Recombinase Polymerase Amplification (RPA) versus PCR for ancient DNA library amplification

Stephen M Richards Corresp., [1] , **Kieren J Mitchell** [1] , **Raymond Tobler** [1] , **Alan Cooper** [1]

[1] School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia

Corresponding Author: Stephen M Richards
Email address: steve.richards@adelaide.edu.au

Background: Recombinase Polymerase Amplification (RPA) is a relatively new isothermal methodology for amplifying DNA. RPA is similar to traditional PCR in that it produces an amplicon that is defined by the annealing of two opposing oligonucleotide primers. However, while PCR relies on repeated heating and cooling cycles to denature and amplify DNA fragments, RPA is performed at a single moderate temperature and uses enzymatic activity to drive amplification. While RPA is commonly used in field-based monitoring of pathogens, it is unknown whether RPA is a viable alternative to PCR for the amplification of ancient DNA.

Methods: In this study, PCR and RPA were used to amplify shotgun and mitochondrial DNA enriched libraries made from extracts from four ancient bison bone samples. Sequencing data from the amplified libraries were examined for biases in sequence content (read length and GC content), fraction of unique reads mapping to a reference sequence, and mitochondrial polymorphisms detection accuracy.

Results: In comparison to PCR, RPA had a variable effect on sequence content, except in the mitochondrial DNA enriched libraries where RPA consistently reduced mean read length by approximately 30 bp. RPA increased the number of unique shotgun reads that mapped to a cattle nuclear reference by between 9% and 99% versus PCR. In contrast, RPA reduced the fraction of unique mitochondrial DNA enriched reads by > 26%, possibly due to the preferential amplification of small unmappable molecules.Both RPA and PCR data allowed the identification of similar variants in mitochondrial DNA enriched libraries, suggesting that the accuracy of the two amplification methods is comparable. Importantly, RPA was able to generate sequencing libraries at approximately a sixth of the cost of PCR. These results indicate that RPA is a viable alternative to PCR for amplification of shotgun libraries made from ancient DNA but may not be suitable for all ancient DNA applications.

Recombinase Polymerase Amplification (RPA) versus PCR for ancient DNA library amplification

Stephen M. Richards, Kieren J. Mitchell, Raymond Tobler, and Alan Cooper

Australian Centre for Ancient DNA, School of Biological Science, University of Adelaide, Adelaide, South Australia, Australia

Corresponding Author:
Stephen M. Richards
University of Adelaide, North Terrace Campus, Darling Building Rm 205b, Adelaide, South Australia, Australia 5005
Email: steve.richards@adelaide.edu.au

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27544v1 | CC BY 4.0 Open Access | rec: 19 Feb 2019, publ: 19 Feb 2019

Abstract

Background: Recombinase Polymerase Amplification (RPA) is a relatively new isothermal methodology for amplifying DNA. RPA is similar to traditional PCR in that it produces an amplicon that is defined by the annealing of two opposing oligonucleotide primers. However, while PCR relies on repeated heating and cooling cycles to denature and amplify DNA fragments, RPA is performed at a single moderate temperature and uses enzymatic activity to drive amplification. While RPA is commonly used in field-based monitoring of pathogens, it is unknown whether RPA is a viable alternative to PCR for the amplification of ancient DNA.

Methods: In this study, PCR and RPA were used to amplify shotgun and mitochondrial DNA enriched libraries made from extracts from four ancient bison bone samples. Sequencing data from the amplified libraries were examined for biases in sequence content (read length and GC content), fraction of unique reads mapping to a reference sequence, and mitochondrial polymorphisms detection accuracy.

Results: In comparison to PCR, RPA had a variable effect on sequence content, except in the mitochondrial DNA enriched libraries where RPA consistently reduced mean read length by approximately 30 bp. RPA increased the number of unique shotgun reads that mapped to a cattle nuclear reference by between 9% and 99% versus PCR. In contrast, RPA reduced the fraction of unique mitochondrial DNA enriched reads by > 26%, possibly due to the preferential amplification of small unmappable molecules. Both RPA and PCR data allowed the identification of similar variants in mitochondrial DNA enriched libraries, suggesting that the accuracy of the two amplification methods is comparable. Importantly, RPA was able to generate sequencing libraries at approximately a sixth of the cost of PCR. These results indicate that RPA is a viable alternative to PCR for amplification of shotgun libraries made from ancient DNA but may not be suitable for all ancient DNA applications.

Introduction

High-throughput sequencing and hybridization capture enrichment are techniques that have become routine in ancient DNA (aDNA) research. After DNA extracted from an ancient sample has been converted to a sequencing library the DNA concentration is low and the library must be amplified in order to produce sufficient material for shotgun sequencing and/or hybridization capture. Ancient extracts usually contain high proportions of non-target DNA, which can often make two sequential rounds of hybridization capture necessary (Fu et al. 2013; Handt et al. 1994; Li et al. 2013). Since relatively small quantities of DNA are recovered following hybridization capture, ancient libraries will require additional amplification after each enrichment. Consequently, aDNA libraries undergo several rounds of amplification prior to sequencing and

PCR is by far the most common method used for these amplification steps. However, PCR does not amplify with complete fidelity and will alter the composition of a library by introducing sequence biases (Dabney & Meyer 2012).

PCR biases result in certain DNA molecules being preferentially amplified. PCR is known to preferentially amplify sequences with a GC content in the 50% to 60% range (Benjamini & Speed 2012). This is because the replication complex in DNA with lower GC content has a greater chance to disassociate and produce a truncated amplicon that cannot act as template in further cycles (Su et al. 1996), while DNA molecules with higher GC content tend to form secondary structures which can block the full extension of the primer by the DNA polymerase (Guido et al. 2016). In aDNA mammalian studies, the GC preference of PCR is known to shift the content of a sequencing library away from endogenous DNA to environmental contamination (Dabney & Meyer 2012). The loss of endogenous DNA is caused by the preferential amplification of microbial DNA, which can have a GC content of >70%, over mammalian nuclear DNA that has a GC content of ≈ 40% to 50% (Hildebrand et al. 2010; Romiguier et al. 2010). PCR is also known to have a bias towards shorter molecules (Dabney & Meyer 2012) as DNA polymerases have a greater chance of disassociating from longer templates and producing truncated amplicons that cannot participate in downstream amplification (Debode et al. 2017).

Isothermal amplification comprises a group of methodologies that are performed at a constant relatively low temperature and use enzymatic activity to denature and amplify DNA instead of heat. Several isothermal methodologies are currently used for DNA amplification (Gill & Ghaemi 2008). For example, Multiple Displacement Amplification (MDA) is used in many

whole genome amplification protocols because of the high fidelity and large yield the method produces (Dean et al. 2002). While MDA is known to introduce sequence biases, these biases have been shown to be less severe than for PCR (Pinard et al. 2006). However, MDA has not been applied to aDNA research because the amplification efficiency of this methodology is positively correlated with template length (Lage et al. 2003), and the latter is characteristically low in aDNA extracts (Brotherton et al. 2007).

One isothermal methodology that can be easily substituted for PCR is Recombinase Polymerase Amplification (RPA), which uses proteins involved in genetic recombination to amplify target DNA. In RPA, recombinase enzymes first form a complex with a primer that scans the template DNA for complimentary sequences. Once found, the primer is annealed to the complimentary sequence and the non-complimentary template strand is displaced. The recombinase enzymes then disassociate from the primer and a DNA polymerase with a strand displacement activity, such as *Bst* DNA polymerase, binds to the double-stranded DNA formed by the primer and template. DNA single strand binding proteins attach to the displaced strand stabilizing the formation of a replication fork. The DNA polymerase then extends the primer to produce a copy of the original template. Repeated extension of two opposing primers produces exponential amplification of the target DNA (Piepenburg et al. 2006). Like PCR, RPA produces an amplicon constrained in size to the binding sites of the primers. While RPA is commonly used for field-based monitoring of pathogens (Liljander et al. 2015), the potential of this technology in aDNA research remains largely unknown.

RPA is a relatively new amplification methodology and there are a limited number of studies that have investigated the biases introduced by this technology. RPA is known to be biased towards smaller DNA molecules (Santiago-Felipe et al. 2015). RPA is performed at a constant temperature so theoretically a molecule can be amplified at most $2^{T/t}$ times, where T is the length of the incubation and t is the time it takes to replicate a template. Since length of t will be influenced at least in part by the size of a template (short molecule = low t), so too will the final copy number. The maximum copy number for short templates will therefore be higher than for long molecules. Further, RPA has also been shown to be biased against sequences with high GC content (Hansen et al. 2016) as the strand exchange activity of some recombinase proteins has been shown to be inhibited by GC content of 70% (Patil et al. 2011). Previous studies have successfully used isothermal amplification on degraded DNA from forensic samples and formalin fixed paraffin embedded (FFPE) tissues (Tate et al. 2011; Wang et al. 2004), as well as modern sequencing libraries (Jasmine et al. 2008; Lou et al. 2013; Ma et al. 2013; Oyola et al. 2012). While, isothermal methods have been also used in the hybridization capture of targets from sequencing libraries made from aDNA (Haak et al. 2015), their performance compared to PCR was not evaluated.

In the current study, we directly compare the performance of the isothermal methodology RPA [using a TwistAmp Basic kit (TwistDx)] and conventional PCR [using Platinum *Taq* DNA Polymerase High Fidelity (ThermoFisher)] for amplifying shotgun and mitochondrial DNA (mtDNA) enriched libraries made from aDNA. Shotgun and mtDNA-enriched libraries made from four ancient bison bones (> 12,000 years old) were amplified using both RPA and PCR, and we compared sequence content biases (i.e. length and GC content), mapping statistics, and

variant calling. We also compared damage profiles for each mtDNA enriched library, as deamination of cytosine to uracil is common in aDNA templates (Brotherton et al. 2007), particularly towards the ends of molecules (Briggs et al. 2010; Lindahl 1993), and many DNA polymerases will read uracil as thymine and misincorporate an adenosine in the complimentary strand. Finally, we compared the cost of RPA versus PCR in our study.

Materials & Methods

Samples

Four ancient bison bones from North America and Asia were selected for study (Table 1) and all of the samples were radiocarbon dated by the Oxford Radiocarbon Accelerator Unit to the Pleistocene. Extraction of aDNA from the bison samples and library construction were performed in the ancient DNA laboratory of the Australian Centre for Ancient DNA (University of Adelaide), which is dedicated to low concentration DNA research and is regularly cleaned with bleach and exposed to UV light. We followed strict guidelines established for aDNA studies (Cooper & Poinar 2000) including extraction blanks and negative controls for all amplifications.

Table 1. Bison samples
*Previously published carbon dates (Soubrier et al. 2016).

DNA extraction

A small block of roughly 1 cm by 1 cm was cut from each bison bone using a Dremel tool with a carborundum cutting disk (Dremel). The resulting bone section was wrapped in aluminium foil and crushed with a hammer. One hundred mg of the resulting bone fragments were extracted using a standard silica-based binding protocol that has been previously described (Brotherton et al. 2013; Rohland & Hofreiter 2007).

Library construction

Two libraries were made for each bison extraction (20 µL of input DNA per library) following previously published method employing truncated Illumina adapters with unique dual 7-mer barcodes (Table S1) (Knapp et al. 2012; Llamas et al. 2016; Meyer & Kircher 2010). Library construction included a treatment with an enzyme cocktail to partially remove uracils (Rohland et al. 2015) and all libraries were taken to the adapter fill-in step which included heat inactivation of the *Bst* DNA polymerase.

Recombinase polymerase amplification

Because exponential amplification (Ahmed et al. 2014) and high yields have been reported for RPA (TwistDx: TwistAmp manual), a single 50 µL TwistAmp reaction was used for all isothermal amplification steps in this study. To concentrate the DNA libraries in order to meet the sample volume requirement of the TwistAmp protocol, one post-*Bst* (40 µL) library per bison was concentrated using a MinElute PCR Purification Kit (Qiagen) by binding to a spin-column and then eluting in a volume of 13.2 µL EB (Qiagen) + 0.05% Tween-20 (Rohland & Reich 2012). RPA was performed in a reaction containing: 13.2 µL library, 2.4 µL each of 10 µM IS7 and IS8 primers (Meyer & Kircher 2010), 29.5 µL TwistAmp Rehydration Buffer, 2.5 µL of 280 mM, and lyophilized TwistAmp reagents. The RPA reactions were mixed well via pipetting and incubated at 39° C for 15 min in a heated-lid thermal cycler. Amplification was stopped by adding 5 µL 0.5 M EDTA and vortexing. Fifty microliters of H_2O was added to the RPA to thin the viscous reaction and the product was purified using 1.8x volumes of Sera-Mag SpeedBeads (GE Healthcare) (Rohland & Reich 2012) and eluted with 30 µL EB + 0.05% Tween. The RPA

product was electrophoresed on a 2% agarose gel, visualized with GelRed staining (Biotium), and quantified using a Qubit dsDNA broad range assay (ThermoFisher). Portions of the truncated library produced by this initial RPA were then used to produce shotgun and mtDNA enriched libraries.

An indexed shotgun library was constructed for each bison sample by combining the following: 30 ng of the initial RPA product, 2.4 µL of 10 µM IS4 primer, 2.4 µL of 10 µM indexing primer (Meyer & Kircher 2010), 29.5 µL TwistAmp Rehydration Buffer, 2.5 µL of 280 mM, lyophilized TwistAmp reagents, and H_2O to 50 µL and then incubated at 39° C for 5 min in a heated-lid thermal cycler. Shotgun amplifications were stopped, purified with 1.8x volumes Sera-Mag beads, and quantified with a Qubit assay as in the initial RPA.

Mitochondrial DNA enrichment was performed using 150 ng of the initial RPA product and an in-house hybridization capture procedure (Richards et al. 2019) with a modification to the final elution: at the end of enrichment procedure, the streptavidin magnetic beads with the captured library were suspended in 13.2 µL EB (Qiagen) + 0.05% Tween-20 and heated at 95°C for 5 min to release the DNA. The beads were pelleted with a magnetic rack and the supernate containing the released DNA was transferred to a new 1.5 ml tube for storage at -20°C. To amplify the enriched mtDNA the following were combined: 13.2 µL enriched mtDNA library, 2.4 µL of 10 µM IS4 primer, 2.4 µL of 10 µM indexing primer, 29.5 µL TwistAmp Rehydration Buffer, 2.5 µL of 280 mM, and lyophilized TwistAmp reagents, and then incubated at 39° C for 20 min in a heated-lid thermal cycler. Enriched library amplifications were stopped, purified, and quantified as in the initial RPA.

PCR amplification

To minimize the introduction of biases, all PCRs were performed with low cycle amplifications and multiple replicates (Polz & Cavanaugh 1998; Rohland et al. 2015). The 40 µL heat inactivated *Bst* reaction from library construction was divided into 8 x 25 µL PCR reactions each containing: 5 µL DNA, 2.5 µL 10x High Fidelity PCR Buffer, 1 µL 50 mM $MgSO_4$, 0.2 µL 25 mM dNTPs, 0.5 µL each of 10 µM IS7 and IS8 primers (Meyer & Kircher 2010), 0.1 µL Platinum *Taq* DNA Polymerase High Fidelity (5 U/µL), and molecular biology grade H_2O to 25 µL. The PCRs were amplified in a heated-lid thermal cycler programmed as follows: initial denaturation 94º C for 2 min; 9 cycles at 94º C for 15 seconds, 58º C for 30 seconds, then 68º C for 45 seconds; and a final extension at 68º C for 2 min. PCRs from the same library were pooled, purified with 1.8x volumes Sera-Mag beads as before, and eluted with 30 µL EB + 0.05% Tween. One microliter of this elution was quantified using qPCR to determine the minimum number of PCR cycles to amplify this library to produce sufficient DNA for downstream procedures (Carøe et al. 2017). Libraries from bison-875 and bison-3133 required dilution (6 µL library added to 24µL EB + 0.05% Tween) and 1µL of this dilution was re-assayed with qPCR as above. Libraries were further amplified in 6 x 25 µL PCRs containing 2.5 µL 10x High Fidelity PCR Buffer, 1 µL 50 mM $MgSO_4$, 0.2 µL 25 mM dNTPs, 0.5 µL each of 10 µM IS7 and IS8 primers, 0.1 µL Platinum *Taq* DNA Polymerase High Fidelity (5 U/µL), 3 µL eluted library (bison-885 and bison-4089) or 3 µL diluted library (bison-875 and bison-3133), and molecular biology grade H_2O to 25 µL. The PCRs were amplified in a heated-lid thermal cycler programmed as follows: initial denaturation 94ºC for 2 min; various cycles at 94ºC for 15 seconds, 58º C for 30 seconds, then 68ºC for 45 seconds; and a final extension at 68º C for 2

min. The cycle numbers used were: extraction blank – 12 cycles, bison 875 – 8 cycles, bison 885 – 6 cycles, bison 3133 – 8 cycles, and bison 4089 – 13 cycles. PCRs from the same library were pooled, purified with Sera-Mag SpeedBeads, eluted in 30 µL EB + 0.05% Tween, and processed as above. The truncated library produced by this second round of amplification was used in all further steps.

To generate indexed shotgun libraries, 30 ng of the truncated library was divided among 4 x 25 µL replicates each containing: 2.5 µL 10x Taq HiFi Buffer, 1µL 50 mM $MgSO_4$, 0.5 µL, 10 mm dNTPs, 0.5 µL each of 10 µM IS4 primer, 2.4 µL of 10 µM indexing primer, 0.1 µL Taq HiFi (5 U/µL), 7.5 ng library from the initial PCR amplification, and H_2O to 25 µL. PCR amplification was performed in a heated-lid thermal cycler programed as follows: initial denaturation at 94°C for 2 min; 6 cycles at 94° C for 10 seconds, 58° C for 30 seconds, 68° C for 30 seconds; and a final extension at 68° C for 2 min. PCRs from the same library were pooled and purified with 1.8x volumes Sera-Mag beads as before.

Mitogenomes were enriched using 150 ng of the PCR truncated library following the same in-house hybridization capture procedure used with the RPA libraries but with a modification to the final recovery of mtDNA. At the end of the enrichment procedure the streptavidin magnetic beads with the captured library were suspended in 30 µL EB + 0.05% Tween-20 and then heated at 95° C for 5 min to release the captured mtDNA. The beads were pelleted with a magnetic rack and the supernate was transferred to a new 1.5 ml tube for storage at -20° C. One microliter of the captured mtDNA was quantified as before to determine cycle number required for the PCR. Amplification was performed in 4 x 25 µL PCRs each containing: 2.5 µL 10x Taq HiFi Buffer,

1µL 50 mM $MgSO_4$, 0.5 µL, 10 mm dNTPs, 0.5 µL each of 10 µM IS4 primer, 2.4 µL of 10 µM indexing primer, 0.1 µL Taq HiFi (5 U/µL), 7.25 µL captured mtDNA, and H_2O to 25. The PCRs were amplified in a heated-lid thermal cycler programmed as follows: initial denaturation 94º C for 2 min; various cycles at 94º C for 15 seconds, 58º C for 30 seconds, then 68º C for 45 seconds; and a final extension at 68º C for 2 min. The cycle numbers used were: extraction blank – 24 cycles, bison 875 - 20 cycles, bison 885 – 22 cycles, bison 3133 – 21 cycles, and bison 4089 – 22 cycles. PCRs from the same library were pooled and purified with 1.8x volumes Sera-Mag SpeedBeads as before.

Sequencing

Indexed libraries were quantified with a Qubit Broad Range dsDNA assay (ThermoFisher) and then diluted to 5 ng/uL with EB + 0.05% Tween-20. The diluted libraries were assayed with a D1000 ScreenTape and 2200 TapeStation (Agilent) and pooled in equimolar amounts according to the ScreenTape results. Pooled libraries were sent to the Kinghorn Centre for Clinical Genomics for sequencing on an Illumina HiSeq X Ten run using paired-end 2 x 150 bp (300 cycle) chemistry.

Analysis of sequencing data

Fastq files from the sequencer were initially demultiplexed according to the internal 7-mer internal barcodes using Sabre (version 1.0: https://github.com/najoshi/sabre). Reads were further processed using AdapterRemoval (2.2.1) (Schubert et al. 2016) to trim adapters, collapse overlapping read pairs, discard reads < 25 bp, and remove reads of low quality (Phred < 40). To eliminate the impact of differences in sequencing depths on mapping results, reads from both

shotgun and mtDNA enriched libraries were randomly subsampled to the level of the least-deeply sequenced library (1,850,000 and 1,000,000 for the shotgun and mtDNA enriched libraries respectively) using the reformat command of BBTools (v36.62-intel-2017.01: https://jgi.doe.gov/data-and-tools/bbtools/) and Java (v1.8.0_121).

To examine the accuracy of the amplification methods, all of the collapsed reads in each of the mtDNA-enriched libraries were mapped because subsampling would limit our ability to call variants. Collapsed reads were mapped to a cattle nuclear genome reference (UMD Bos_taurus 3.1) and a *Bison bison* mitochondrial genome reference (GenBank number: GU947006.1) using BWA aln (0.5.11-foss-2016b) with parameters recommended for aDNA (Li & Durbin 2009; Schubert et al. 2012). After mapping, duplicate reads were removed using the SortSam and MarkDuplicates packages of Picard Tools v2.2.4: (https://broadinstitute.github.io/picard/index.html). Damage profiles for the mtDNA enriched reads mapped to the mitochondrial reference were generated using mapDamage2.0 (Jónsson et al. 2013). Variants were called using Geneious v10.0.8 and the default parameters with a minimum read coverage of ≥ 5 (Kearse et al. 2012).

To examine the sequence content biases introduced by the amplification methods, the shotgun and enriched bison libraries were examined for differences in sequence composition. All libraries were subsampled to a maximum 25,000 reads using the reformat command of BBTools and Java to provide a comparable number of sequences in each dataset. Read length and GC content were extracted from the subsampled data using SeqKit (v0.7.2: https://github.com/shenwei356/seqkit), which were then used to generate violin plots and perform Wilcoxon signed rank tests with R (v3.4.2).

Results

Sequence content biases

To examine the introduction of sequence biases by PCR and RPA four types of data were compared: unmapped reads from shotgun libraries (shotgun-unmapped), reads mapped to a nuclear genome from shotgun libraries (shotgun-mapped), unmapped reads from mtDNA-enriched libraries (mtDNA-unmapped), and reads mapped to a mitogenome from mtDNA-enriched libraries (mtDNA-mapped). Split violin plots were generated to allow pair-wise comparison of the sequence biases (read length and GC content) introduced by the amplification methods (Figures 1 and 2). For this discussion, "mean" refers to the mean across all four bison combined (the "All' violins in Figures 1 and 2), while "average" is used to refer to the mean of an individual bison sample.

As illustrated by the violin plots, amplification method impacted sequence biases differently. When comparing mean read length, RPA tended to produce shorter reads than PCR (Figures 1 and 2). The smallest difference in mean read length between the amplification methods differed by 0.98 bp in the shotgun-mapped data (Figure 1C). The mtDNA-enriched data exhibited considerably larger differences in mean read length with RPA libraries being on average ≈ 30 bp shorter than the corresponding PCR data [29.12 bp for the mtDNA-unmapped (Figure 2A) and 28.09 bp for mtDNA-mapped (Figure 2C) data]. The difference in mean read length was significant for all four comparisons (Wilcoxon signed rank test scores: shotgun-unmapped: $W = 5.7 \times 10^9$, p-value $< 2.2 \times 10^{-16}$; shotgun-mapped: $W = 5.1 \times 10^9$, p-value $< 2.2 \times 10^{-16}$; mtDNA-unmapped: $W = 7.4 \times 10^9$, p-value $< 2.2 \times 10^{-16}$; mtDNA-mapped: $W = 2.0 \times 10^9$, p-value $< 2.2 \times 10^{-16}$). In contrast, RPA did not consistently reduce average read length relative to PCR in the

317 shotgun data at the level of the individual bison in the shotgun data [4.48 bp increase for bison-
318 3133 in shotgun-unmapped data (Figure 1A) and an increase of 3.71 bp and 3.01 bp respectively
319 for bison-885 and bison-3133 in shotgun-mapped data (Figure 1C)].

320
321
322
323 **Figure 1. Shotgun library split violin plots**
324 Read length and GC content distributions of four shotgun libraries constructed from bison aDNA and amplified with
325 either PCR or RPA. Bison sample numbers are plotted on the x-axis and the "All" violin figures represent the mean
326 of all four bison. The colored area represents the frequency distribution of read length or GC content of the
327 amplification method. The diamonds in the violin plots are the mean of each amplification method and the red
328 dotted line represents the 41.89% GC content of the nuclear cattle reference (GenBank number: UMD 3.1) used for
329 mapping. In comparison to PCR, amplification with RPA did not have a consistent effect on read length or GC
330 content. Amplification of shotgun libraries with RPA increased the fraction of unique reads by > 9% in comparison
331 to PCR (Table 2). **A**) Shotgun-unmapped: Read Length, **B**) Shotgun-unmapped: GC Content, **C**) Shotgun-mapped:
332 Read Length, and **D**) Shotgun-mapped: GC Content.
333
334

335 Amplification method also produced variable results with regards to the frequency in read GC
336 content, with the largest differences occurring in unmapped data (Figures 1B and 2B) where
337 PCR increased the frequency of higher GC content reads in comparison to RPA. Of particular
338 note is a strong shift towards reads with a GC content of ≈ 62% produced by PCR in bison-4089.

339

340 In further comparison of the GC biases, RPA reduced the mean GC content of the shotgun-
341 unmapped, shotgun-mapped, and mtDNA-unmapped data by 8.13%, 3.59%, and 5.42%
342 respectively in comparison to PCR (Figures 1B, 1D, and 2B). In contrast, RPA produced an
343 increase of 0.35% relative to PCR in the mean GC content in the mtDNA-mapped dataset
344 (Figure 2D). The differences in mean GC content between PCR and RPA were significant in all
345 cases with Wilcoxon signed rank test scores that ranged from $W = 1.4x10^9$, p-value = $3.1x10^{-2}$
346 for the mtDNA-mapped data and $W = 6.8x10^9$, p-value < $2.2x10^{-16}$ for shotgun-unmapped data.
347 Again, however, the effect of RPA on GC content was not always consistent at the sample level.
348 In the shotgun-mapped libraries for example, RPA increased the average GC content of bison-

4089 by 1.15% compared to PCR, while decreasing the average GC content between 3.77% and 6.02% in the other bison samples (Figure 1D). Further, the difference between PCR and RPA in average GC content of the mtDNA-mapped libraries showed no consistent pattern and tended to be small, with a maximum difference of 1.6% across the four bison (Figure 2D).

Figure 2. mtDNA-enriched library split violin plots
Read length and GC content of ancient bison libraries enriched for mtDNA using hybridization capture and amplified with either PCR or RPA. Bison sample numbers are plotted on the x-axis and the "All" violin figures represent the mean of all four bison. The colored area represents the frequency distribution of read length or GC content of the amplification method. The diamonds in the violin plots are the mean of each amplification method and the red dotted line represents the 39.14% GC content of the bison mitochondrial reference (GenBank number: GU947006.1) used for mapping. RPA reduced the mean read length of mtDNA-enriched data by ≈ 30 bp in comparison to PCR amplification. The reduction in read length produced by the isothermal method appears to have favored the amplification of small unmappable DNA molecules and reduced the fraction of unique mapped reads (by > 26%) in the RPA mtDNA enriched data (Table 3). **A**) mtDNA-unmapped: Read Length, **B**) mtDNA-unmapped: GC Content, **C**) mtDNA-mapped: Read Length, and **D**) mtDNA-mapped: GC Content.

Shotgun library mapping

Subsampled shotgun libraries were mapped to a cattle nuclear reference and a bison mitogenome reference using BWA aln and parameters commonly used with aDNA (Schubert et al. 2012). After the removal of duplicate reads, the fraction of unique mapped reads was calculated by dividing the number of unique mapped reads by the number of collapsed reads (Tables 2 and 3). In comparison to PCR, RPA consistently increased the fraction of uniquely mapped reads that aligned with the nuclear cattle reference (from 9% to 99%). In contrast, the effect of amplification methods was inconsistent regarding the fraction of unique reads from shotgun libraries that mapped to the mitogenome reference. These mitochondrial results likely stemmed from the generally low numbers of mitochondrial sequences in the shotgun libraries, which made the mapping results stochastic.

mtDNA-enriched library mapping

Subsampled mtDNA enriched libraries were mapped to a bison mitogenome reference using BWA as with the shotgun data analysis. In all bison, amplification with RPA reduced the fraction of unique reads by >26% when compared to PCR (Table 3). This reduction of unique reads is likely the result of the greater bias of RPA towards smaller DNA fragments, which preferentially amplified reads that were too short to accurately map (Figure 2). To compare the accuracy of RPA to PCR, variants in the mtDNA-enriched data were called against a bison mitochondrial reference (SI File). To maximize the number of variants that could be identified, all the reads in the mtDNA-enriched libraries were mapped. Except in loci of low coverage, PCR and RPA performed in a similar manner and identified the same variants in each bison sample (Table 4).

Table 2. Mapping statistics of shotgun libraries

Shotgun libraries were made from four ancient bison and amplified with either PCR or RPA. To eliminate the effect of differences in sequencing depth all libraries were subsampled to 1,850,000 collapsed reads (a number determined from the lowest number of reads for any sample) and then mapped to a bison mitochondrial reference genome (GenBank number: GU947006.1) and a cattle reference genome (GenBank number: UMD 3.1) using BWA and parameters standardly used with aDNA (Schubert et al. 2012). The fraction of unique mapped reads was determined by dividing the number of unique mapped reads by the number of collapsed reads. In comparison to PCR, RPA had a variable effect on the unique reads mapped to a mitochondrial reference, while RPA consistently increased unique reads that mapped to a nuclear reference. Libraries with lower proportions of endogenous DNA experienced the greatest increase (> 60%) in unique reads that mapped to the nuclear reference with RPA. Mapping statistics for the entire shotgun dataset are given in Table S2. ExB = Extraction Blank

Table 3. Mapping statistics mtDNA-enriched libraries

Four ancient bison sequencing libraries were enriched for mtDNA using hybridization capture and amplified with either PCR or RPA. To eliminate the effect of differences in sequencing depth all libraries were subsampled to 1,000,000 collapsed reads (a number determined from the lowest number of reads for any sample) and mapped to a bison reference (GenBank number: GU947006.1) using BWA and parameters standardly used with aDNA (Schubert et al. 2012). The fraction of unique mapped reads was determined by dividing the number of unique mapped read by the number of collapsed reads. In comparison to PCR, RPA reduced the fraction of unique reads by >26% in the mapped data. The drop in unique reads observed in the RPA mtDNA-enriched libraries likely stemmed from the amplification of reads that were too short to accurately map. The mapping statistics for the entire mtDNA-enriched dataset is given in Table S3. ExB = Extraction Blank

Table 4. Number of variants detected in mtDNA-enriched mapped data

The number of sequence variants (e.g. single nucleotide polymorphism) detected in four ancient bison sequencing libraries enriched for mtDNA, amplified with either PCR or RPA, and mapped to a modern bison reference mitogenome (GenBank number: GU947006.1) using BWA and parameters standardly used with aDNA (Schubert et al. 2012). For this analysis all data from the mtDNA-enriched libraries were used to maximize the number of variants, which were called using Geneious software (v10.0.08) with the default parameters and a minimum read depth of ≥ 5(Kearse et al. 2012). The denominator of the fraction is the total number of variants detected in the sample and the numerator is the number of variants called in the mtDNA enriched data. The difference between the amplification methods represents the loci that could not be called because of low coverage.

Damage profiles

To determine if the DNA polymerase included in the TwistAmp kit will read through deaminated cytosine, the damage profiles generated by PCR and RPA were examined. Libraries amplified with PCR and RPA produced damage profiles typical of aDNA, indicating the polymerase in the TwistDx kit does read through deaminated cytosine and misincorporate a thymine (Figure 3). These damage results indicate that the library preparation and bioinformatics safeguards commonly used with PCR amplified data to minimize the impact of deaminated cytosine on sequencing data must also be considered when amplifying aDNA using a TwistAmp kit. As with PCR, damage profiles in TwistAmp amplified libraries can be used to aid with the authentication of aDNA.

Figure 3. PCR and RPA damage profiles
Damage profiles of four bison mtDNA-enriched libraries amplified with either PCR or RPA and mapped to a modern bison reference mitogenome (GenBank number: GU947006.1) were generated using mapDamage 2.0 (Jónsson et al. 2013). The similar profiles between the two amplification methods indicates that the DNA polymerase provided in the TwistAmp kit will read through and misincorporate an adenosine when encountering a deaminated cytosine.
A) Bison 875-PCR, **B**) Bison 875-RPA, **C**) Bison 885-PCR, **D**) Bison 885-RPA, **E**) Bison 3313-PCR, **F**) Bison 3313-RPA, **G)** Bison 4089-PCR, and **H**) Bison 4089-RPA

Cost

The estimated cost of a single amplification reaction in this study was $2.60 USD for PCR and $3.70 USD for RPA (TwistAmp kit), which makes PCR less expensive on per reaction basis. However, we performed multiple parallel PCRs per library, as is typical for aDNA studies (Polz

& Cavanaugh 1998; Rohland et al. 2015), which increased the overall cost of the PCR protocol. For example, to produce an mtDNA-enriched library with the TwistAmp kit required two amplification reactions at a cost of $7.40 USD (2 x $3.70 USD), while to generate the equivalent library with PCR required a total of 18 amplification reactions at a cost of $ 46.80 USD (18 x $2.60 USD). The cost of PCR amplification will vary depending on the DNA polymerase and the number of reactions used at each step, however, TwistAmp will remain competitive cost-wise in any protocol that uses multiple PCRs for amplification.

Discussion

Bison shotgun libraries amplified with RPA had an increased fraction of unique reads that mapped to a nuclear reference compared to samples amplified with PCR. This increase likely stems from several biases inherent in the amplification methods. First, RPA appears to have the stronger bias towards smaller DNA molecules, which would have favoured the amplification of fragmented endogenous DNA over more intact environmental contamination. Second, the conflicting biases of the methods towards DNA with higher GC content would have favoured the amplification of endogenous bison molecules by RPA. PCR is biased towards sequences with higher GC content and favoured the amplification of microbial DNA, while RPA is biased against sequences with higher GC content, which preferentially amplified endogenous bison DNA. In a project that requires a large volume of sequencing, such as generating an ancient genome, the small increase in unique reads produced by RPA may lead to a significant reduction in study costs.

We found RPA to be competitive cost wise to any aDNA protocol that amplifies with multiple PCRs. Importantly, amplification with a single RPA reaction produced an increased number of reads mapping to a nuclear reference compared to a multiple PCR protocol. Applying a single RPA reaction approach will reduce the labor effort and streamline library amplification procedures. Further, amplifying with a single RPA reaction will reduce the risk of contaminating a sample especially in the early stages of library preparation.

While RPA performed well with shotgun libraries, this method does not appear to be suited for amplification of mtDNA-enriched libraries because of a strong bias towards small DNA molecules. In hybridization capture of degraded DNA there are two opposing biases with regards to read length. First, hybridization capture favors larger DNA fragments because longer sequences form more stable complexes with probe molecules leading to a higher frequency of recovery (Brotherton et al. 2013). Second, PCR and RPA are both biased towards smaller DNA molecules. In our mtDNA-enriched data, the large reduction in mean read length (≈ 30 bp) produced by RPA suggest that this isothermal method has a much stronger size bias than PCR and is able to overcome the size preference of hybridization capture towards longer DNA molecules. The strong size bias of RPA in our study likely led to a greater amplification of short DNA molecules the mtDNA-enriched libraries and in the case of endogenous DNA many of these small reads were too short to map causing a reduction in the coverage of the mitogenome reference in mapping analysis. The strong size bias of RPA would also make the isothermal method unsuitable for amplifying shotgun libraries with a high proportion of small DNA molecules. It is not entirely clear why RPA did not produce a stronger size bias in the shotgun data. Sequence diversity in shotgun libraries is much greater than in enriched libraries and this

increased complexity may have dampened the size bias of RPA. The size bias of RPA is dependent on the replication complex repeatedly attaching to small DNA molecules, which is less likely in a complex shotgun library than a library that has been enriched for short fragmented endogenous DNA.

Conclusion

This study sought to determine if RPA was a viable alternative to PCR in the amplification of shotgun and mtDNA enriched libraries made from aDNA. Our data indicates that RPA can be substituted for PCR for the amplification of aDNA shotgun libraries. In comparison PCR, RPA produced an increase of unique reads from ancient bison libraries that mapped to a cattle nuclear reference genome at a reduced cost. However, RPA does not appear to a suitable method to amplify libraries after hybridization capture as the amplification method has a strong bias towards short molecules that are not mappable.

In this study aDNA from sub-fossil was examined but, RPA may also be suitable for the amplification of degraded DNA from other sources. FFPE tissues from clinical procedures represents a largely underutilized genetic resource for pathological conditions, mainly because of the difficulty of processing the DNA from these samples (Tang et al. 2009). Similar to aDNA from sub-fossil bones, the DNA from FFPE tissues is damaged, fragmented, and found at low concentrations (Munchel et al. 2015). To overcome these difficulties, high-throughput sequencing is now being applied to DNA extracted from FFPE samples to characterize the mutations involved with diseases such as cancer (Munchel et al. 2015). Amplification of shotgun

libraries prepared from FFPE tissues with RPA may produce comparable increases in endogenous mapped reads as observed in the current study.

Acknowledgements: The authors of this manuscript would like to thank Holly Heiniger and Nicole Moore for their endless technical help in the laboratory. The authors would also like to thank Matt Gilliham, Geoff Fincher, Birgitte Skadhauge, and Birger Møller for their help in obtaining the funding which supported this study.

References

Ahmed A, van der Linden H, and Hartskeerl RA. 2014. Development of a Recombinase Polymerase Amplification Assay for the Detection of Pathogenic Leptospira. *International Journal of Environmental Research and Public Health* 11:4953-4964. 10.3390/ijerph110504953

Benjamini Y, and Speed TP. 2012. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Research* 40:e72-e72. 10.1093/nar/gks001

Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, and Pääbo S. 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Research* 38:1-12. 10.1093/nar/gkp1163

Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, and Cooper A. 2007. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Research* 35:5717-5728.

Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, Jane Adler C, Richards SM, Sarkissian CD, Ganslmeier R, Friederich S, Dresely V, van Oven M, Kenyon R, Van der Hoek MB, Korlach J, Luong K, Ho SYW, Quintana-Murci L, Behar DM, Meller H, Alt KW, and Cooper A. 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat Commun* 4:1764. 10.1038/ncomms2656

Carøe C, Gopalakrishnan S, Vinner L, Mak SST, Sinding MHS, Samaniego JA, Wales N, Sicheritz-Pontén T, and Gilbert MTP. 2017. Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution*:n/a-n/a. 10.1111/2041-210X.12871

Cooper A, and Poinar HN. 2000. Ancient DNA: Do it right or not at ALL. *Science* 289:1139-1139.

Dabney J, and Meyer M. 2012. Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* 52:87-94. 10.2144/000113809

Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, and Lasken RS. 2002. Comprehensive human genome amplification using multiple displacement amplification. *Proceedings of the National Academy of Sciences of the United States of America* 99:5261-5266.

Debode F, Marien A, Janssen E, Bragard C, and Berben G. 2017. The influence of amplicon length on real-time PCR results. *Biotechnology, Agronomy, Society and Environment* 21:3-11.

Fu QM, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, and Pääbo S. 2013. DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences of the United States of America* 110:2223-2227. 10.1073/pnas.1221359110

Gill P, and Ghaemi A. 2008. Nucleic acid isothermal amplification technologies - A review. *Nucleosides Nucleotides & Nucleic Acids* 27:224-243. 10.1080/15257770701845204

575 Guido N, Starostina E, Leake D, and Saaem I. 2016. Improved PCR Amplification of Broad
576 Spectrum GC DNA Templates. *PLoS One* 11:e0156478.
577 10.1371/journal.pone.0156478
578 Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S,
579 Harney E, Stewardson K, Fu Q, Mittnik A, Banffy E, Economou C, Francken M,
580 Friederich S, Pena RG, Hallgren F, Khartanovich V, Khokhlov A, Kunst M, Kuznetsov
581 P, Meller H, Mochalov O, Moiseyev V, Nicklisch N, Pichler SL, Risch R, Rojo Guerra
582 MA, Roth C, Szecsenyi-Nagy A, Wahl J, Meyer M, Krause J, Brown D, Anthony D,
583 Cooper A, Alt KW, and Reich D. 2015. Massive migration from the steppe was a
584 source for Indo-European languages in Europe. *Nature* advance online publication.
585 Handt O, Höss M, Krings M, and Pääbo S. 1994. Ancient DNA: Methodological challenges.
586 *Cellular and Molecular Life Sciences* 50:524-529. 10.1007/bf01921720
587 Hansen S, Schäfer J, Fechner K, Czerny C-P, and Abd El Wahed A. 2016. Development of a
588 Recombinase Polymerase Amplification Assay for Rapid Detection of the
589 Mycobacterium avium subsp. paratuberculosis. *PLoS One* 11:e0168733.
590 10.1371/journal.pone.0168733
591 Hildebrand F, Meyer A, and Eyre-Walker A. 2010. Evidence of Selection upon Genomic GC-
592 Content in Bacteria. *PLoS Genetics* 6:e1001107. 10.1371/journal.pgen.1001107
593 Jasmine F, Ahsan H, Andrulis IL, John EM, Chang-Claude J, and Kibriya MG. 2008. Whole-
594 genome amplification enables accurate genotyping for microarray-based high-
595 density single nucleotide polymorphism array. *Cancer Epidemiology, Biomarkers &*
596 *Prevention* 17:3499-3508.
597 Jónsson H, Ginolhac A, Schubert M, Johnson PLF, and Orlando L. 2013. mapDamage2.0: fast
598 approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*
599 29:1682-1684. 10.1093/bioinformatics/btt193
600 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A,
601 Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, and Drummond A. 2012.
602 Geneious Basic: an integrated and extendable desktop software platform for the
603 organization and analysis of sequence data. *Bioinformatics (Oxford, England)*
604 28:1647-1649. 10.1093/bioinformatics/bts199
605 Knapp M, Stiller M, and Meyer M. 2012. Generating barcoded libraries for multiplex high-
606 throughput sequencing. *Methods in Molecular Biology* 840:155-170.
607 Lage JM, Leamon JH, Pejovic T, Hamann S, Lacey M, Dillon D, Segraves R, Vossbrinck B,
608 Gonzalez A, Pinkel D, Albertson DG, Costa J, and Lizardi PM. 2003. Whole genome
609 analysis of genetic alterations in small DNA samples using hyperbranched strand
610 displacement amplification and array-CGH. *Genome Research* 13:294-307.
611 Li C, Hofreiter M, Straube N, Corrigan S, and Naylor GJ. 2013. Capturing protein-coding
612 genes across highly divergent species. *Biotechniques* 54:321-326.
613 10.2144/000114039
614 Li H, and Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler
615 transform. *Bioinformatics* 25:1754-1760. 10.1093/bioinformatics/btp324
616 Liljander A, Yu M, O'Brien E, Heller M, Nepper JF, Weibel DB, Gluecks I, Younan M, Frey J,
617 Falquet L, and Jores J. 2015. Field-Applicable Recombinase Polymerase
618 Amplification Assay for Rapid Detection of Mycoplasma capricolum subsp.
619 capripneumoniae. *Journal of Clinical Microbiology* 53:2810-2815.
620 10.1128/jcm.00623-15

Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature* 362:709-715.

Llamas B, Fehren-Schmitz L, Valverde G, Soubrier J, Mallick S, Rohland N, Nordenfelt S, Valdiosera C, Richards SM, Rohrlach A, Romero MIB, Espinoza IF, Cagigao ET, Jiménez LW, Makowski K, Reyna ISL, Lory JM, Torrez JAB, Rivera MA, Burger RL, Ceruti MC, Reinhard J, Wells RS, Politis G, Santoro CM, Standen VG, Smith C, Reich D, Ho SYW, Cooper A, and Haak W. 2016. Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances* 2. 10.1126/sciadv.1501385

Lou DI, Hussmann JA, McBee RM, Acevedo A, Andino R, Press WH, and Sawyer SL. 2013. High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. *Proceedings of the National Academy of Sciences* 110:19872-19877. 10.1073/pnas.1319590110

Ma Z, Lee RW, Li B, Kenney P, Wang Y, Erikson J, Goyal S, and Lao K. 2013. Isothermal amplification method for next-generation sequencing. *Proceedings of the National Academy of Sciences* 110:14320-14323. 10.1073/pnas.1311334110

Meyer M, and Kircher M. 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols* 2010:pdb.prot5448. 10.1101/pdb.prot5448

Munchel S, Hoang Y, Zhao Y, Cottrell J, Klotzle B, Godwin AK, Koestler D, Beyerlein P, Fan J-B, Bibikova M, and Chien J. 2015. Targeted or whole genome sequencing of formalin fixed tissue samples: potential applications in cancer genomics. *Oncotarget* 6:25943-25961. 10.18632/oncotarget.4671

Oyola SO, Otto TD, Gu Y, Maslen G, Manske M, Campino S, Turner DJ, MacInnis B, Kwiatkowski DP, Swerdlow HP, and Quail MA. 2012. Optimizing illumina next-generation sequencing library preparation for extremely at-biased genomes. *BMC Genomics* 13:12. 110.1186/1471-2164-13-1

Patil KN, Singh P, and Muniyappa K. 2011. DNA Binding, Coprotease, and Strand Exchange Activities of Mycobacterial RecA Proteins: Implications for Functional Diversity among RecA Nucleoprotein Filaments. *Biochemistry* 50:300-311. 10.1021/bi1018013

Piepenburg O, Williams CH, Stemple DL, and Armes NA. 2006. DNA Detection Using Recombination Proteins. *PLoS Biology* 4:e204.

Pinard R, de Winter A, Sarkis GJ, Gerstein MB, Tartaro KR, Plant RN, Egholm M, Rothberg JM, and Leamon JH. 2006. Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics* 7:1-21. 10.1186/1471-2164-7-216

Polz MF, and Cavanaugh CM. 1998. Bias in Template-to-Product Ratios in Multitemplate PCR. *Applied and Environmental Microbiology* 64:3724-3730.

Richards SM, Hovhannisyan N, Gilliham M, Ingram J, Skadhauge B, Heiniger H, Llamas B, Mitchell KJ, Meachen J, Fincher GB, Austin JJ, and Cooper A. 2019. Low-cost cross-taxon enrichment of mitochondrial DNA using in-house synthesised RNA probes. *PLoS One* 14:e0209499. 10.1371/journal.pone.0209499

Rohland N, Harney E, Mallick S, Nordenfelt S, and Reich D. 2015. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 370. 10.1098/rstb.2013.0624

667 Rohland N, and Hofreiter M. 2007. Ancient DNA extraction from bones and teeth. *Nature*
668 *Protocols* 2:1756-1762.
669 Rohland N, and Reich D. 2012. Cost-effective, high-throughput DNA sequencing libraries for
670 multiplexed target capture. *Genome Research* 22:939-946.
671 Romiguier J, Ranwez V, Douzery EJP, and Galtier N. 2010. Contrasting GC-content dynamics
672 across 33 mammalian genomes: Relationship with life-history traits and
673 chromosome sizes. *Genome Research* 20:1001-1009. 10.1101/gr.104372.109
674 Santiago-Felipe S, Tortajada-Genaro LA, Morais S, Puchades R, and Maquieira Á. 2015.
675 Isothermal DNA amplification strategies for duplex microorganism detection. *Food*
676 *Chemistry* 174:509-515. http://dx.doi.org/10.1016/j.foodchem.2014.11.080
677 Schubert M, Ginolhac A, Lindgreen S, Thompson JF, AL-Rasheid KA, Willerslev E, Krogh A,
678 and Orlando L. 2012. Improving ancient DNA read mapping against modern
679 reference genomes. *BMC Genomics* 13:178. 10.1186/1471-2164-13-178
680 Schubert M, Lindgreen S, and Orlando L. 2016. AdapterRemoval v2: rapid adapter
681 trimming, identification, and read merging. *BMC Research Notes* 9:88.
682 10.1186/s13104-016-1900-2
683 Soubrier J, Gower G, Chen K, Richards SM, Llamas B, Mitchell KJ, Ho SYW, Kosintsev P, Lee
684 MSY, Baryshnikov G, Bollongino R, Bover P, Burger J, Chivall D, Crégut-Bonnoure E,
685 Decker JE, Doronichev VB, Douka K, Fordham DA, Fontana F, Fritz C, Glimmerveen J,
686 Golovanova LV, Groves C, Guerreschi A, Haak W, Higham T, Hofman-Kamińska E,
687 Immel A, Julien M-A, Krause J, Krotova O, Langbein F, Larson G, Rohrlach A, Scheu A,
688 Schnabel RD, Taylor JF, Tokarska M, Tosello G, van der Plicht J, van Loenen A, Vigne
689 J-D, Wooley O, Orlando L, Kowalczyk R, Shapiro B, and Cooper A. 2016. Early cave
690 art and ancient DNA record the origin of European bison. *Nat Commun* 7:13158.
691 10.1038/ncomms13158:
692 http://www.nature.com/articles/ncomms13158#supplementary-information
693 Su XZ, Wu Y, Sifri CD, and Wellems TE. 1996. Reduced extension temperatures required for
694 PCR amplification of extremely A+T-rich DNA. *Nucleic Acids Research* 24:1574-1575.
695 Tang W, David FB, Wilson MM, Barwick BG, Leyland-Jones BR, and Bouzyk MM. 2009. DNA
696 Extraction from Formalin-Fixed, Paraffin-Embedded Tissue. *Cold Spring Harbor*
697 *Protocols* 2009:pdb.prot5138. 10.1101/pdb.prot5138
698 Tate CM, Nuñez AN, Goldstein CA, Gomes I, Robertson JM, Kavlick MF, and Budowle B.
699 2011. Evaluation of circular DNA substrates for whole genome amplification prior to
700 forensic analysis. *Forensic Science International: Genetics* 6:185-190.
701 Wang G, Maher E, Brennan C, Chin L, Leo C, Kaur M, Zhu P, Rook M, Wolfe JL, and
702 Makrigiorgos GM. 2004. DNA amplification method tolerant to sample degradation.
703 *Genome Research* 14:2357-2366. 10.1101/gr.2813404
704
705

Table 1(on next page)

Table 1. Bison samples

*Previously published carbon dates (Soubrier et al. 2016).

1
2

ACAD Number	Species	Tissue	Country, state	Location	Calibrated Carbon Dates (ORAU lab number)
875	*Bison priscus*	Metacarpal	Russia, Siberia	Alyoshkina Zaimka	>50,000* (OxA-29064)
885	Bison sp.	Humerus	USA, Alaska	Lost Chicken Creek	12,465 ± 75 (OxA-11245)
3133	*Bison priscus*	Astragalus	Canada, Yukon Territory	Irish gulch	26,360 ± 220* (OxA-22141)
4089	Bison sp.	Long bone fragment	Russia, Adyghe	Mezmaiskaya Cave	>50,000* (OxA-19197)

3

Table 2(on next page)

Table 2. Mapping statistics of shotgun libraries

Shotgun libraries were made from four ancient bison and amplified with either PCR or RPA. To eliminate the effect of differences in sequencing depth all libraries were subsampled to 1,850,000 collapsed reads (a number determined from the lowest number of reads for any sample) and then mapped to a bison mitochondrial reference genome (GenBank number: GU947006.1)and a cattlereference genome (GenBank number: UMD 3.1) using BWA and parameters standardly used with aDNA (Schubert et al. 2012). The fraction of unique mapped reads was determined by dividing the number of unique mapped reads by the number of collapsed reads. In comparison to PCR, RPA had a variable effect on the unique reads mapped to a mitochondrial reference, while RPA consistently increased unique reads that mapped to a nuclear reference. Libraries with lower proportions of endogenous DNA experienced the greatest increase (> 60%) in unique reads that mapped to the nuclear reference with RPA. Mapping statistics for the entire shotgun dataset are given in Table S2. ExB = Extraction Blank

Library	Mapped Reads	Unique Mapped Reads	Fraction Unique Mapped reads	% Change in Fraction Unique Mapped Reads in Comparison to PCR
Mapped to Mitochondrial Reference Genome				
PCR - ExB	0	0	0	
PCR -875	594	589	0.000318378	
PCR -885	138	137	7.41E-05	
PCR -3133	109	108	5.84E-05	
PCR -4089	1011	831	0.000449189	
RPA - ExB	0	0	0	
RPA-875	548	519	0.000280541	-13.29
RPA-885	212	188	0.000101622	+27.13
RPA-3133	100	94	5.08E-05	-14.89
RPA-4089	2177	1027	0.000555135	+19.08
Mapped to Nuclear Reference Genome				
PCR - ExB	1055	959	0.000518378	
PCR -875	1030040	1019825	0.551256757	
PCR -885	91919	90799	0.049080541	
PCR -3133	1028050	1017857	0.550192973	
PCR -4089	94391	85283	0.046098919	
RPA - ExB	3628	1175	0.000635135	
RPA-875	1174194	1130883	0.611288108	+10.98
RPA-885	163111	147280	0.079610811	+62.2
RPA-3133	1163979	1114441	0.602400541	+9.49
RPA-4089	322564	169819	0.091794054	+99.12

1

Table 3(on next page)

Table 3. Mapping statistics mtDNA-enriched libraries

Four ancient bison sequencing libraries were enriched for mtDNA using hybridization capture and amplified with either PCR or RPA. To eliminate the effect of differences in sequencing depth all libraries were subsampled to 1,000,000 collapsed reads (a number determined from the lowest number of reads for any sample) and mapped to a bison reference (GenBank number: GU947006.1)using BWA and parameters standardly used with aDNA (Schubert et al. 2012). The fraction of unique mapped reads was determined by dividing the number of unique mapped read by the number of collapsed reads. In comparison to PCR, RPA reduced the fraction of unique reads by >26% in the mapped data. The drop in unique reads observed in the RPA mtDNA-enriched libraries likely stemmed from the amplification of reads that were too short to accurately map. The mapping statistics for the entire mtDNA-enriched dataset is given in Table S3. ExB = Extraction Blank

Library	Mapped Reads	Unique Mapped Reads	Fraction Unique Mapped Reads	% Change in Fraction Unique Mapped Reads in Comparison to PCR
ExB-PCR	1301	5	0.000005	
875-PCR	111165	22998	0.022998	
885-PCR	54943	9532	0.009532	
3133-PCR	25356	10766	0.010766	
4089-PCR	186830	4656	0.004656	
ExB-RPA	0	0	0	
875-RPA	48040	14712	0.014712	-26.03
885-RPA	64621	3370	0.00337	- 64.65
3133-RPA	19910	7135	0.007135	-33.73
4089-RPA	212137	2641	0.002641	-43.28

1

Table 4(on next page)

Table 4. Number of variants detected in mtDNA-enriched mapped data

The number of sequence variants (e.g. single nucleotide polymorphism) detected in four ancient bison sequencing libraries enriched for mtDNA, amplified with either PCR or RPA, and mapped to a modern bison reference mitogenome (GenBank number: GU947006.1)using BWA and parameters standardly used with aDNA (Schubert et al. 2012). For this analysis all data from the mtDNA-enriched libraries were used to maximize the number of variants, which were called using Geneious software (v10.0.08) with the default parameters and a minimum read depth of ³5(Kearse et al. 2012). The denominator of the fraction is the total number of variants detected in the sample and the numerator is the number of variants called in the mtDNA enriched data. The difference between the amplification methods represents the loci that could not be called because of low coverage.

Bison Sample	PCR	RPA
875	78/78	78/78
885	61/61	60/61
3133	59/59	55/59
4089	613/614	376/614

1

Figure 1

Figure 1.Shotgun library split violin plots

Read length and GC content distributions of four shotgun libraries constructed from bison aDNA and amplified with either PCR or RPA. Bison sample numbers are plotted on the x-axis and the “All” violin figures represent the mean of all four bison. The colored area represents the frequency distribution of read length or GC content of the amplification method. The diamonds in the violin plots are the mean of each amplification method and the red dotted line representsthe 41.89% GC content of the nuclear cattle reference (GenBank number: UMD 3.1) used for mapping. In comparison to PCR, amplification with RPA did not have a consistent effect on read length or GC content. Amplification of shotgun libraries with RPA increased the fraction of unique reads by > 9% in comparison to PCR (Table 2). **A**) Shotgun-unmapped: Read Length, **B**) Shotgun-unmapped: GC Content, **C**) Shotgun-mapped: Read Length, and **D**) Shotgun-mapped: GC Content.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27544v1 | CC BY 4.0 Open Access | rec: 19 Feb 2019, publ: 19 Feb 2019

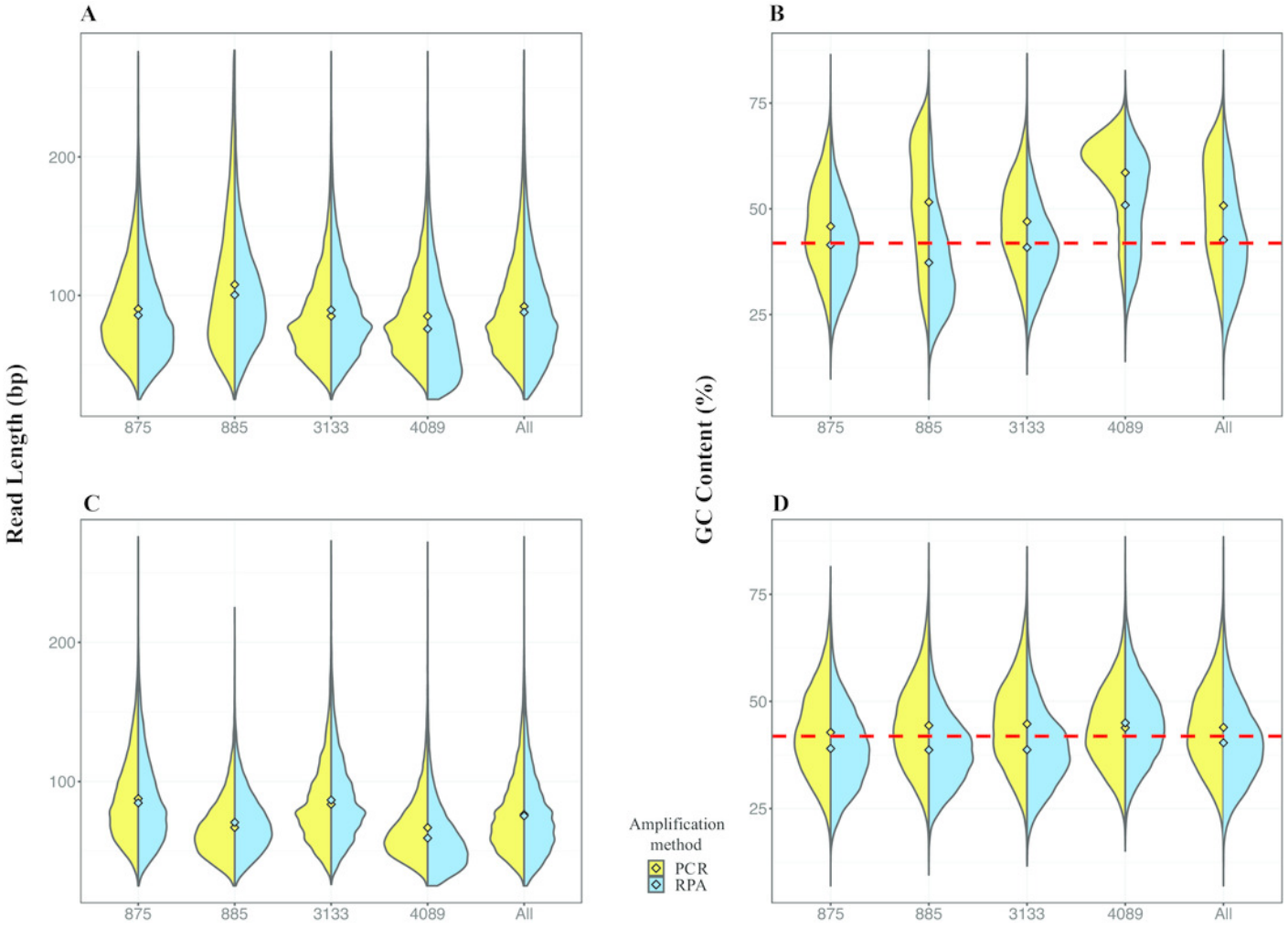


Figure 2

Figure 2.mtDNA-enriched library split violin plots

Read length and GC content of ancient bison libraries enriched for mtDNA using hybridization capture and amplified with either PCR or RPA. Bison sample numbers are plotted on the x-axis and the “All” violin figures represent the mean of all four bison. The colored area represents the frequency distribution of read length or GC content of the amplification method. The diamonds in the violin plots are the mean of each amplification method and the red dotted line represents the 39.14% GC content of the bison mitochondrialreference (GenBank number: GU947006.1) used for mapping. RPA reduced the mean read length of mtDNA-enriched data by ≈ 30 bp in comparison to PCR amplification. The reduction in read length produced by the isothermal method appears to have favored the amplification of small unmappable DNA molecules and reduced the fraction of unique mapped reads (by > 26%) in the RPA mtDNA enriched data (Table 3). **A**) mtDNA-unmapped: Read Length, **B**) mtDNA-unmapped: GC Content, **C**) mtDNA-mapped: Read Length, and **D**) mtDNA-mapped: GC Content.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27544v1 | CC BY 4.0 Open Access | rec: 19 Feb 2019, publ: 19 Feb 2019

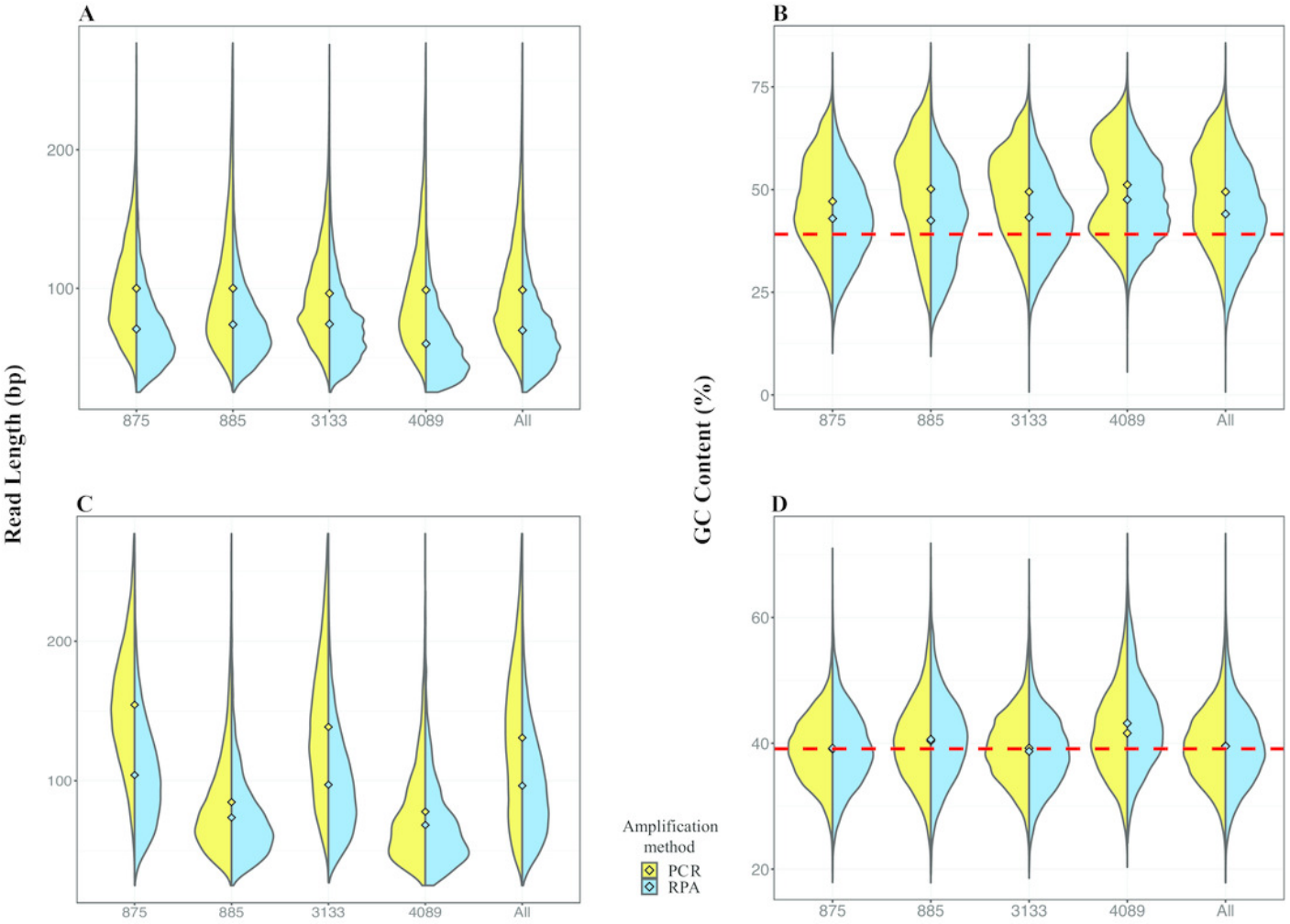
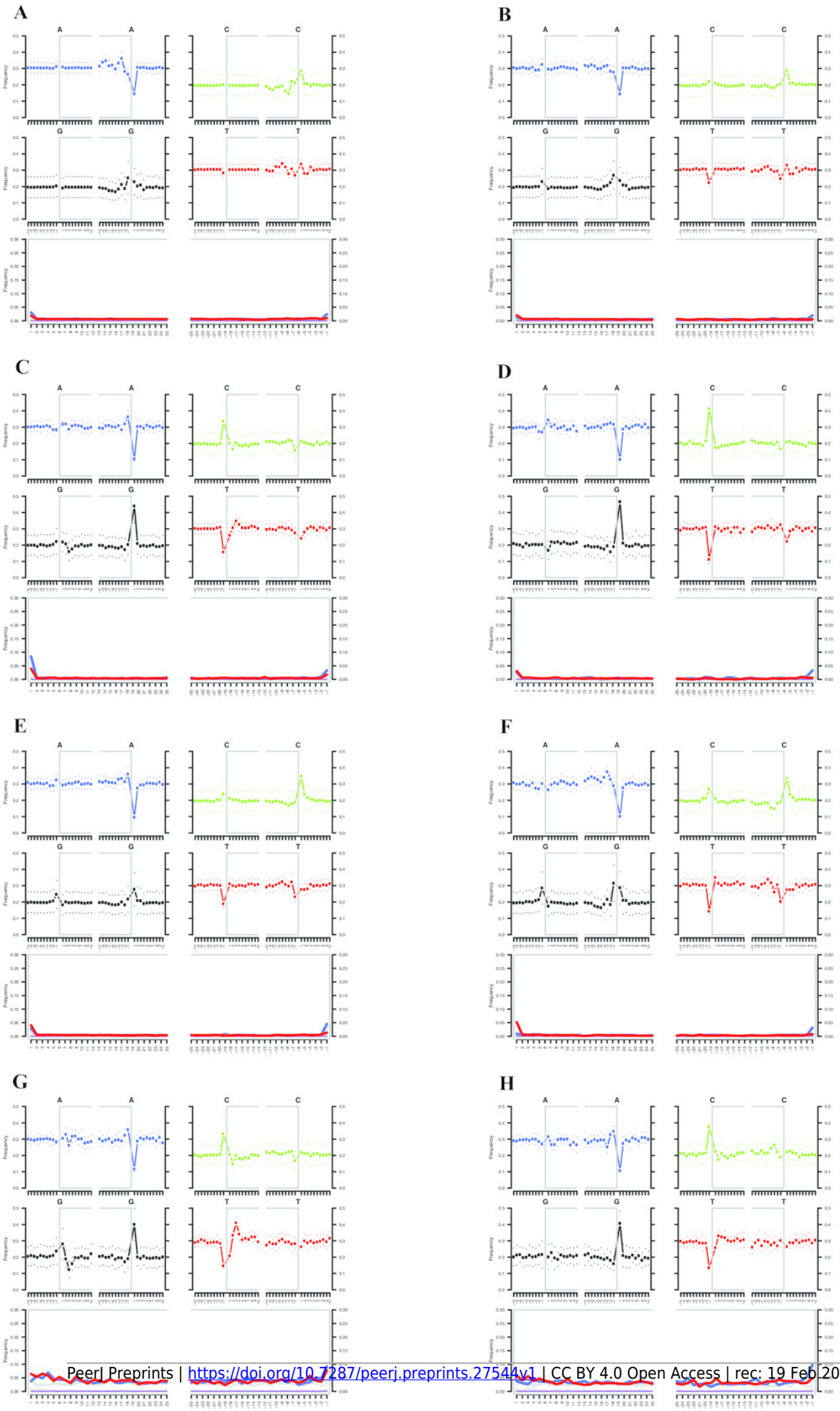

A
B
C
D
Read Length (bp)
GC Content (%)
200
100
75
50
25
0
60
40
20
875
885
3133
4089
All
Amplification method
PCR
RPA

Figure 3

Figure 3. PCR and RPA damage profiles

Damage profiles of four bison mtDNA-enriched libraries amplified with either PCR or RPA and mapped to a modern bison reference mitogenome (GenBank number: GU947006.1) were generated using mapDamage 2.0 (Jónsson et al. 2013). The similar profiles between the two amplification methods indicates that the DNA polymerase provided in the TwistAmp kit will read through and misincorporate an adenosine when encountering a deaminated cytosine.**A**) Bison 875-PCR,**B**) Bison 875-RPA, **C**) Bison 885-PCR, **D**) Bison 885-RPA, **E**) Bison 3313-PCR, **F**) Bison 3313-RPA, **G)**Bison 4089-PCR, and **H**) Bison 4089-RPA


A
B
C
D
E
F
G
H