

Library construction for ancient genomics: Single strand or double strand?

E. Andrew Bennett, Diyendo Massilani, Giulia Lizzo, Julien Daligault, Eva-Maria Geigl, and Thierry Grange

Institut Jacques Monod, CNRS, Université Paris Diderot, Paris, France

BioTechniques 56:289-300 (June 2014) doi 10.2144/000114176

Keywords: ancient DNA; palaeogenomics; DNA library preparation; next-generation sequencing

Supplementary material for this article is available at www.BioTechniques.com/article/114176.

A novel method of library construction that takes advantage of a single-stranded DNA ligase has been recently described and used to generate high-resolution genomes from ancient DNA samples. While this method is effective and appears to recover a greater fraction of endogenous ancient material, there has been no direct comparison of results from different library construction methods on a diversity of ancient DNA samples. In addition, the single-stranded method is limited by high cost and lengthy preparation time and is restricted to the Illumina sequencing platform. Here we present in-depth comparisons of the different available library construction methods for DNA purified from 16 ancient and modern faunal and human remains, covering a range of different taphonomic and climatic conditions. We further present a DNA purification method for ancient samples that permits the concentration of a large volume of dissolved extract with minimal manipulation and methodological improvements to the single-stranded method to render it more economical and versatile, in particular to expand its use to both the Illumina and the Ion Torrent sequencing platforms. We show that the single-stranded library construction method improves the relative recovery of endogenous to exogenous DNA for most, but not all, of our ancient extracts.

The use of high-throughput sequencing techniques in the field of ancient DNA research has been essential for reconstructing the genomes of ancient or extinct organisms, as well as offering insight into past climates (1), ancient population demographics (2), and historic pathogens (3). However, current protocols for constructing libraries, which require modifications at the ends of double-stranded molecules and multiple purification steps, are particularly poorly suited for the extremely low quantities of fragmented and damaged DNA found in ancient and degraded organic material.

The common strategy for preparing DNA libraries for high-throughput sequencing ligates sequencing adaptors to fragmented DNA by one of two methods. Blunt-end ligation of partially double-stranded adaptor pairs (4–5) requires fewer steps to prepare the DNA for ligation than alternative methods, but due to the indiscriminate nature of this ligation, the 50% of molecules that by chance receive non-distinct adapters are lost from the library. Y-shaped adapters, introduced by Illumina, use an A-tailing reaction to create overhangs in order to introduce directionality into the ligation step and

ensure that each molecule is ligated to distinct adaptor pairs (6).

In contrast to both of these methods, a novel method of library construction has been recently described and used to generate high-resolution genomes from two ancient hominins from the Denisova Cave (7–8), as well as mitochondrial genomes from 30–50 bp DNA fragments from the bones of a 300,000 year-old cave bear (9), and a 400,000 year-old hominin (10). This method makes use of a single-stranded DNA ligase and a 5' phosphorylated and biotinylated adapter oligonucleotide to first capture and immobilize to

METHOD SUMMARY

Here we compare the results of double- and single-stranded sequencing library preparations from 16 diverse ancient and modern samples. We also experimentally investigate purported limitations of the single-stranded DNA ligase used for the single-stranded library preparation protocol and offer methods to expand and facilitate the method.

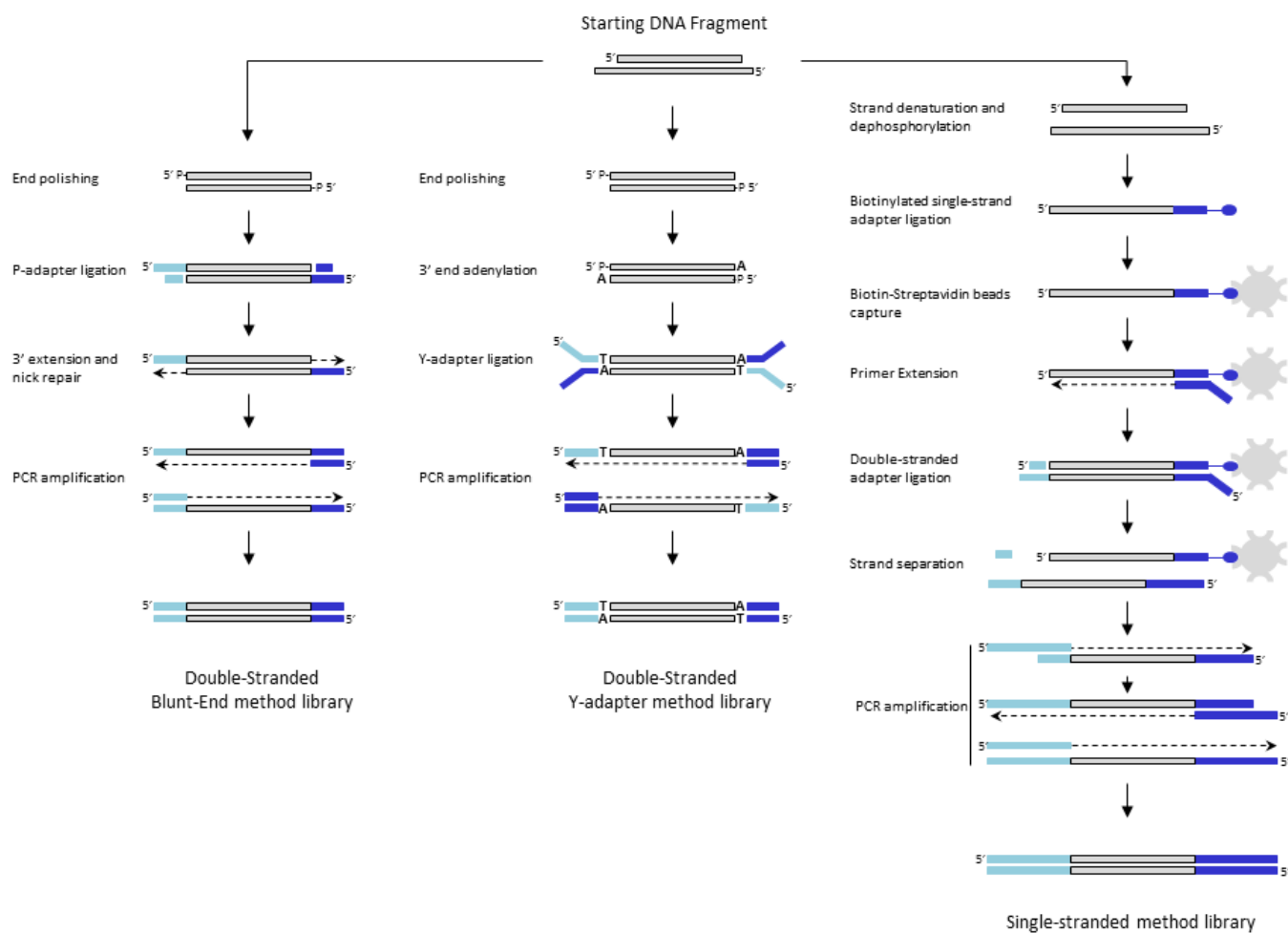


Figure 1. Current library preparation methods for high-throughput sequencing of damaged DNA. Blunt-end method: 3' overhangs are extended and 5' overhangs are removed to create blunt ends with 5' phosphates and 3' OH groups. Complementary non-phosphorylated P-adapters are then ligated to the ends of double-stranded molecules, and the 3' ends of ligated fragments are extended. Use of a polymerase with nick-translation or strand-displacement activity at this step can also repair nicks occurring within the insert, if preceded by a 3' OH. The library is then ready for amplification, but any nicks or gaps that were not successfully extended will be lost during denaturation, and inserts with identical adapters at each end (50% of the final library) will not be sequenced. Y-adaptor method: DNA fragments are end-repaired as with the blunt-end method, followed by an A-tailing reaction, which requires an additional purification step. Y-adapters with 3' dT overhangs are ligated, thus ensuring ligation of distinct adapters at each end. Strands containing nicks or gaps will be lost during denaturation. Single-stranded library method: DNA fragments are dephosphorylated and denatured, then ligated to the 5' phosphate of a 3'-biotinylated oligonucleotide using a single-stranded DNA ligase. These products are then bound to streptavidin beads to allow bead-purification and extended using a primer with 5' adapter tail. After removing the 3' dA, a second adaptor is ligated, and the finished product is eluted to allow PCR amplification using barcoded primers.

beads single-stranded DNA molecules without prior end-repair. A primer uses the products of this ligation to generate complementary strands, resulting in double-stranded DNA, which then receives a second adaptor via blunt-end ligation. The molecules are finally heated to release the finished single strand, which is used to complete the adaptor sequence through an amplification reaction (11) (Figure 1). While the advantages of this method in the recovery of ancient DNA have been demonstrated, a systematic comparison with traditional methods on a diverse array of samples has yet to be performed. In addition,

higher cost, a longer preparation time, and sequencing platform restrictions (11) remain possible obstacles to its broader application.

The ability to incorporate ancient DNA into DNA libraries for sequencing is hindered by the degradation of these molecules over time. This process predominantly results in nicks initiated through the generation of abasic sites from the hydrolysis of N-glycosyl bonds, followed by hydrolysis of the deoxyribose, and the eventual fragmentation of the DNA into progressively smaller molecules with single-stranded overhangs.

Cytosine deamination, which occurs more frequently at the single-stranded ends of these molecules, is a common process, and the accumulating uracils not only can lead to C to T transitions in the final sequences (12–14), but may also inhibit copying by most proof-reading polymerases (15). Depending on the nature of the damage and the techniques used during library preparation, ancient DNA fragments containing nicks or gaps may be lost when using library construction methods that require double-stranded molecules for ligation. The blunt-end ligation method may avoid some of

this loss if, prior to the ligation step, a polymerase with strand displacement or nick translation activity is used in combination with enzymes that remove internal 3'-phosphates, which inhibit extension (Figure 1). Additionally, certain AT-rich double-stranded DNA molecules may be subject to denaturation during silica-based DNA purification due to chaotropic agents present in binding buffers (16), rendering them incapable of double-stranded ligation. The single-stranded method, however, allows each of these single-stranded fragments to be incorporated into the library. Another challenge, common to all methods of library construction, is that ancient DNA is often found in trace quantities, and the purification steps required during preparation of the ends for ligation can further reduce these amounts, diminishing the complexity of the final library.

In order to characterize the effect different library construction methods have on high-throughput sequencing results from ancient DNA, we tested the double-stranded blunt-end ligation method and the single-stranded method on a variety of ancient samples from various burial contexts containing DNA with diverse ages, average fragment sizes, and levels of background environmental DNA. We found the single-stranded method allows the incorporation of a higher proportion of endogenous DNA relative to environmental DNA in most, but not all, samples.

Materials and methods

DNA Extraction and purification

DNA was prepared from teeth and bones from bovines and woolly mammoths between 5200 and 42,000 years old recovered from various archeological sites from Turkey, Greece, Austria, France, and Belgium, a 1400 year-old human bone sample from northern France, a 100 year-old chimpanzee bone, and present-day human DNA from blood. Bovine, mammoth, and the 1400 year-old human samples were extracted and purified in a separate, contained, dedicated ancient DNA laboratory where protective clothing and decontamination of reagents, surfaces and equipment

was employed (17–18). After removing surfaces of the teeth and bones, the underlying areas were ground into powder either by low-speed drilling with a heat-sterilized drill bit using a Dremel Fortiflex (Dremel Europe, Breda, The Netherlands), or cut into fragments with a Dremel 4000 (Dremel Europe) equipped with a diamond saw blade and then powdered using a freezer mill (SPEX CertiPrep 6750, Metuchen, NJ). The resulting powder was incubated in 5–10 mL extraction buffer (0.5 M EDTA, 0.25 M $\text{Na}_2\text{HPO}_4^{3-}$, pH 8.0, 1% beta-mercaptoethanol) for 48–70 h at 37°C on a rotating wheel. Samples were then pelleted, and DNA was purified from the supernatant according to a protocol based on the QIAquick Gel Extraction kit protocol (Qiagen, Hilden, Germany), which included additional washing steps with 2 mL QG binding buffer and 2 mL PE wash buffer (Qiagen). Total volumes of 25–50 mL extract and binding buffer were passed through the silica columns on a vacuum manifold (Qiagen) with the aid of 20 mL tube extenders (Qiagen) (see comparisons in Supplementary Figure S1). This modification allows increased volumes of extract to be processed on a single silica column through the use of both a vacuum manifold and disposable column extenders. It can be applied to any large-volume extraction strategy, can accommodate multiple samples with minimal manipulation, and can be easily assembled in a laminar flow hood without contacting any surfaces exposed to extracts, minimizing the risk of introducing contamination. After the PE wash step, columns were then transferred from the manifold to a new 2 mL collection tube, and purification was continued according to the manufacturer's instructions using a bench-top centrifuge. DNA was eluted by 2 elution steps, each using 27 μL EB elution buffer (Qiagen) heated to 65°C.

Library preparation and sequencing

For both double-stranded and single-stranded libraries, all bovine samples were first treated with USER enzyme prior to library preparation (New England Biolabs, Ipswich, MA). The USER enzyme treatment was omitted for the other samples. Illumina and Ion Torrent double-stranded libraries were

prepared from 10–130 ng of DNA, water, or mock controls, then end-repair was performed using NEBNext End Repair Module (New England Biolabs) according to the manufacturer's instructions and purified using MinElute silica columns (Qiagen) according to the manufacturer's instructions but with 2 elutions of 17 μL each of EB buffer heated to 65°C. Blunt-end ligations were performed using the NEBNext Quick Ligation Module (New England Biolabs) according to the manufacturer's instructions. Since the single-stranded method uses a modified P5 Illumina adapter that requires a custom sequencing primer (7), double-stranded libraries were also prepared with this modified P5 Illumina adaptor in order to allow libraries generated using either single-stranded or double-stranded methods to be pooled and sequenced together (Supplementary Material). Following ligation, non-ligated adapters were removed through the addition of 1 μL exonuclease I (New England Biolabs), incubated for 10 min at 37°C, then brought to 50°C for 20 min. Then, elongation of the adapters (Figure 1) was performed by adding 1 volume of OneTaq DNA polymerase 2x Master Mix (New England Biolabs) and incubating for 10 min at 68°C. Reactions were then purified on a MinElute column (Qiagen) with 2 elutions of 17 μL each of EB buffer heated to 65°C.

Illumina and Ion Torrent single-stranded libraries were prepared from 10 and 25 ng of DNA, water, or mock controls according to Gansauge et al. (11) with the following modifications: (i) Single-strand ligations were performed with the addition of 1 μL instead of 4 μL of CircLigase II (Epicenter, Madison, WI) and incubated 3 h instead of 1 h, which reduced total reagent cost for this method by ~45%. (ii) The Bst 2.0 extension step using a 5' tailed primer was performed with incubation at a constant temperature of 15°C for 30 min instead of a gradual increase of 1°C per minute from 15°C to 37°C and 5 min at 37°C (similar results were also obtained after 30 min at room temperature).

To convert the single-stranded libraries for use with the Ion Torrent platform, single-stranded library products were amplified using custom

barcoded primers (Supplementary Figure S2). Amplified libraries from all methods were visualized and quantified using a High Sensitivity DNA Assay Chip kit on a Bioanalyzer 2100 (Agilent, Santa Clara, CA), pooled into equimolar amounts, and size-selected using a DNA 300 Chip on a LabChip XT fractionation system (Perkin Elmer, Waltham, MA), to minimize inclusion of adapter-dimers. Illumina libraries were pooled and paired-end sequenced (2×150) on a MiSeq platform, with sequencing primer CL72 replacing the first read sequencing primer (7). Pooled Ion Torrent libraries were sequenced on a 316 chip (500 run flows). Further methodological details can be found in the Supplementary Material.

Sequence analysis

After trimming adapter sequences and removing poor quality and duplicate reads, the remaining reads were mapped with BWA version 1.2.3 against elephant (*LoxAfr 3.0*), cow (*BosTau31*), chimpanzee (*panTro4*), or human (*hg19*), with the seeding deactivated, an edit distance of 0.04, and 1 gap opening allowed. These settings were compared with previously published and recommended settings for mapping ancient DNA reads with BWA and found to optimize the number of correctly

mapped reads uniformly for all size lengths (Supplementary Material), an important consideration for the present work. A minimum read size of 28 bp was determined by remapping (BWA with parameters as above) the subset of reads that mapped to their respective reference genomes to an additional reference sequence consisting of 2702 concatenated bacterial genomes with a total size of 9.4 gigabases (courtesy of Olivier Gorgé) and determining at which size bin the proportion of reads mapping to both reference sequences falls below 1% of total mapped reads. This minimum read size was found optimal for the mapping parameters of BWA we used, but a change of these parameters may lead to a change in the optimal cut-off length (Supplementary Material and Supplementary Figure S3).

Results and discussion

We first explored two library construction strategies using double-stranded DNA as a substrate that differ by the nature of the ligated adapter, either with a pair of dephosphorylated partially double-strand adapters or with a Y adapter (Figure 1). We found the Y-adapter method (TruSeq kit, Illumina Inc., San Diego, CA) to be poorly suited for low-quantity starting material due to

intrinsic adapter-dimer artifacts, which derive from trace amounts of adapter sequences with incorrect nucleotides (see Supplementary Material and Supplementary Figure S4). We therefore recommend against its use for creating libraries from poorly preserved samples.

We then compared libraries constructed with the method using blunt-end ligation of adaptors to double-stranded DNA with that using single-stranded DNA as a substrate. For all samples, libraries prepared from the same DNA extracts using both methods show a striking difference in the distribution of the lengths of the inserts they incorporate. Libraries produced using the single-stranded method contain a larger fraction of shorter molecules than those produced by the double-stranded method, and this fraction also includes a large proportion of molecules that are too short to be informatively mapped to a reference sequence (less than 28 bp with the mapping parameters used). A typical distribution of the different fragment lengths observed for each library construction method from the same sample is shown in Figure 2. For the samples tested in this study, the proportion of reads from single-stranded libraries, which could not

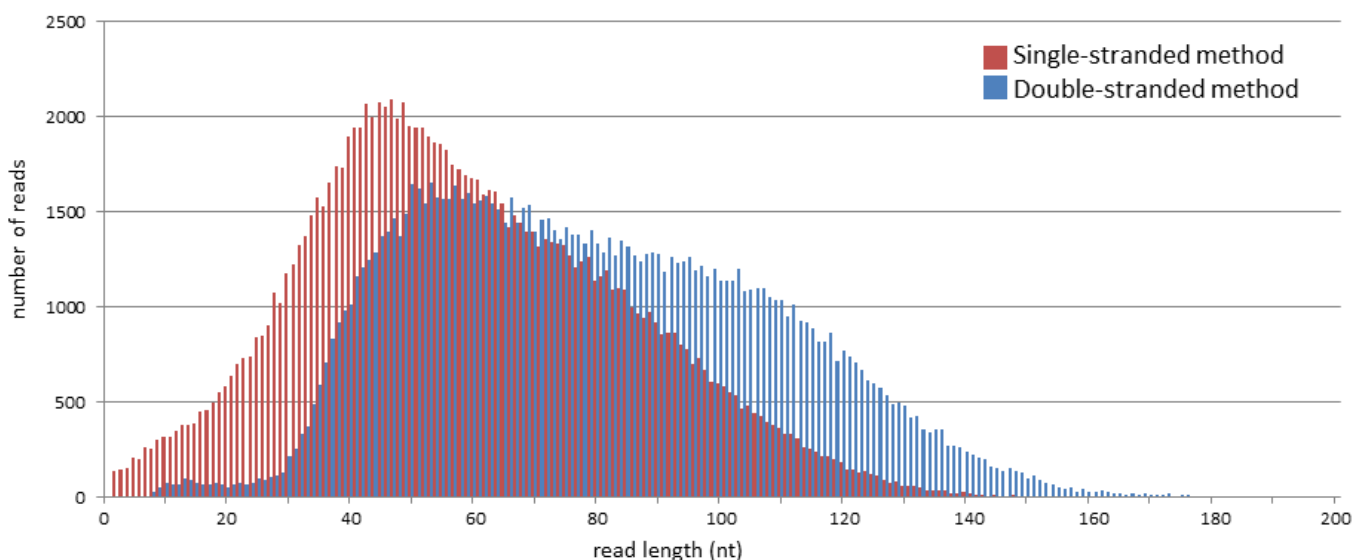


Figure 2. Typical size distributions of raw reads from single-stranded and double-stranded libraries. Overlapping histograms of the distribution of insert sizes for Ion Torrent libraries prepared from sample Mam2 with either single-stranded (red) or double-stranded (blue) libraries show typical characteristics of insert size incorporation observed for each method. Adapter sequence has been trimmed by the Ion Torrent Software Suite, which also removes inserts 4 bp or less for the double-stranded library. The 34 bp sequences flanking the insert for the single-strand procedure (see Supplementary Material) and PCR duplicates for both libraries have been removed. The total number of reads has been normalized between the two libraries.

Table 1. Sequencing results of double-stranded and single-stranded libraries

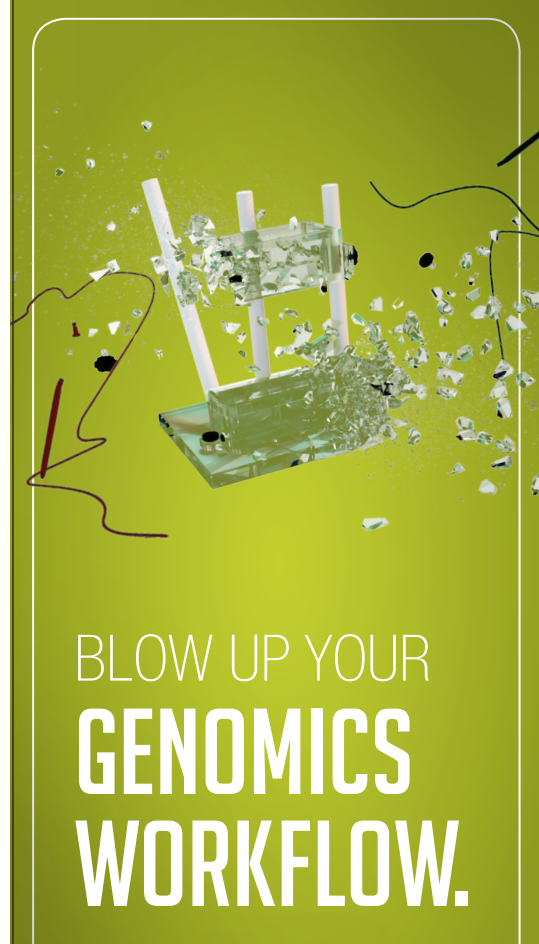
Sample ID	Source	Approximate age (years)	Origin	Library method	Platform	Total reads	Reads <28bp	Avg length of reads to be mapped (bp)	Endogenous mapped reads	Avg size mapped reads (bp)	Endogenous mapped bases	SS/DS (fold difference)
Bos5	Cattle	8,700	SWA	DS	Illumina	554142	0.2%	103	8.99%	99	8.70%	
Bos5				SS	Illumina	626066	17.4%	63	38.58%	56	34.59%	4.0
Bos16	Cattle	8,700	SWA	DS	Illumina	643585	0.3%	97	0.27%	94	0.26%	
Bos16				SS	Illumina	488403	12.9%	65	1.79%	55	1.50%	5.8
Bos17	Cattle	8,800	SWA	DS	Illumina	524156	0.0%	99	0.03%	82	0.03%	
Bos17				SS	Illumina	617035	10.4%	60	0.72%	50	0.59%	22.6
Bos21	Cattle	8,700	SWA	DS	Illumina	404785	0.0%	105	0.02%	86	0.01%	
Bos21				SS	Illumina	561434	10.1%	68	0.11%	56	0.09%	7.0
Bos12	Cattle	5,000	E	DS	Illumina	533535	0.1%	113	0.20%	69	0.12%	
Bos12				SS	Illumina	640447	17.2%	75	3.19%	56	2.41%	20.4
Bos15	Cattle	5,000	E	DS	Illumina	503348	0.0%	106	0.04%	76	0.03%	
Bos15				SS	Illumina	468114	9.1%	72	0.40%	51	0.29%	9.7
Bos16k	Cattle	5,800	SWA	DS	Illumina	799885	0.1%	94	0.02%	73	0.02%	
Bos16k				SS	Illumina	503613	26.9%	67	0.61%	54	0.49%	25.9
Bos4v	Cattle	7,300	SE	DS	Illumina	479543	0.1%	112	0.01%	91	0.00%	
Bos4v				SS	Illumina	491533	11.3%	67	0.04%	50	0.03%	5.6
Bos4g	Cattle	6,500	E	DS	Illumina	422594	0.3%	100	0.05%	92	0.05%	
Bos4g				SS	Illumina	583403	24.9%	60	0.34%	55	0.32%	6.6
Bos9	Cattle	6,500	E	DS	Illumina	448530	0.1%	102	0.01%	76	0.01%	
Bos9				SS	Illumina	670732	23.6%	58	0.08%	51	0.07%	8.5
Mam1	Mammoth	21,000	E	DS	Ion Torrent	202289	1.7%	98	0.24%	80	0.20%	
Mam1				SS	Ion Torrent	294931	10.2%	84	0.06%	56	0.05%	0.2
Mam2	Mammoth	21,000	E	DS	Ion Torrent	123575	1.6%	89	0.00%	0	0.00%	
Mam2				SS	Ion Torrent	267057	6.4%	85	0.02%	55	0.01%	-
Mam4	Mammoth	42,000	E	DS	Ion Torrent	89326	1.5%	81	0.40%	73	0.36%	
Mam4				SS	Ion Torrent	194510	2.3%	78	3.98%	60	3.05%	8.4
Hom81	Human	1,400	E	DS	Illumina	1520879	0.1%	131	0.17%	119	0.15%	
Hom81				SS	Illumina	717402	5.5%	88	0.15%	83	0.14%	0.9
Pan1	Chimpanzee	100	M	DS	Illumina	456127	0.0%	94	88.41%	92	86.64%	
Pan1				SS	Illumina	380837	9.1%	67	87.14%	64	83.83%	1.0
Hom1	Human	modern	E	DS	Illumina	251596	0.0%	103	89.74%	100	87.82%	
Hom1				SS	Illumina	569510	13.6%	69	90.45%	66	87.00%	1.0

DS = double stranded, SS = single stranded library preparation methods. "Reads < 28 bp" is the percent of insert sizes less than 28 bp after removing low quality and duplicate reads. "Avg length of reads to be mapped" is the average read length after removal of low quality, duplicates, and reads <28bp. "Endogenous mapped reads" is the percent of reads mapping to either the cow (Bos) or elephant (Mam), chimp (Pan) or human (Hom) reference genome. "Endogenous mapped bases" is the percentage of unique mapped bases from the total unique bases sequenced after removal of low quality, duplicates, and reads <28bp. "SS/DS" is the relative improvement of the percent of mapped bases between libraries. E, Europe; SE, Southern Europe; SWA, South West Asia; M, museum sample.

be reliably mapped due to insufficient length, comprised 5%–27% of the total reads after quality filters and duplicate removal, whereas the corresponding size category in double-stranded libraries prepared from the same samples was consistently less than 2% (Table 1). Furthermore, it was found that after removal of these shorter reads, the average insert length of the remaining reads was reduced in all of our single-stranded libraries by 3%–43% compared with libraries prepared with the double-stranded method from the same sample, with an average reduction of 32% (Table 1). The use of UDG and endonuclease VIII, a damage treatment step often used with ancient DNA to remove deaminated cytosines, may create short single-stranded DNA by-products, which could conceivably be ligated with the single-stranded ligase and inflate the amount of short fragments observed. Libraries made from samples with and without UDG/endonuclease VIII treatment, however, showed no greater proportion of these short reads (Supplementary Figure S5). Although for this study, libraries were size selected only to remove adapter-dimers in order to facilitate direct comparisons between methods, a stricter size-selection may be used for single-stranded libraries to reduce the quantity of shorter inserts that cannot be reliably mapped. Care must be taken with this approach however, as many ancient DNA libraries contain a large proportion of informative fragments near this lower size limit, a fraction of which may be also removed.

After the removal of reads shorter than 28 bp, the remaining reads were mapped to the corresponding reference genomes in order to determine the proportion of endogenous ancient DNA to environmental or exogenous DNA in each sample. Despite both the overall reduction in the number of reads of sufficient length for mapping and the reduction in average length of those reads, the incorporation of shorter DNA molecules with the single-stranded method appears to be an advantage for 12 out of 14 of the ancient samples analyzed. These samples show a higher proportion of mapped reads to total reads with the single-stranded

than with the double-stranded libraries from the same extracts, normalized by percentiles. Additionally, the average lengths of these mapped reads were found to be 18%–45% shorter than the average lengths of the reads mapped from double-stranded libraries, with an average reduction of 33% (Table 1). Since the additional reads were generally of a shorter length compared with those mapped with the double-stranded library, we calculated the overall difference in the mapped bases between the two methods, normalized as a percentage of the total bases available for mapping for each library. These results show increases in the proportion of mapped bases between 4- and 26-fold, with an average increase of 11-fold for those samples that gave a higher proportion of mapped bases with the single-stranded library, while the two ancient samples that yielded proportionally fewer mapped bases with the single-stranded method show decreases of 0.2- and 0.9-fold (Table 1). It should be noted that this value is dependent on the separate efficiencies of two individual library preparations as well as sample-dependent DNA characteristics that may favor one method over the other. Since the proportion of mapped reads to total reads is frequently used to calculate the percentage of endogenous DNA of a given ancient sample, this value may also fluctuate for the same extract depending on the library construction method used. Typical differences in the distribution of the total and endogenous fragment lengths recovered with each library construction method from the same extract are presented for three ancient samples (Figure 3A–C) and two recent ones (Supplementary Figure S6). The proportion of mapped reads for each 10 bp bin was compared for both methods (Figure 3B). This comparison reveals that the single-stranded method allows recovery of a higher proportion of mapped reads at almost every bin size. This is observed for both shorter and longer reads, although this improvement generally diminishes for increasing fragment lengths. It should be noted that there are sample-specific differences in the extent of the improvements for these



BLOW UP YOUR GENOMICS WORKFLOW.

Automate nucleic acid
QC and get on with your
life sciences.

Wipe out handling errors and streamline lab operations with Fragment Analyzer. Just pipette once and it automates capillary electrophoresis.

**No chips. No tapes.
No compromises.**

- Setup in seconds
- Get resolution down to 2 base pairs
- Detection starts at 5 pg/μL

DISPENSE WITH YOUR TIRED
OLD WORKFLOW AT
AATI-US.COM.



For inquiries, call
+1 515-296-6600

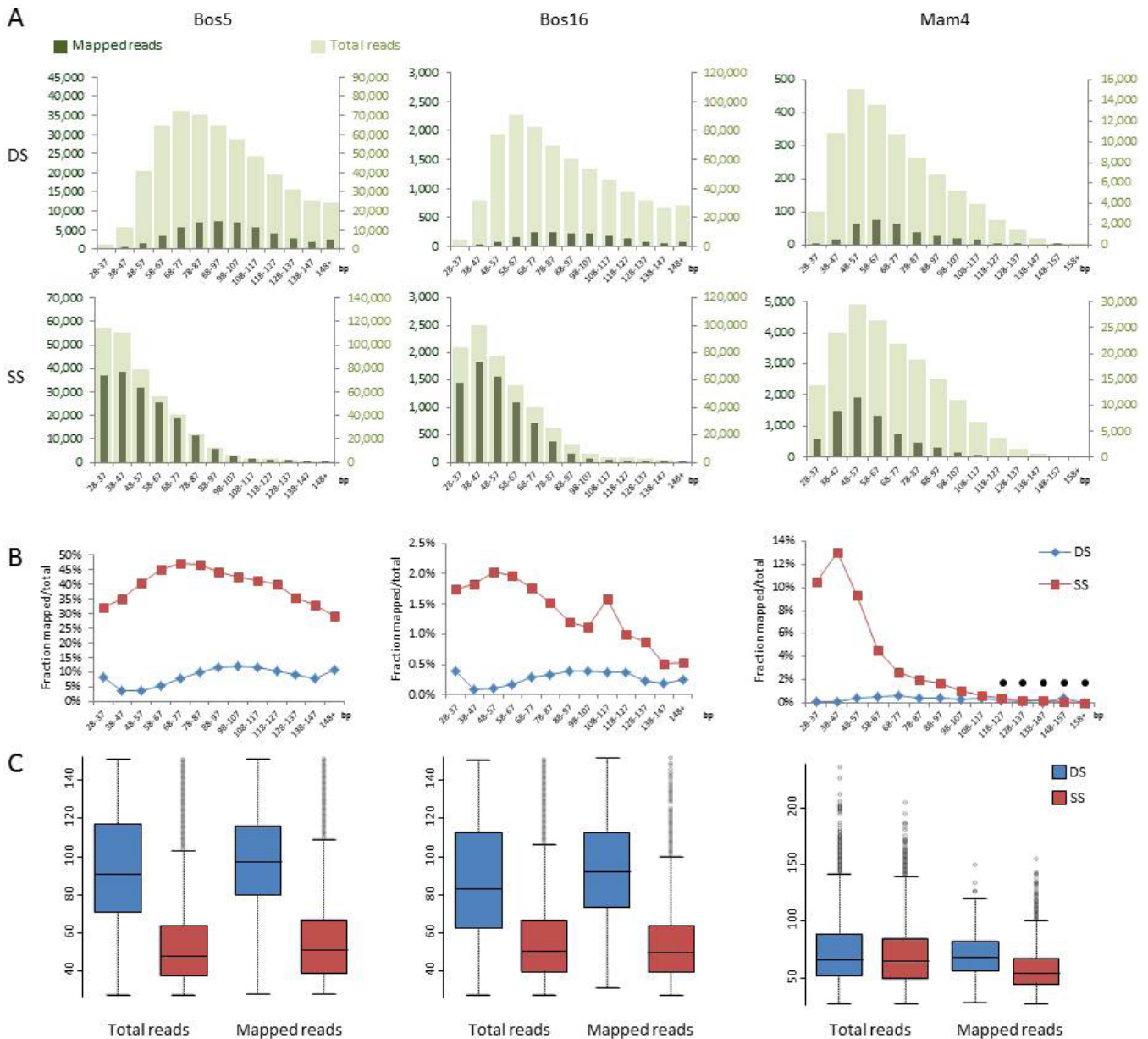


Figure 3. Insert sizes of total and mapped reads from single-stranded and double-stranded libraries. (A) Insert size distribution results from 3 ancient samples sequenced on different platforms (Bos5, Bos16: 8000 year-old cattle sequenced on Illumina MiSeq, Mam4: 42,000 year-old mammoth sequenced on Ion Torrent PGM). Top row, double-stranded libraries (DS), bottom row, single-stranded libraries (SS). Horizontal axes show distribution of insert sizes in 10 bp bins (after removing reads shorter than 28 bp). Left vertical axes show numbers of non-redundant endogenous (mapped) reads for each 10 bp bin are given for the 2 different library preparation methods for each sample. The bins for which the statistical test did not reveal a significant difference in the proportions between the methods are indicated by a black dot (Supplementary Material and Supplementary Table S1). (C) Box plots of size distributions of the total numbers of informative reads and mapped endogenous reads from both single-stranded (SS) and double-stranded (DS) sequencing runs. The whiskers indicate the shortest sequence still within the 1.5 interquartile range (IQR) of the lower quartile, and the longest sequence still within the 1.5 IQR of the upper quartile, whereas the outliers are individually represented by a circle.

longer reads. For the oldest sample analyzed here (a 42,000 year-old mammoth from a temperate site), the extent of DNA degradation is more pronounced, and neither the single-stranded nor the double-stranded methods show significant improvement in the recovery of larger fragments,

which may be due to the fact that very few larger fragments are preserved in this sample (Figure 3B).

For the more recent samples in this study, DNA purified from a 100 year-old chimpanzee bone and from a blood sample from a present-day human, the latter being fragmented by

sonication prior to library construction, we find no change in the proportion of mapped reads when using the single-stranded method over the double-stranded method (Supplementary Figure S6). This is due to the fact that these samples contain the same low levels of exogenous DNA that could be

Sequencing-Ready DNA Libraries from 1 ng – 1 µg of DNA in 2 Hours or Less

The **NEXTflex™ Rapid DNA-Seq Kit** offers sequencing-ready, Illumina-compatible libraries from 1 ng – 1 µg of DNA in two hours or less. Proven Enhanced Adapter Ligation Technology provides the largest number of unique sequencing reads available. With low input requirements, a streamlined library prep protocol and a cost-effective price point, the NEXTflex Rapid DNA-Seq Kit will help you meet your research goals faster than you ever thought possible.

Fast - Libraries in 2 hours or less

Multiplex - Up to 192 samples

Flexible - Input DNA from 1 ng to 1 µg

Compatible - Ideal for clinical & FFPE samples

Available Now - Next day delivery

Visit BioNGS.com to accelerate your library prep.

Protocols now available for Beckman Biomek® FX and Biomek® FX^P Laboratory Automation Workstations

Amplicon-Seq • DNA-Seq • Pre-Target Capture
 Bisulfite-Seq • Methyl-Seq • RNA-Seq
 Rapid RNA-Seq • Directional RNA-Seq
 Small RNA-Seq • Rapid Directional RNA-Seq
 PCR-Free DNA-Seq • ChIP-Seq • qRNA-Seq
Rapid DNA-Seq • Multiple Platform Compatibility

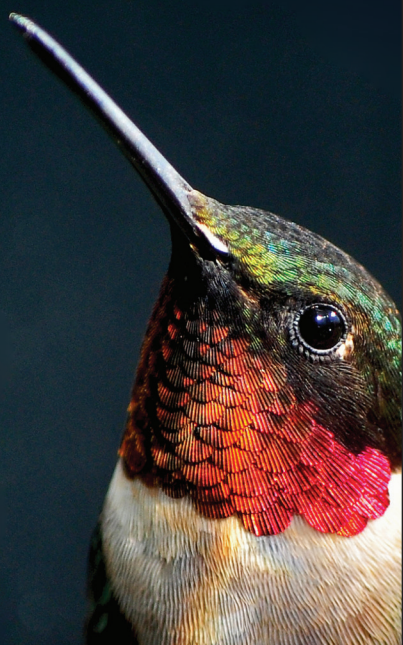
incorporated into libraries with either method.

It is noteworthy that two ancient samples did not show an increase in the proportion of mapped bases when prepared with the single-stranded as opposed to the double-stranded method. This result may be best explained by the distribution of the fragment lengths of exogenous or environmental DNA with relation to the endogenous DNA in the sample. In the case of sample Hom81, a 1400 year-old human sample from Northeastern France, qPCR analysis confirms extremely well-preserved human DNA. Indeed, identical human mitochondrial sequences above 300 bp could regularly be amplified as single products from different samples of this individual. The difference in the average lengths of pre-mapped inserts recovered from this sample between the single- and double-stranded methods (88 bp and 131 bp respectively, Table 1) attests to the presence of larger fragments in the extract. Similar to the modern samples, the results from both libraries reveal roughly an equal proportion of endogenous/exogenous DNA content for the different size fractions incorporated by the two library construction methods. In contrast, for sample Mam1, a 21,000 year-old mammoth from a temperate climate (Austria), the 5-fold increase of the percentage of mapped reads found with the double-stranded method among larger inserts (averaging 98 bp) as opposed to those found among shorter inserts obtained with the single-stranded method (averaging 84 bp), may indicate the presence of a large fraction of degraded exogenous DNA content in this sample that reduces the relative presence of endogenous DNA of smaller sizes. As seen with these examples, the lack of predictable concordance between fragment length and age of endogenous molecules (19), coupled with the presence of exogenous DNA in equally unknown fragment lengths makes it difficult to predict which ancient samples will be less amenable to the improvements expected by using single-stranded library construction.

It remained unclear to what extent the differences in fragment length

distributions observed between these two library construction methods was due to an increased availability of molecules in the shorter size ranges or a reduced efficiency of the single-stranded ligase to ligate longer molecules. It had been previously reported that due to the inefficiency of the single-strand DNA ligase to ligate single-stranded DNA exceeding 120 bp (20), this method may not be suitable for preparing libraries from ancient samples known to contain endogenous molecules larger than this limit (11). We explored this proposed limitation by performing a series of single-stranded ligation experiments using synthetic oligonucleotides of varying lengths in both competitive and non-competitive assays (Supplementary Material). Surprisingly, these experiments show no loss of ligation efficiency correlated to size for oligonucleotides between 90 and 358 bp (Supplementary Figure S7). These results indicate that the increase in the proportion of shorter molecules we observe with the libraries prepared from the single-stranded method is likely to be a consequence of a greater proportion of shorter DNA molecules present in the purified extracts, which are intractable to being incorporated into libraries prepared with the double-stranded method.

Finally, we examined the impact of platform selection for sequencing diverse ancient DNA libraries. We compare Illumina MiSeq with Ion Torrent PGM, which we make possible through adaptations to the single-stranded method. Both of these sequencing platforms use a sequence-by-synthesis approach, but they differ in throughput capabilities, sample preparation protocols, detection technology, quality-assignment software, and accuracy (see Reference 21 for in-depth platform comparisons). We found no notable difference in insert size when analyzing the same sample with these different platforms (Supplementary Figure S8). We do observe, however, a clear difference between the distributions of quality scores over the read lengths, which requires a slight modification of quality trimming parameters in order to maximize the number of mapped reads



for each platform (see Supplementary Material).

In conclusion, we present a systematic comparison of traditional high-throughput sequencing library building techniques with a novel method developed specifically to address the challenges encountered when preparing libraries from ancient or degraded material, and show the results of these methods when applied to a broad collection of ancient and recent samples containing DNA with various degrees of preservation. We confirm the utility of the single-stranded library preparation method in recovering short endogenous reads that would otherwise be lost using the traditional double-stranded method and that the single-stranded method also allows incorporation of a higher ratio of endogenous to environmental DNA for the majority of our ancient samples, even when larger reads are considered. While we demonstrate that the single-stranded ligase is equally efficient with longer DNA fragments, we caution against the exclusive adoption of this method on uncharacterized samples, since the improvements appear to depend upon the relative size distributions of the endogenous and environmental DNA in a given sample.

Author contributions

Conceived and designed the experiments: T.G., E.M.G., and E.A.B. Performed the experiments: E.A.B., D.M., and G.L. Analyzed the data: E.A.B., D.M., G.L., and T.G. Performed sequencing and facilitated data analysis: JD. Wrote the paper: E.A.B., T.G., and E.M.G.

Acknowledgments

We are grateful to Rose-Marie Arbogast, Ben Arbuckle, H el ene Barrand Emam, Ursula G ohlich, Lamys Hachem, Nadja Hoke, Joris Peters, Patrick Semal, and Frank Zachos for providing the samples for the study, Olivier Gorge for technical assistance, and Ga elle Lelanda s for helpful advice on the statistical analyses. We also would like to thank the four anonymous reviewers for their helpful suggestions and comments. EAB and GL were supported by the

Labex « Who am I ? ». The sequencing facility is supported by grants from the University Paris Diderot, the Fondation pour la Recherche M edicale, and the R egion Ile-de-France.

Competing interests

The authors declare no competing interests.

References

1. Willerslev, E., J. Davison, M. Moora, M. Zobel, E. Coissac, M.E. Edwards, E.D. Lorenzen, M. Vestergaard, et al. 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* 506:47-51.
2. Orlando, L., A. Ginolhac, G. Zhang, D. Froese, A. Albrechtsen, M. Stiller, M. Schubert, E. Cappellini, et al. 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499:74-78.
3. Bos, K.I., V.J. Schuenemann, G.B. Golding, H.A. Burbano, N. Waglechner, B.K. Coombes, J.B. McPhee, S.N. DeWitte, et al. 2011. A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* 478:506-510.
4. Margulies, M., M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bembien, J. Berka, M.S. Braverman, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376-380.
5. Blow, M.J., T. Zhang, T. Woyke, C.F. Speller, A. Krivoschapkin, D.Y. Yang, A. Derevianko, and E.M. Rubin. 2008. Identification of ancient remains through genomic sequencing. *Genome Res.* 18:1347-1353.
6. Bentley, D.R., S. Balasubramanian, H.P. Swerdlow, G.P. Smith, J. Milton, C.G. Brown, K.P. Hall, D.J. Evers, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53-59.
7. Meyer, M., M. Kircher, M.T. Gansauge, H. Li, F. Racimo, S. Mallick, J.G. Schraiber, F. Jay, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338:222-226.
8. Pr ufer, K., F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505:43-49.
9. Dabney, J., M. Knapp, I. Glocke, M.T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. Garcia, et al. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA* 110:15758-15763.
10. Meyer, M., Q. Fu, A. Aximu-Petri, I. Glocke, B. Nickel, J.L. Arsuaga, I. Martinez, A. Gracia, et al. 2014. A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* 505:403-406.

11. Gansauge, M.T. and M. Meyer. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* 8:737-748.
12. Lindahl, T. 1993. Instability and decay of the primary structure of DNA. *Nature* 362:709-715.
13. Hofreiter, M., V. Jaenicke, D. Serre, A. von Haeseler, and S. Paabo. 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* 29:4793-4799.
14. Briggs, A.W., U. Stenzel, P.L. Johnson, R.E. Green, J. Kelso, K. Prufer, M. Meyer, J. Krause, et al. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. USA* 104:14616-14621.
15. Heyn, P., U. Stenzel, A.W. Briggs, M. Kircher, M. Hofreiter, and M. Meyer. 2010. Road blocks on paleogenomes--polymerase extension profiling reveals the frequency of blocking lesions in ancient DNA. *Nucleic Acids Res.* 38:e161.
16. Prevorovsk y, M. and F. Puta. 2003. A/T-rich inverted DNA repeats are destabilized by chaotrope-containing buffer during purification using silica gel membrane technology. *Biotechniques* 35:698-702.
17. Charruau, P., C. Fernandes, P. Orozco-Terwengel, J. Peters, L. Hunter, H. Ziaie, A. Jourabchian, H. Jowkar, et al. 2011. Phylogeography, genetic structure and population divergence time of cheetahs in Africa and Asia: evidence for long-term geographic isolates. *Mol. Ecol.* 20:706-724.
18. Champlot, S., C. Berthelot, M. Pruvost, E.A. Bennett, T. Grange, and E.M. Geigl. 2010. An efficient multistage DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS ONE* 5:e13042.
19. Sawyer, S., J. Krause, K. Guschanski, V. Savolainen, and S. Paabo. 2012. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE* 7:e34131.
20. Li, T.W. and K.M. Weeks. 2006. Structure-independent and quantitative ligation of single-stranded DNA. *Anal. Biochem.* 349:242-246.
21. Loman, N.J., R.V. Misra, T.J. Dallman, C. Constantinidou, S.E. Gharbia, J. Wain, and M.J. Pallen. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* 30:434-439.

Received 02 December 2013; accepted 27 May 2014.

Address correspondence to E. Andrew Bennett, Eva-Maria Geigl, or Thierry Grange, Institut Jacques Monod, Universit e Paris Diderot, Paris, France. E-mail: bennett@ijm.univ-paris-diderot.fr, geigl@ijm.univ-paris-diderot.fr, grange@ijm.univ-paris-diderot.fr

To purchase reprints of this article, contact: biotechniques@fosterprinting.com