Supplemental Material

Extending the spectrum of DNA sequences retrieved from ancient bones and teeth

Isabelle Glocke and Matthias Meyer

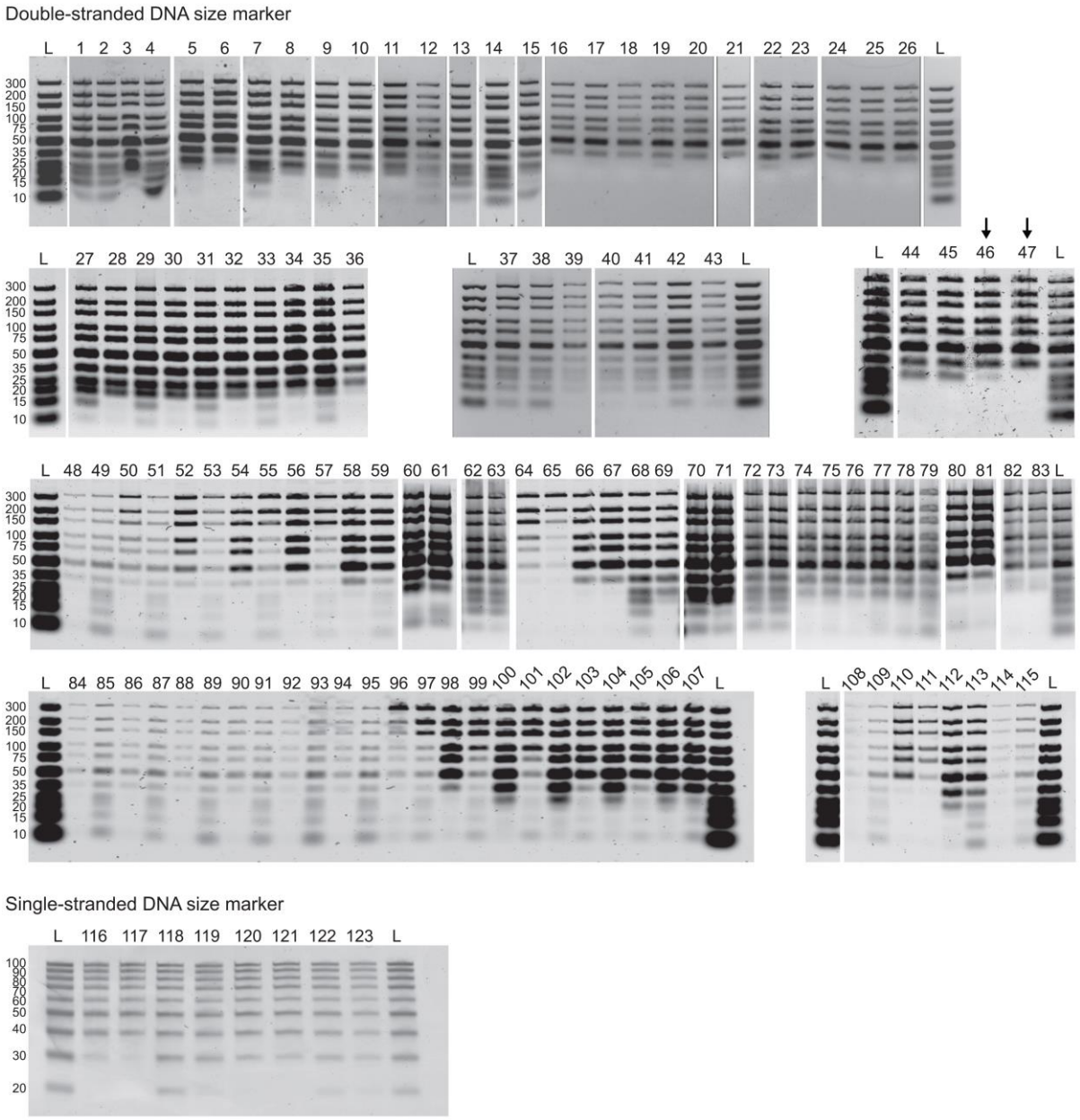

Figure S1: Recovery of a double- and a single-stranded DNA size marker purified on silica spin columns with different binding buffers. The double-stranded DNA size marker was visualized on 4% agarose gels and the single-stranded DNA size marker on a 10% TBE-Urea gel (Criterion, Bio-Rad). Sample loading volumes were adjusted so that they correspond to 1 µg of double-stranded DNA / 20 ng of single-stranded DNA if recovery in the extracts was 100%. The same amount of DNA was loaded in the size marker lanes (L). Intensities of bands on different gels cannot be compared. Arrows point to purification conditions of method A. For a description of the experiments see Table S1.

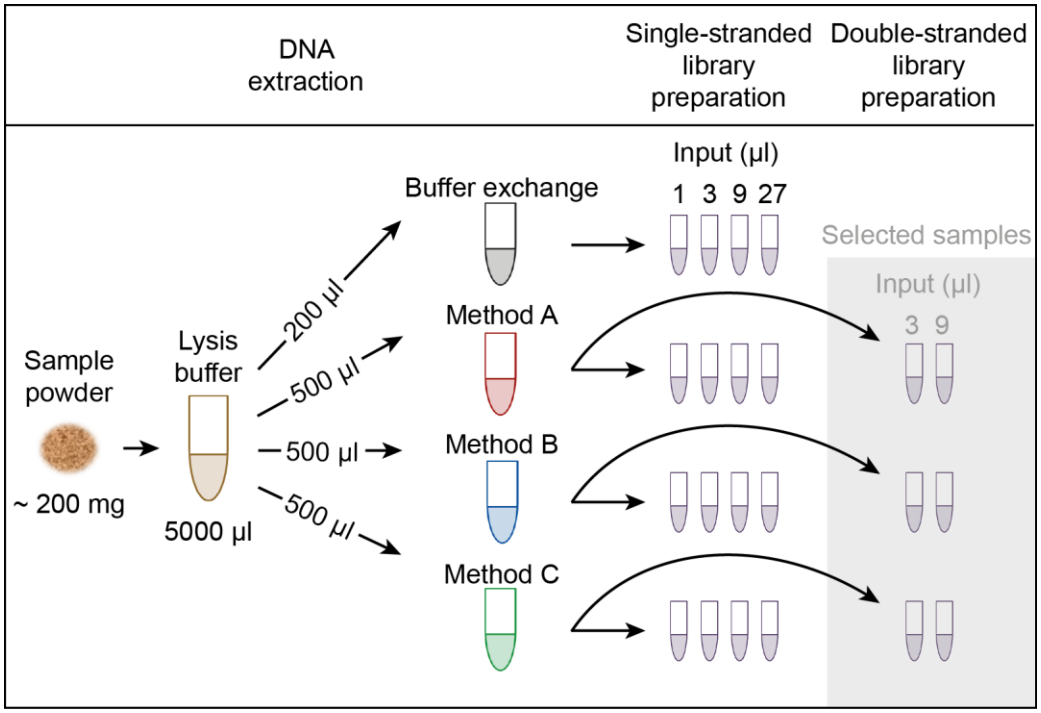

Figure S2: Overview of the experiment design.

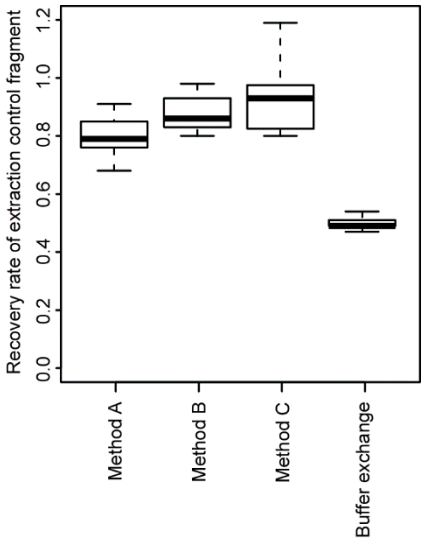

Figure S3: Recovery rate of the extraction control DNA fragment in the extracts prepared from seven ancient samples. Thick black lines depict medians, boxes represent the lower and upper quartile and whiskers extend to values within 1.5 times the interquartile range.

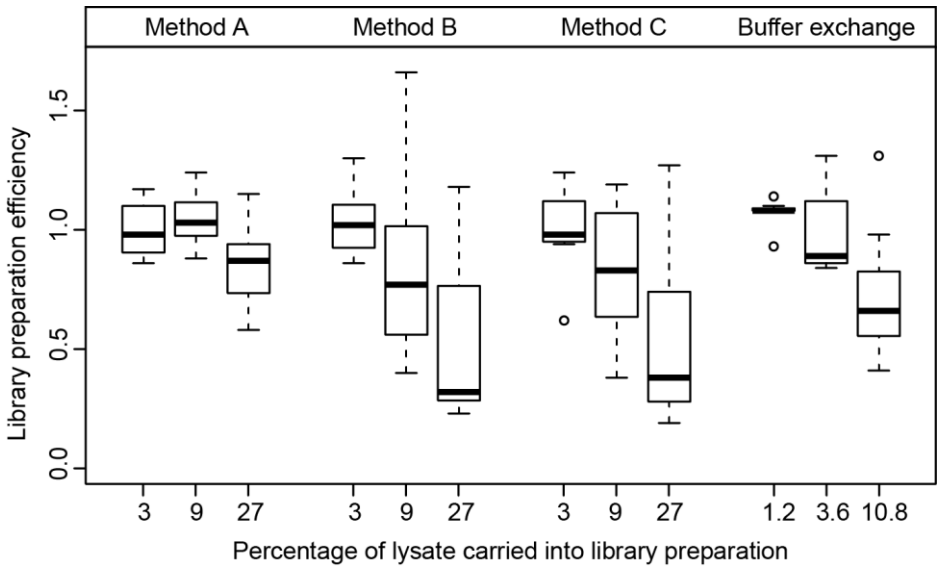

Figure S4: Library preparation efficiencies for each of the extraction methods inferred from the input-output relationship of library preparation from seven ancient samples. 500 µl of bone/tooth powder lysate were used as input for silica-based DNA extraction and 200 µl for buffer exchange. Thick black lines depict medians, boxes represent the lower and upper quartile, whiskers extend to values within 1.5 times the interquartile range and circles represent outliers.

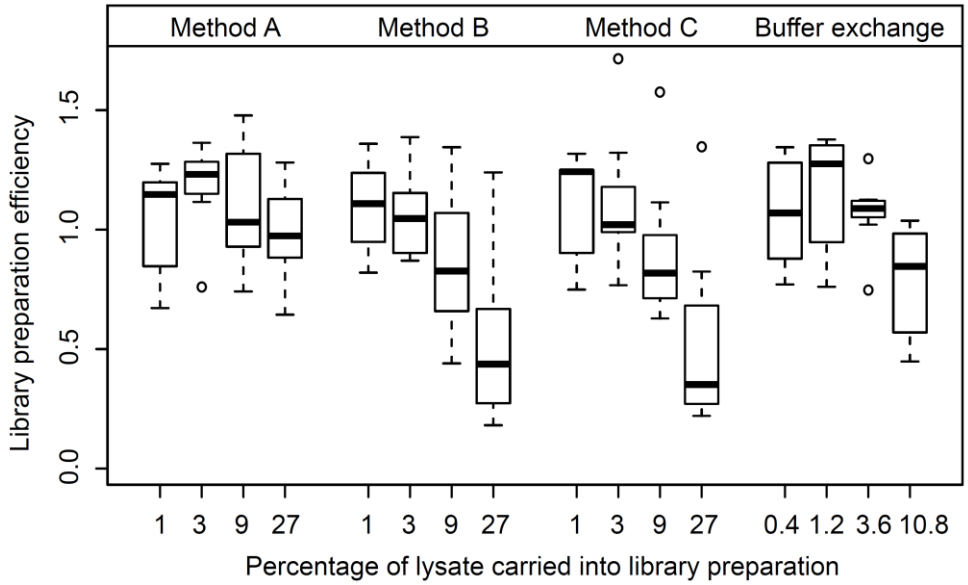

Figure S5: Library preparation efficiencies for each of the extraction methods inferred from the conversion rate of the library control oligonucleotide for the seven ancient samples Thick black lines depict medians, boxes represent the lower and upper quartile, whiskers extend to values within 1.5 times the interquartile range and circles represent outliers.

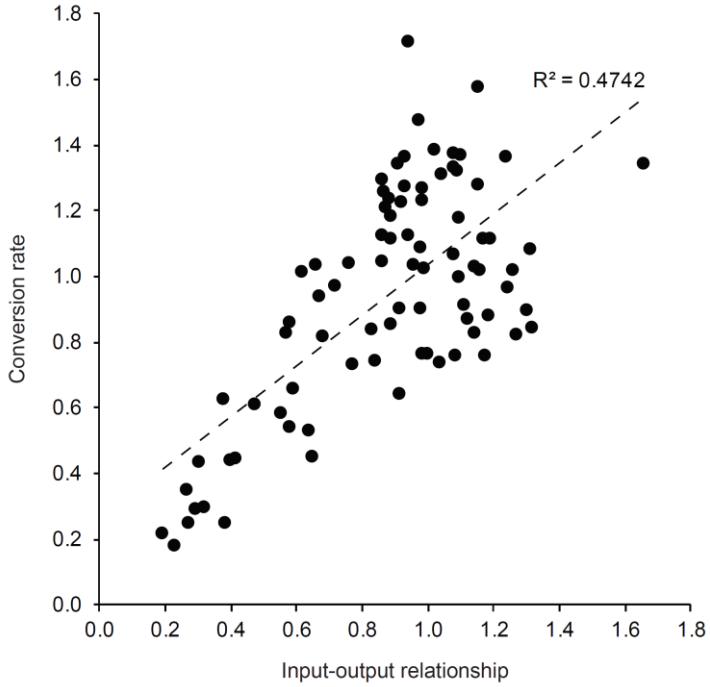

Figure S6: Correlation between the two quality control measures used to detect inhibition during library preparation: conversion rate of the library control oligonucleotide and library preparation efficiency inferred from the input-output relationship.

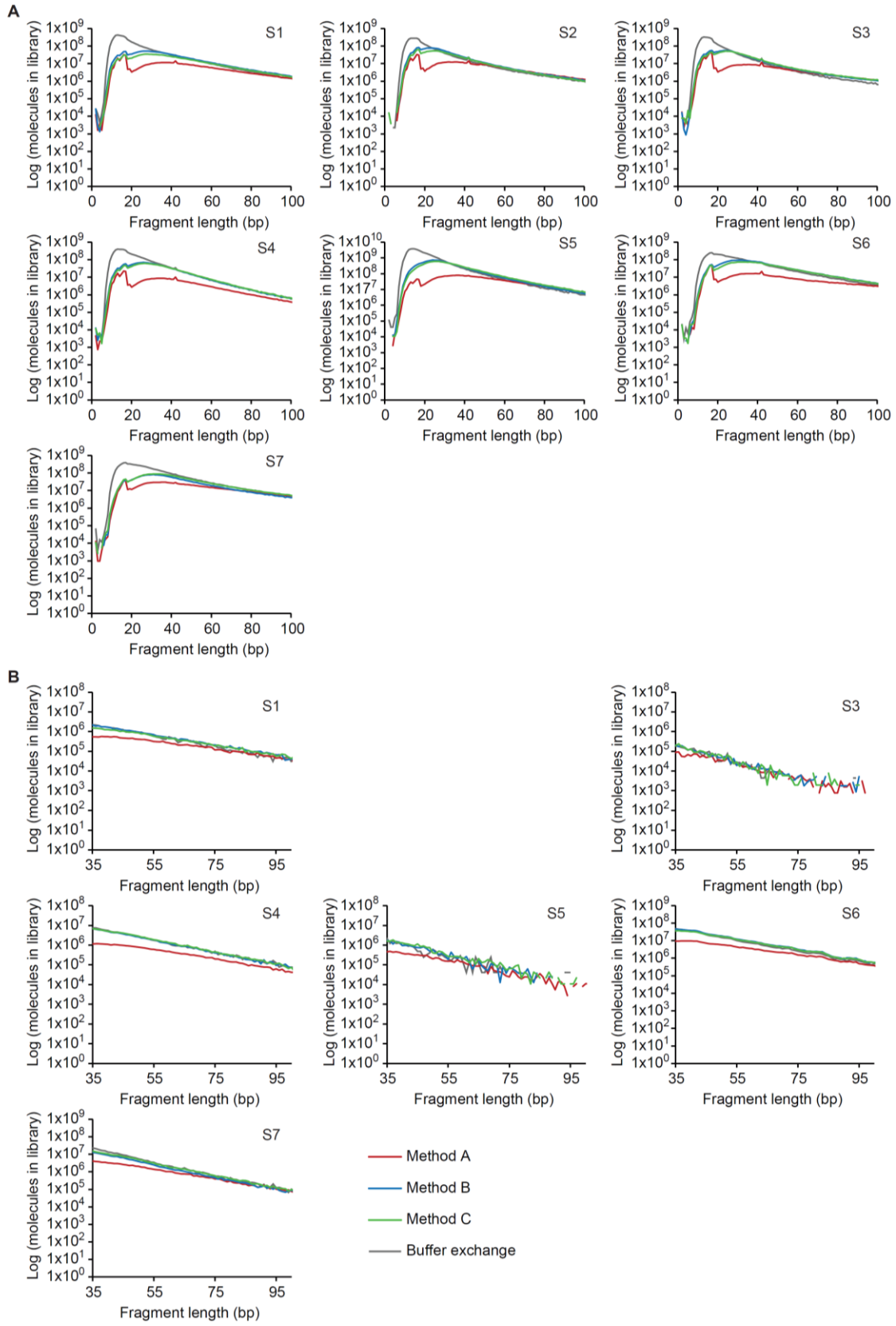

Figure S7: Estimates of the number of molecules in each library binned by size and plotted on a log-scale for **(A)** all molecules and **(B)** 'informative' molecules, i.e. molecules that aligned to the respective reference genome. Numbers for the libraries prepared from buffer exchange extracts were multiplied by 2.5 to compensate for the smaller volume of lysate used. Peaks below 20 bp are due to artifacts from library preparation, which appear only if very little input DNA is used.

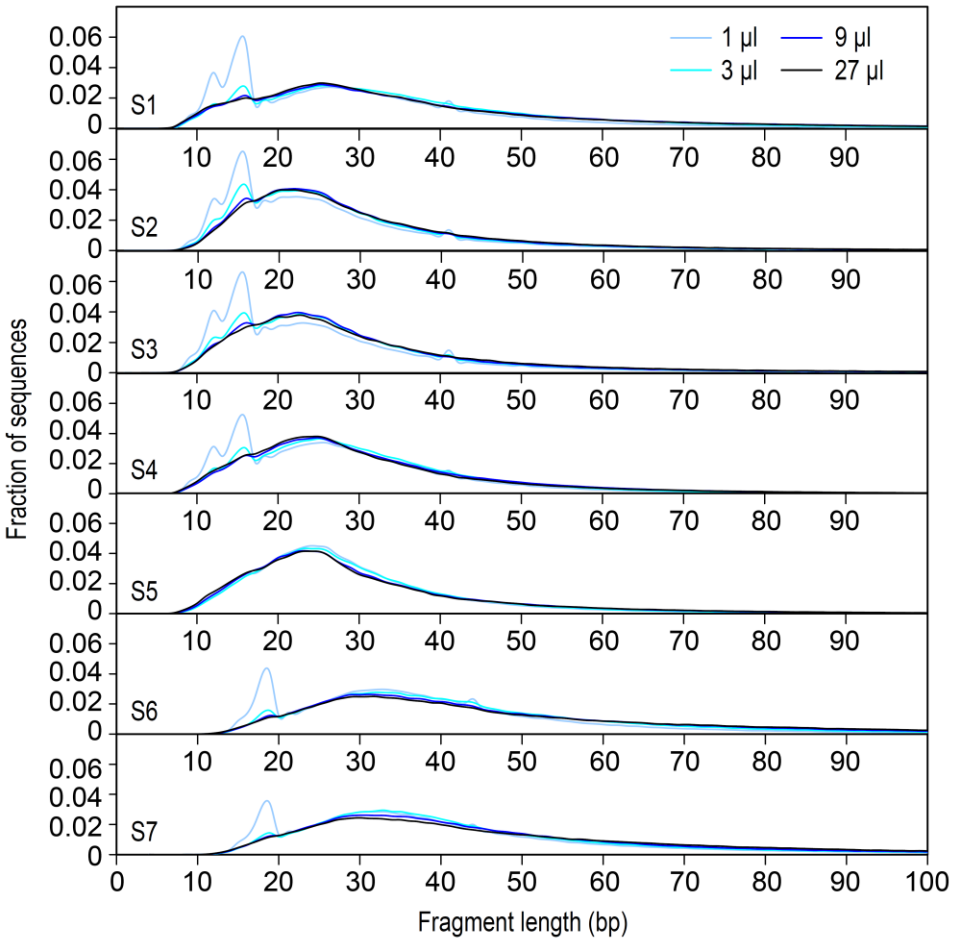

Figure S8: Fragment length distribution in libraries prepared from different volumes of extract prepared with method B. Artifacts from library preparation are visible in libraries prepared from small quantities of input DNA.

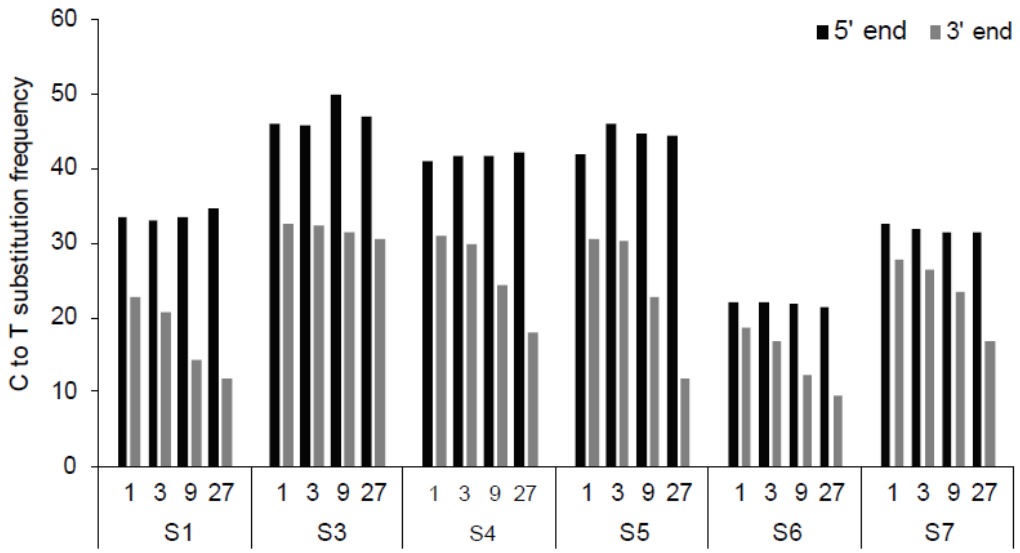

Figure S9: Terminal cytosine (C) to thymine (T) substitution frequencies in aligned sequences ≥ 35 bp of six ancient samples. Libraries were prepared from different volumes of extract (1, 3, 9 and 27 µl). Extracts were generated using method B. The deprivation in C to T substitution frequencies at 5' ends compared to 3' ends of sequence alignments is likely due to inefficient ligation of uracil-containing 3' ends by the *T4* DNA ligase used in library preparation (Gansauge et al. 2017).

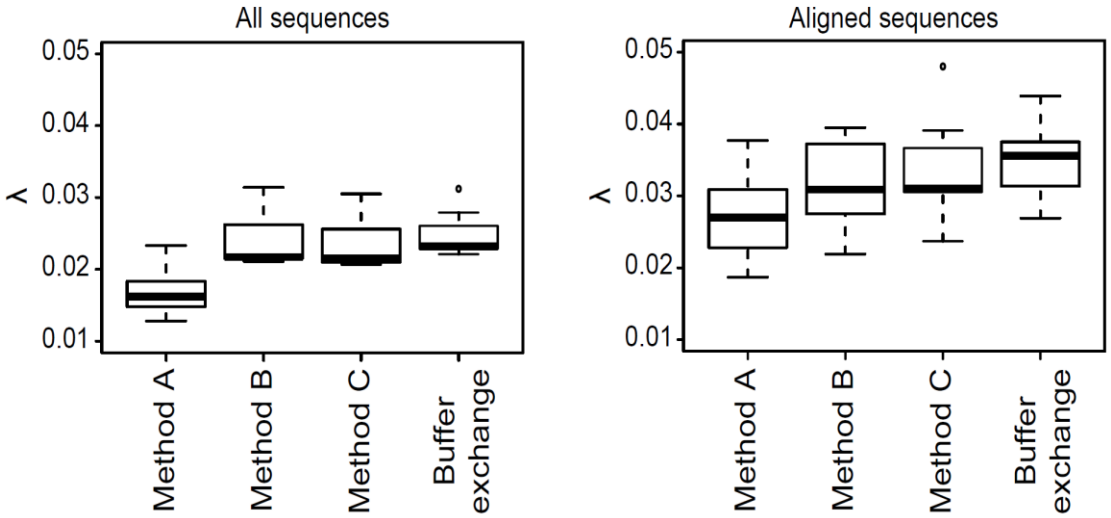

Figure S10: Boxplots showing the damage frequencies (λ) in the DNA of seven ancient samples extracted with the different methods. λ was inferred from all sequences (left plot) and those that aligned to a respective reference genome (right plot). Thick black lines depict medians, boxes represent the lower and upper quartile, whiskers extend to values within 1.5 times the interquartile range and circles represent outliers.

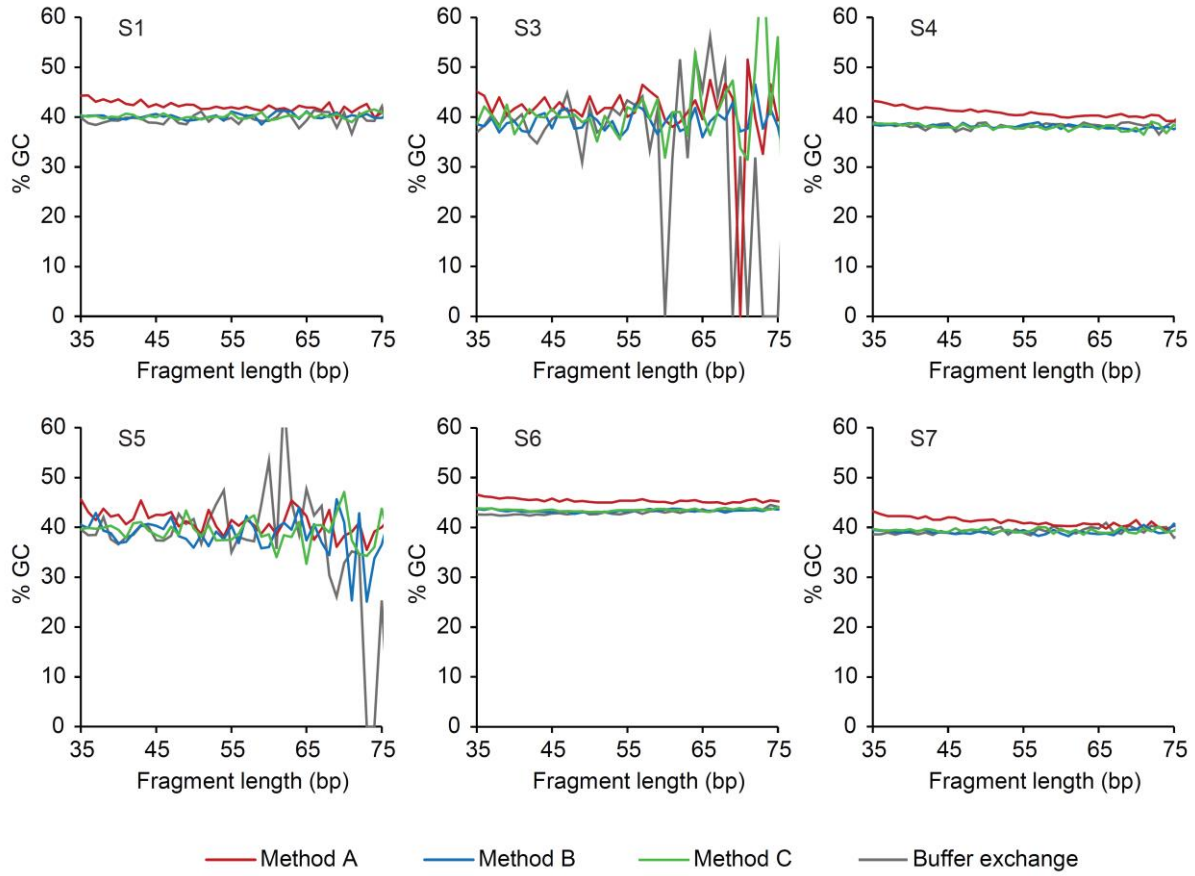

Figure S11: GC content of aligned DNA sequences binned by size. Plots for samples 3 and 5 are noisy due to the small number of aligned sequences available for analysis.

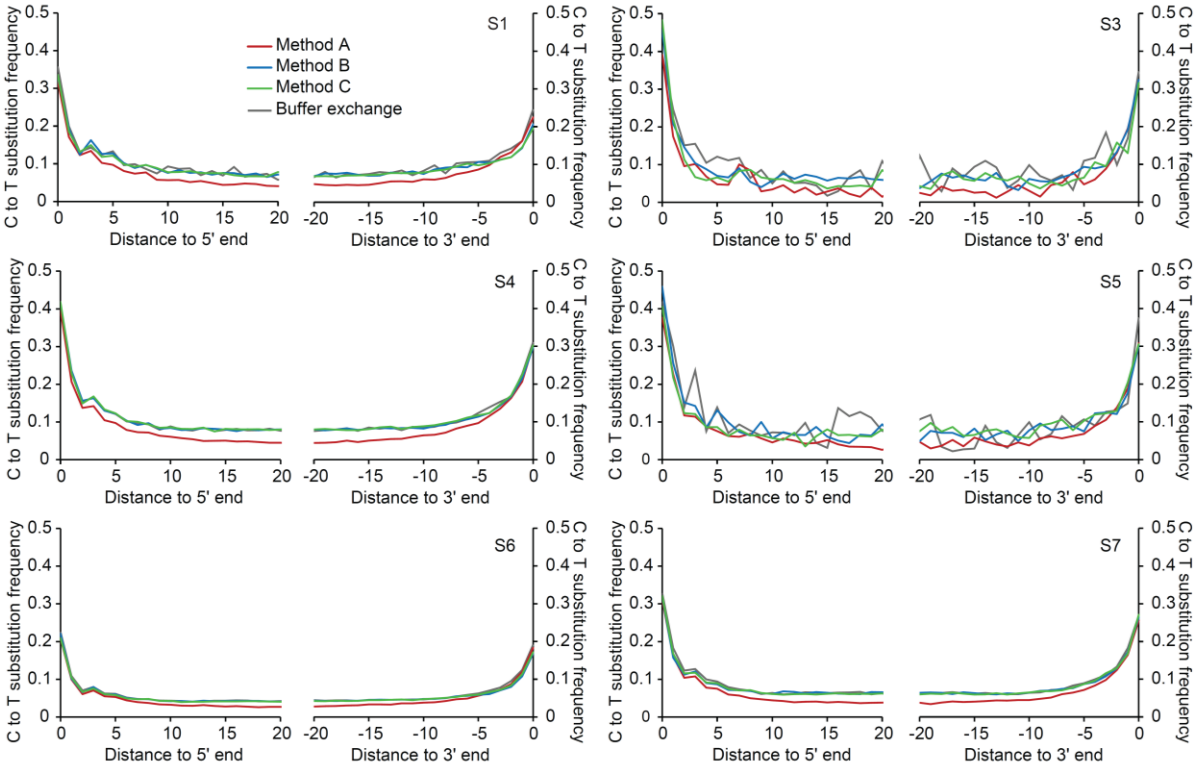

Figure S12: Cytosine (C) to thymine (T) substitution frequencies in aligned sequences ≥ 35 bp plotted as a function of the distance to the 5’ and 3’ ends. Plots for samples 3 and 5 are noisy due to the small number of aligned sequences available for analysis.

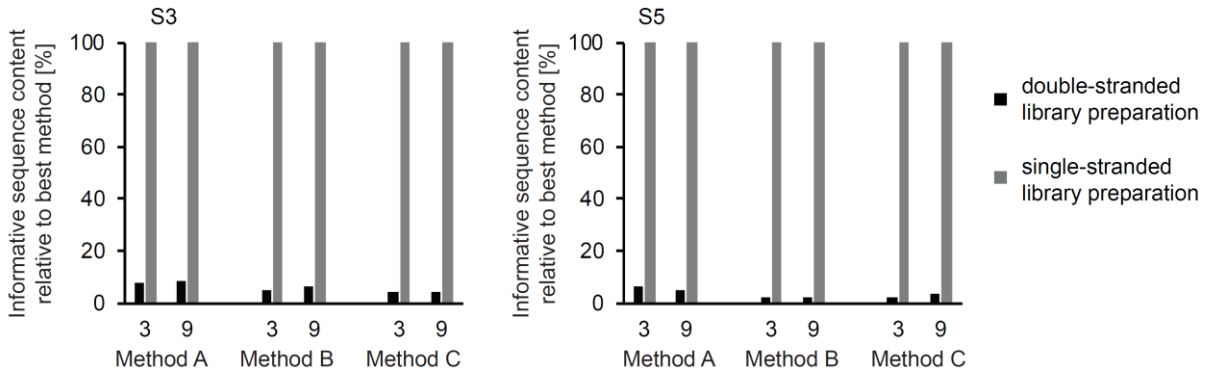

Figure S13: Informative sequence content relative to the best method (the method recovering the highest number of informative sequences) in libraries prepared from the extracts of two ancient samples using single- and double-stranded library preparation. Two volumes of extract (3 and 9 µl) were used for library preparation.

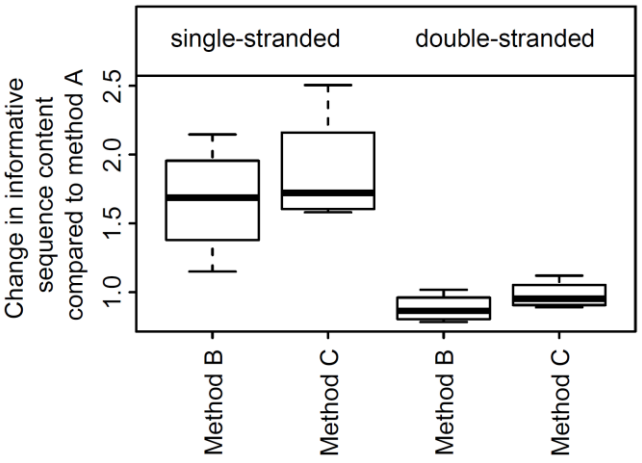

Figure S14: Boxplot depicting the change in informative sequence content obtained by using extraction methods B and C instead of method A for two library preparation methods. The comparison is based on only two samples (S3 and S5) of which double-stranded libraries were prepared. Thick black lines depict medians, boxes represent the lower and upper quartile and whiskers extend to values within 1.5 times the interquartile range.

Table S2: Samples used in this study.

Sample	Sample ID	Type of sample	Age	Powder (mg)	Lysis buffer (µl)	Extraction method	Lysis buffer used (µl)	Corres-ponds to sample material (mg)	Extract ID	ss library preparation	ds library preparation
1: Cave bear (Gamsulzen cave)	SP891	bone	~44 kya	224	5000	Buffer Exchange	200	9.0	E3418	+	-
						Method A	500	22.4	E3426	+	-
						Method B	500	22.4	E3434	+	-
						Method C	500	22.4	E3442	+	-
2: Cave bear (Sima de los Huesos cave)	SP3177	bone	~430 kya	210	5000	Buffer Exchange	200	8.4	E3419	+	-
						Method A	500	21.0	E3427	+	-
						Method B	500	21.0	E3435	+	-
						Method C	500	21.0	E3443	+	-
3: Cave bear (Vindija cave)	SP2689	bone	~47 kya	196	5000	Buffer Exchange	200	7.8	E3420	+	-
						Method A	500	19.6	E3428	+	+
						Method B	500	19.6	E3436	+	+
						Method C	500	19.6	E3444	+	+
4: Brown bear (Denisova cave)	SP3386	tooth	Late Pleistocene	204	5000	Buffer Exchange	200	8.2	E3421	+	-
						Method A	500	20.4	E3429	+	-
						Method B	500	20.4	E3437	+	-
						Method C	500	20.4	E3445	+	-
5: Yak (Denisova cave)	SP3391	bone	> 49 kya	197	5000	Buffer Exchange	200	7.9	E3422	+	-
						Method A	500	19.7	E3430	+	+
						Method B	500	19.7	E3438	+	+
						Method C	500	19.7	E3446	+	+
6: Bison (Yukon permafrost)	SP3424	bone	Late Pleistocene	222	5000	Buffer Exchange	200	8.9	E3423	+	-
						Method A	500	22.2	E3431	+	-
						Method B	500	22.2	E3439	+	-
						Method C	500	22.2	E3447	+	-
7: Beluga whale (North Sea)	SP1108	bone	Holocene	234	5000	Buffer Exchange	200	9.4	E3424	+	-
						Method A	500	23.4	E3432	+	-
						Method B	500	23.4	E3440	+	-
						Method C	500	23.4	E3448	+	-

Table S4: Summary of sequencing results obtained from double-stranded library preparation from a subset of extracts.

DNA extraction and library preparation					Shotgun sequencing								
Sample	**Extraction method**	**Volume of extract used for library prep-aration [µl]**	**Library ID**	**№ library molecules**	**Reference genome**	**Sequences**	**Overlap-merged sequences**	**Overlap-merged sequences ≥35bp**	**Aligned sequences ≥35bp**	**Average length of aligned sequences ≥35bp**	**Informative sequence content in library**	**5' C to T substi-tution frequency**	**3' G to A substi-tution frequency**
3: Cave bear (Vindija cave)	Method A	3	R9166	$2.20x10^8$	ursMar0	$9.22x10^5$	$8.67x10^5$	$8.13x10^5$	$3.34x10^2$	52.70	$4.20x10^6$	26.7	35.4
		9	R9167	$3.39x10^8$	ursMar0	$1.13x10^6$	$1.04x10^6$	$9.62x10^5$	$1.01x10^3$	52.86	$1.60x10^7$	33.2	35.0
	Method B	3	R9169	$1.94x10^8$	ursMar0	$7.20x10^5$	$6.80x10^5$	$6.24x10^5$	$3.16x10^2$	50.28	$4.27x10^6$	36.0	39.5
		9	R9170	$2.84x10^8$	ursMar0	$1.03X10^6$	$9.50x10^5$	$8.63x10^5$	$9.12x10^2$	52.64	$1.32x10^7$	34.6	39.3
	Method C	3	R9172	$1.98x10^8$	ursMar0	$9.23x10^5$	$8.65x10^5$	$7.97x10^5$	$3.74x10^2$	51.43	$4.13x10^6$	32.7	33.3
		9	R9173	$3.25x10^8$	ursMar0	$9.25x10^5$	$8.50 x10^5$	$7.74x10^5$	$7.99x10^2$	52.75	$1.48x10^7$	37.2	32.3
5: Yak (Denisova cave)	Method A	3	R9175	$3.73x10^8$	bosTau6	$1.15x10^6$	$1.10x10^6$	$1.00x10^6$	$1.49x10^3$	53.69	$2.61x10^7$	37.3	36.0
		9	R9176	$8.31x10^8$	bosTau6	$1.13x10^6$	$1.08x10^6$	$9.75x10^5$	$1.53x10^3$	53.53	$6.03x10^7$	36.1	39.3
	Method B	3	R9177	$3.56x10^8$	bosTau6	$1.17x10^6$	$1.13x10^6$	$9.51x10^5$	$1.29x10^3$	51.90	$2.04x10^7$	39.6	34.3
		9	R9178	$8.13x10^8$	bosTau6	$1.25x10^6$	$1.22x10^6$	$9.87x10^5$	$1.63x10^3$	51.84	$5.46x10^7$	37.5	36.2
	Method C	3	R9179	$4.19x10^8$	bosTau6	$1.18x10^6$	$1.14x10^6$	$9.83x10^5$	$1.27x10^3$	51.43	$2.32x10^7$	34.4	35.7
		9	R9180	$9.50x10^8$	bosTau6	$1.26x10^6$	$1.22x10^6$	$1.01x10^6$	$1.71x10^3$	52.34	$6.76x10^7$	34.2	34.3
8: extraction blank	Method A	9	R9163	$1.89x10^8$	hg19	$1.01x10^5$	9.77E+04	$9.23x10^4$	26.00	50.65	$2.45x10^6$	0.0	0.0
	Method B	9	R9164	$1.78x10^8$	hg19	$1.18x10^5$	$1.13x10^5$	$1.06x10^5$	12.00	42.17	$7.64x10^5$	0.0	0.0
	Method C	9	R9165	$1.45x10^8$	hg19	$1.37x10^5$	$1.32 x10^5$	$1.25x10^5$	15.00	41.33	$6.57x10^5$	0.0	0.0
library blank			R9181	$1.74x10^8$	hg19	$7.38x10^4$	$7.10x10^4$	$6.61x10^4$	2.00	37.50	$1.77x10^5$	0.0	0.0

Table S5: Oligonucleotides used in this study.

Name	Description	Sequence (5' - 3')	Published in
CL53	Library preparation double-stranded adapter, strand 1	CGACGCTCTTC-ddC	Gansauge and Meyer 2013
CL73	Library preparation double-stranded adapter, strand 2	[phosphate]GGAAGAGCGTCGTGTAGGGAAAGAG*T*G*T*A	Gansauge and Meyer 2013
CL78	Library preparation adapter	[phosphate]AGATCGGAAG[C3Spacer]$_{10}$-TEG-biotin	Gansauge and Meyer 2013
CL104	Library preparation control	[phosphate]TCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGA[phosphate]	Gansauge and Meyer 2013
CL107	qPCR primer	TCATGTAACTCGCCTTGATCGT	Gansauge and Meyer 2013
CL118	qPCR probe	[FAM]TTCAGCTCCGGTTCCCAACGAT[BHQ1]	
CL130	Library preparation extension primer	GTGACTGGAGTTCAGACGTGTGCTCTTCC*GA*TC*T	Korlevic et al. 2015
CL200	Extraction control, strand 1	[phosphate]TATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTA[phosphate]	
CL201	Digital PCR primer	TATCCGCTCACAATTCCACA	
CL202	Digital PCR primer	TAGGCACCCCAGGCTTTAC	
CL204	Extraction control, strand 2	[phosphate]TAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATA[phosphate]	
CL304	Library preparation control	[phosphate]ATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGA[phosphate]	
IS5	PCR primer	AATGATACGGCGACCACCGA	Meyer and Kircher 2010
IS6	PCR primer	CAAGCAGAAGACGGCATACGA	Meyer and Kircher 2010
IS7	qPCR primer	ACACTCTTTCCCTACACGAC	Meyer and Kircher 2010
IS8	qPCR primer	GTGACTGGAGTTCAGACGTGT	Meyer and Kircher 2010
P5 index primer	Index primer	AATGATACGGCGACCACCGAGATCTACAC[index]ACACTCTTTCCCTACACGACGCTCTT	Kircher et al. 2012
P7 index primer	Index primer	CAAGCAGAAGACGGCATACGAGAT[index]GTGACTGGAGTTCAGACGTGT	Kircher et al. 2012
TL110	Library preparation splinter	SpacerC12-AA[SpacerC12]CTTCCGATCTNNNNNN-AmC6	Gansauge et al. 2017

AmC6 – 3' amino modifier C6, BHQ1 – black hole quencher, ddC – dideoxy cytosine, N – mixture of adenine, thymine, cytosine and guanine, FAM – fluorescein, * – phosphorothioate bond, TEG – triethylene glycol.

REFERENCES

Gansauge MT, Gerber T, Glocke I, Korlevic P, Lippik L, Nagel S, Riehl LM, Schmidt A, Meyer M. 2017. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res* doi:10.1093/nar/gkx033.

Gansauge MT, Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature protocols* **8**: 737-748.

Kircher M, Sawyer S, Meyer M. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* **40**.

Korlevic P, Gerber T, Gansauge MT, Hajdinjak M, Nagel S, Aximu-Petri A, Meyer M. 2015. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**: 87-93.

Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* **2010**.