–25°C) for up to 1 week with some potential decrease in DNA quality, most likely yielding extracts still within acceptable parameters for current long-read sequencing platforms. This condition is less likely to yield acceptable results with tissue samples.

fish whole body), 1 day at 4°C (fish ovary), 1 week at 4°C (mouse muscle, frog blood and muscle, bird blood and muscle, painted turtle blood and muscle), 1 day at room temperature (fish muscle and ovary), 1 week at room temperature (mouse muscle, frog blood and muscle, bird blood and muscle, painted turtle blood and muscle, sea turtle RBCs, fish muscle and ovary), 4 weeks at room temperature (fish muscle and ovary), 5 months at room temperature (sea turtle RBCs), and 1 week at 37°C (mouse muscle). Storage time at –80°C after treatment and before DNA extraction varied slightly between samples, but such variation is expected to have a negligible impact on sample quality.

The preservation methods tested here include flash-freezing in liquid nitrogen, no added preservative agent, 95% EtOH, 20–25% DMSO-EDTA (DMSO), DNAgard tissue and cells (DNAgard; cat. no. #62001–046, Biomatrix), Allprotect Tissue Reagent (Allprotect; cat. no. 76405, Qiagen), and RNAlater Stabilization Solution (RNAlater; cat. no. AM7021, Invitrogen). Our DMSO recipe was 20–25% DMSO, 25% 0.5 M EDTA, remaining 50–55% H₂O, saturated with NaCl. Flash-freezing, EtOH, and DNAgard were tested on all included species and tissue types. DMSO was tested on all species and tissue types except sea turtle RBCs. No-preservative treatments were tested on frog blood, bird blood, painted turtle blood, and sea turtle RBCs. Allprotect was tested on mouse spleen and muscle and fish body. RNAlater was tested on fish ovary and muscle samples.

To gain insights into variation within these treatments, isolated RBC samples were collected from 3 different sea turtle individuals and processed separately as biological and technical replicates. The third replicate had insufficient material to test all treatments.

DNA extraction

We extracted DNA from tissue samples using the agarose plug protocol as below at VGP data production hubs at the Rockefeller University, Wellcome Sanger Institute, and Max Planck Institute of Molecular Cell Biology and Genetics Dresden (Supplementary

Table S1). This method was established, at the time of this experiment, as standard protocol for long-read sequencing in all VGP projects [4]. From each tissue sample, a 30- to 40-mg piece was weighed and then processed using the Bionano Prep™ Animal Tissue DNA Isolation Fibrous Tissue Protocol (Bionano document number 30071) and Soft Tissue Protocol (Bionano document number 30077). Briefly, the fibrous tissue (muscle, whole) pieces were further cut into 3-mm pieces and fixed with 2% formaldehyde and Bionano Prep Animal Tissue Homogenization Buffer. Tissue was blended into a homogenate with a Qiagen Rotor-Stator homogenizer and embedded in 2% agarose plugs cooled to 43°C. Plugs were treated with Proteinase K and RNase A and washed with 1× Bionano Prep Wash Buffer and 1× TE Buffer (pH 8.0). DNA was recovered with 2 μl of 0.5 U/μl Agarase enzyme per plug for 45 minutes at 43°C and further purified by drop dialysis with 1× TE Buffer. The soft tissue (spleen, ovary) pieces were further cut into 3-mm pieces and then homogenized with a tissue grinder followed by a DNA stabilization step with ethanol. The homogenate pellet was then embedded in 2% agarose plugs as in the fibrous tissue protocol above. For blood samples, DNA was extracted from whole blood or RBCs following the unpublished Bionano Frozen Whole Nucleated Blood Stored in Ethanol—DNA Isolation Guidelines. The ethanol supernatant was removed and the blood pellet was resuspended in Bionano Cell Buffer in a 1:2 dilution. For samples that freeze solidly at –80°C, tubes were thawed at 37°C for 2–4 minutes. The same Bionano guidelines for nucleated blood in ethanol were modified by adding 1–2 additional centrifugation steps at 5,000 × g for 10 minutes prior to removing DNAgard supernatant and homogenizing blood cells in Bionano Cell Buffer in a 1:2 dilution. All samples were mixed with 36 μl agarose and placed in plug molds following the animal tissue protocol.

Assessing sample purity and yield

All extractions had sufficient DNA yield to measure except one: mouse spleen tissue in DMSO. This sample congealed and

solidified in such a way that no DNA could be extracted. To measure DNA yield and purity, we used both the fluorescence-based Broad Range Qubit®, Invitrogen, Waltham, MA, USA assay and absorbance-based Nanodrop One™, Thermo Fisher Scientific, Waltham, MA, USA. To measure yield, 2- μ l aliquots of genomic DNA were taken from the top, middle, and bottom of each DNA sample and diluted in a Qubit Working Solution of 1:200 Dye Assay Reagent with BR Dilution Buffer. Sample concentrations were recorded on a Qubit 4 Fluorometer. The concentrations of the top, middle, and bottom readings were averaged to estimate the concentration of each DNA sample. Spectrophotometry was then performed on a Nanodrop One to measure sample purity in terms of the 260/230-nm and 260/280-nm ratios.

Assessing sample fragment size distributions

Fragment length distributions of samples were measured with at least 1 of 2 available methods: PFGE or FEMTO. PFGE was performed using the Sage Science™ Pippin Pulse gel system with the Lambda PFG Ladder (New England Biolabs, Ipswich, MA, USA). To quantify fragment length distribution from PFGE gel images, we compared the proportions of signal above and below 145 kb. This was done using the program ImageJ [30] following Mulcahy et al. [16] based on the Gel Analysis tool in ImageJ. Further quantifying of the PFGE signal below 145 kb, such as the relative amount of low molecular weight DNA, was not robust due to compression or streaking obscuring smaller fragment patterns. Concise visualization of gel plot profiles was produced in the R package ggridges [31] with a custom Python script for piecewise linear scaling across different gels according to a common size standard. Gray-value intensity measured in ImageJ was scaled locally in each lane and cropped to the gel boundary such that, excluding the well, the brightest value along the lane became 100 and the darkest became 0. Analysis of FEMTO outputs was carried out in the ProSize Data Analysis Software. First, each trace was assessed for signs of an unreliable run, including ladder quality, loading concentration, raised baseline, and unusual smear patterns. Runs with these hallmarks were not incorporated further. Because signals above 165 kb are not reliable on FEMTO, we only considered signals within the range of 1.3–165 kb. We then recorded the proportion of the sample measuring above 45 kb. Further visualization of FEMTO traces was made in the same manner as above with a custom Python script and the R package ggridges, except scaling to a size standard was done in ProSize. Yields were insufficient for fragment size analysis from frog muscle in DMSO for 1 week at 4°C and 6 hours at 4°C and mouse spleen in DMSO for 16 hours at 4°C.

Statistical analysis

We used linear modeling in the R statistical package to explore the relative contribution of several factors to the variance in DNA yield and fragment length among tests. The 3 response variables—DNA yield per unit mass (yield), PFGE proportion >145 kb (PFGE), and FEMTO proportion >45 kb (FEMTO)—were modeled separately. The data for each model were samples with those measurements, and all conditions had at least 2 replicates (yield: $n = 139$, PFGE: $n = 102$, FEMTO: $n = 108$). DNA yield was log-transformed using the natural logarithm to satisfy assumptions of normality and modeled with temperature, preservative, vertebrate group, and tissue type included as fixed effects. Homoscedasticity was checked after modeling and found to conform to assumptions. PFGE proportion and FEMTO proportion were modeled with quasibinomial error distributions with tem-

perature, preservation method, and tissue type included as fixed effects. Vertebrate group was not included in the final fragment length models due to collinearity with tissue type. Post hoc tests were done using the `glht` function of the R package `multcomp` to examine differences between the levels of each factor. Further model details including P values and contingency tables are available in the supplementary materials (Supplementary Tables S2 and S3).

Hi-C library preparation and sequencing

Because Hi-C methods require intact cell nuclei, we tested a subset of bird samples from our preservation experiments and 2 additional no-preservative bird muscle samples directly using the Arima-HiC platform. We tested blood and muscle samples in 3 different treatments: without preservatives, in EtOH, and in DNAgard. Each preservation method was subjected to 3 temperature treatments: immediately flash-frozen, 6 hours at 4°C, and 1 week at room temperature (20–25°C). After temperature treatment, each sample was moved to –80°C. Blood with no preservative at room temperature for 1 week was excluded from this set. Two technical replicates of each sample were prepared and sequenced at Arima Genomics following their standard protocol (Arima Genomics, Doc A160177 v00). Briefly, standard protocol for nucleated blood in a solution like EtOH or DNAgard is to pellet the cells, remove the supernatant, wash with 1 \times phosphate-buffered saline solution containing 1% fetal bovine serum, and then carry the washed pelleted cells into cross-linking and then Arima-HiC. We measured the performance of Arima-HiC runs by mapping the sequence reads to the zebra finch reference genome (GCA_003957565.1) to determine the proximity of ligated sequence pairs. Assessments were made based on the ratios of *cis* (intrachromosome) to *trans* (interchromosome) read pairs as well as the total percentage comprising long-distance (>15 kb) *cis* pairs.

Data Availability

Sample information, PFGE measurements, FEMTO measurements, and DNA yield data can be found in the supplemental materials. Raw FEMTO outputs, PFGE gel images, and analysis scripts are available on Dryad [32]. Additional files, including sample information (Supplementary Table S1), detailed statistical outputs (Supplementary Tables S2 and S3), and figure source values are available via GigaDB [33]. Raw Hi-C read-pairs are publicly available on the DDBJ DRA (BioProject: PRJDB13233, BioSample: SAMD00448194–SAMD00448226).

Additional Files

Supplementary Fig. S1. Example of genomic DNA traces of the same sample made with 2 different methods. DNA extract from flash-frozen mouse muscle was measured with (A) pulsed-field gel electrophoresis via the software ImageJ and (B) Agilent FEMTO Pulse via the software ProSize. Traces are displayed as gel images (right) and as plot profiles (left).

Supplementary Fig. S2. Plot profiles of FEMTO results on fish and frog samples. Agilent FEMTO Pulse traces for fish and frog samples are visualized as overlapping ridgeline plots. Each ridgeline plot corresponds to a single sample. The x-axis values are scaled via ProSize Data Analysis Software. The y-axis of each plot is scaled independently to ignore peaks outside the range of 10–165 kb such that the highest value in that range of each plot becomes 100 and the lowest value becomes 0. Plots with excess noise are generally ones that have low signal.

Supplementary Fig. S3. Plot profiles of FEMTO results on turtle, mouse, and bird samples. Agilent FEMTO Pulse traces for turtle, mouse, and bird samples are visualized as overlapping ridgeline plots. Each ridgeline plot corresponds to a single sample. The x-axis values are scaled via ProSize Data Analysis Software. The y-axis of each plot is scaled independently to ignore peaks outside the range of 10–165 kb such that the highest value in that range of each plot becomes 100 and the lowest value becomes 0. Plots with excess noise are generally ones that have low signal.

Supplementary Fig. S4. Testing of different variables on uHMW DNA proportion as measured by FEMTO. Distributions of sample groups are overlaid with results of linear modeling of fragment length ($n = 108$). Shown are box-and-whisker plots, with the median, quartiles, and full range of individual observations. Fragment length was quantified here as the proportion of signal between 45 kb and 165 kb as measured on the Agilent FEMTO Pulse system and modeled in a generalized linear model with temperature (A), preservative (B), and sample type (C) as predictors. Significant relationships from post hoc comparisons are shown as connecting bars with significance levels: **** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$.

Supplementary Fig. S5. Comparisons of replicate sea turtle RBC samples. (A) DNA yields per unit of sample input shown for up to 3 replicates per treatment. (B) Bar plots of PFGE measurements of signal proportion greater than 145 kb, excluding the well. The ordering of sea turtle replicates is plotted consistently from left to right (i.e., the first bar of each treatment is from the same individual).

Supplementary Table S1. Sample information and measurement data. “Total.yield” refers to the amount of DNA recovered from the tissue sample in μg . “Yield.per.mass” is calculated based on the total DNA yield and input sample mass, and “mass.for.5ug” is a calculation of the required sample mass to reach the target of 5 μg DNA. “PFGE.gt145kb.percent” and “PFGE.le145kb.percent” are calculations based on the plot profiles of PFGE gel lanes for percentage greater than or less than 145 kb, respectively. “Femto.perc.1300bp-44999bp,” “Femto.perc.45kb-165kb,” and “Femto.adjusted.45kb-165kb” similarly are calculations of percentage of detected DNA within each indicated fragment size range. The “260.28” and “260.23” columns contain the commonly measured 260/280 and 260/230 ratios used to assess sample purity.

Supplementary Table S2. Statistical analysis output. Full output parameters are reported for statistical analyses of DNA yield, fragment length as measured by PFGE, and fragment length as measured by FEMTO.

Supplementary Table S3. Statistical analysis contingency tables. Each table contains the number of samples in each combination of parameters.

Abbreviations

Allprotect: Allprotect Tissue Reagent cat. no. 76405, Qiagen; d: day; DNAgard: DNAgard tissue and cells cat. no. #62001-046, Biomatrica; frag length-Femto: DNA fragment length measured with Agilent Femto Pulse System; FF: flash frozen; frag length-PFGE: DNA fragment length measured by pulsed-field gel electrophoresis; hr: hour; Isolated RBCs: isolated red blood cells; LN2: Flash frozen in liquid nitrogen; None: no preservation solution added; RNAlater: RNAlater Stabilization Solution (RNAlater; cat. no. AM7021, Invitrogen); RT: room temperature; wk: week; yield: DNA yield; 25% DMSO: MSO, a mix of 20–25% dimethyl sulfoxide, 25% 0.5 M EDTA,

and 50–55% H₂O; 37C, 4C: temperature in Celsius; 95% EtOH: 95% ethanol.

Competing Interests

The authors declare no competing interests.

Funding

This research was supported by Howard Hughes Medical Institute Funds and Rockefeller University Startup funds to E.D.J., institutional funds of the Max Planck Institute of Molecular Cell Biology and Genetics, and funds by the Wellcome Trust made out to DNAP R&D team at Wellcome Sanger Institute. I.B.’s time was supported by Wellcome grants WT207492 and 104640/Z/14/Z, 092096/Z/10/Z. Sampling was facilitated by Leslie Buck and Mouska Patang for the painted turtle and Brian Fabella for the bullfrog. Sea turtle sampling was conducted and generously facilitated by the New England Aquarium and Massachusetts Audubon Wellfleet Bay Wildlife Sanctuary authorized under USFWS permit TE01150C-1; sample transfer to L.M.K. was permitted via a USFWS special authorization letter.

Authors’ Contributions

J.M., S.W., A.F.S., I.B., L.M.K., T.L., A.J.C., R.W.M., A.D.S., P.A.M., E.D.J., and O.F. initially conceptualized the study. H.A.D., J.M., J.B., S.W., A.F.S., S.M., O.V.P., I.B., K.O., M.S., W.T., A.K., L.M.K., E.D.J., and O.F. carried out data collection and preprocessing. H.A.D., J.B., G.F., and A.F.S. analyzed the data and produced the figures. The manuscript was drafted by H.A.D., J.M., J.B., G.F., E.D.J., and O.F., and all authors contributed to revisions.

References

- Koepfli, KP, Paten, B, Genome 10 K Community of Scientists, O’Brien, SJ. The Genome 10 K Project: a way forward. *Annu Rev Anim Biosci* 2015;**3**(1):57–111.
- Ko, BJ, Lee, C, Kim, J, et al. Widespread false gene gains caused by duplication errors in genome assemblies. *bioRxiv* 2021. <https://doi.org/10.1101/2021.04.09.438957>.
- Kim, J, Lee, C, Ko, BJ, et al. False gene and chromosome losses affected by assembly and sequence errors. *bioRxiv* 2021. <https://doi.org/10.1101/2021.04.09.438906>.
- Rhie, A, McCarthy, SA, Fedrigo, O, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 2021;**592**(7856):737–46.
- Sambrook, J, Russell, DW. Purification of nucleic acids by extraction with phenol:chloroform. *CSH Protoc*. 2006 Jun 1;**2006**(1):pdb.prot4455.
- Lahiri, DK, Nurnberger, JI Jr. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res* 1991;**19**(19):5444.
- Zhang, M, Zhang, Y, Scheuring, CF, et al. Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. *Nat Protoc* 2012;**7**(3):467–78.
- Zhang, Y, Zhang, Y, Burke, JM, et al. A simple thermo-plastic substrate containing hierarchical silica lamellae for high-molecular-weight DNA extraction. *Adv Mater* 2016;**28**(48):10630–6.
- Frampton, M, Sam, D. Evaluation of specimen preservatives for DNA analyses of bees. *J Hymenopt Res* 2008;**17**:195–200.

10. Kilpatrick, CW. Noncryogenic preservation of mammalian tissues for DNA extraction: an assessment of storage methods. *Biochem Genet* 2002;**40**(1/2):53–62.
11. Seutin, G, White, BN, Boag, PT. Preservation of avian blood and tissue samples for DNA analyses. *Can J Zool* 1991;**69**(1):82–90.
12. Reiss, RA, Schwert, DP, Ashworth, AC. Field preservation of Coleoptera for molecular genetic analyses. *Environ Entomol* 1995;**24**(3):716–9.
13. Wong, PB, Wiley, EO, Johnson, WE, et al. Tissue sampling methods and standards for vertebrate genomics. *Gigascience* 2012; **1**(1):8.
14. Anchordoquy, TJ, Molina, MC. Preservation of DNA. *Cell Preserv Technol* 2007;**5**(4):180–8.
15. Camacho-Sanchez, M, Burraco, P, Gomez-Mestre, I, et al. Preservation of RNA and DNA from mammal samples under field conditions. *Mol Ecol Resour* 2013;**13**(4):663–73.
16. Mulcahy, DG, Macdonald, KS III, Brady, SG, et al. Greater than kb: a quantitative assessment of preservation conditions on genomic DNA quality, and a proposed standard for genome-quality DNA. *PeerJ* 2016;**4**:e2528.
17. Zhang, Y, Broach, J. Abstract 5125: A novel method for isolating high-quality UHMW DNA from 10 mg of freshly frozen or liquid-preserved animal and human tissue including solid tumors. *Mol Cell Biol Genet* 2019;**79**(13_Supplement):5125.
18. Srinivasan, M, Sedmak, D, Jewell, S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am J Pathol* 2002;**161**(6):1961–71.
19. Oosting, T, Hilario, E, Wellenreuther, M, et al. DNA degradation in fish: Practical solutions and guidelines to improve DNA preservation for genomic research. *Ecol Evol* 2020;**10**(16): 8643–51.
20. Michaud, CL, Foran, DR. Simplified field preservation of tissues for subsequent DNA analyses. *J Forensic Sci* 2011;**56**(4):846–52.
21. Bista, I, McCarthy, SA, Wood, J, et al. The genome sequence of the channel bull blenny, (Günther, 1861). *Wellcome Open Res* 2020;**5**:148.
22. Kronenberg, ZN, Rhie, A, Koren, S, et al. Extended haplotype-phasing of long-read de novo genome assemblies using Hi-C. *Nat Commun* 2021;**12**(1):1935.
23. Balakrishnan, CN, Edwards, SV, Clayton, DF. The zebra finch genome and avian genomics in the wild. *Emu Austral Ornithol* 2010;**110**(3):233–41.
24. Warren, WC, Clayton, DF, Ellegren, H, et al. The genome of a songbird. *Nature* 2010;**464**(7289):757–62.
25. Klingström, T, Bongcam-Rudloff, E, Pettersson, OV. A comprehensive model of DNA fragmentation for the preservation of high molecular weight DNA. *bioRxiv* 2018, <http://dx.doi.org/10.1101/254276>.
26. Elmore, S. Apoptosis: a review of programmed cell death. *Toxicol Pathol* 2007;**35**(4):495–516.
27. Doyle, JJ, Dickson, EE. Preservation of plant samples for DNA restriction endonuclease analysis. *Taxon* 1987;**36**(4): 715–22.
28. Evans, RK, Xu, Z, Bohannon, KE, et al. Evaluation of degradation pathways for plasmid DNA in pharmaceutical formulations via accelerated stability studies. *J Pharm Sci* 2000;**89**(1): 76–87.
29. Formenti, G, Rhie, A, Balacco, J, et al. Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. *Genome Biol* 2021;**22**(1):120.
30. Schneider, CA, Rasband, WS, Eliceiri, KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 2012;**9**(7):671–5.
31. Wilke, CO. *ggridges: Ridgeline Plots in “ggplot2.”* R package version 0.5.3. 2021. <https://cran.r-project.org/package=ggridges>.
32. Dahn, HA, Mountcastle, J, Balacco, J, et al. Data from: Benchmarking ultra-high molecular weight DNA preservation methods for long-read and long-range sequencing, Dryad, Dataset. 2022. <https://doi.org/10.5061/dryad.000000041>.
33. Dahn, HA, Mountcastle, J, Balacco, J, et al. Supporting data for Benchmarking ultra-high molecular weight DNA preservation methods for long-read and long-range sequencing. *GigaScience Database*. 2022. <http://dx.doi.org/10.5524/102202>.