The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils
Morten E. Allentoft1,2,3,*, Matthew Collins4, David Harker4, James Haile1, Charlotte L. Oskam1, Marie L. Hale2, Paula F. Campos3,5, Jose A. Samaniego3, M. Thomas P. Gilbert1,3, Eske Willerslev3, Guojie Zhang6, R. Paul Scofield7, Richard N. Holdaway2,8 and Michael Bunce1, *
1Ancient DNA Laboratory, School of Biological Sciences and Biotechnology, Murdoch University, 90 South Street, Perth, Western Australia 6150, Australia
2School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch, New Zealand
3Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen K, Denmark
4Department of Archaeology, University of York, PO Box 373, York, UK
5Museu da Ciência, University of Coimbra, Laboratorio Chimico, Largo Marquês de Pombal, 3000-272 Coimbra, Portugal
6Beijing Genomics Institute-Shenzhen, Shenzhen 518083, People’s Republic of China
7Canterbury Museum, Rolleston Avenue, Christchurch 8050, New Zealand
8Palaeol Research Ltd, 167 Springs Road, Hornby, Christchurch 8042, New Zealand

Claims of extreme survival of DNA have emphasized the need for reliable models of DNA degradation through time. By analysing mitochondrial DNA (mtDNA) from 158 radiocarbon-dated bones of the extinct New Zealand moa, we confirm empirically a long-hypothesized exponential decay relationship. The average DNA half-life within this geographically constrained fossil assemblage was estimated to be 521 years for a 242 bp mtDNA sequence, corresponding to a per nucleotide fragmentation rate \( k \) of 5.50 \( \times 10^{-6} \) per year. With an effective burial temperature of 13.1°C, the rate is almost 400 times slower than predicted from published kinetic data of \textit{in vitro} DNA depurination at pH 5. Although best described by an exponential model \( R^2 = 0.39 \), considerable sample-to-sample variance in DNA preservation could not be accounted for by geologic age. This variation likely derives from differences in taphonomy and bone diagenesis, which have confounded previous, less spatially constrained attempts to study DNA decay kinetics. Lastly, by calculating DNA fragmentation rates on Illumina HiSeq data, we show that nuclear DNA has degraded at least twice as fast as mtDNA. These results provide a baseline for predicting long-term DNA survival in bone.

\textbf{Keywords:} DNA degradation; aDNA; decay kinetics; DNA half-life

\section{1. INTRODUCTION}

Although early-1990s claims of DNA recovered from million year-old fossils [1–4] are now widely regarded as modern contaminants [5–13], the kinetics of long-term post-mortem DNA decay is still poorly understood. There is a lack of empirical data on which to estimate the rate of \textit{in situ} DNA fragmentation. Because the field of ancient DNA (aDNA) has recently entered the era of whole-genome profiling [14–17], which is dependent on samples of exceptional preservation, understanding the nature and rate of DNA decay is as pertinent as ever, for both predictive and authentication purposes.

DNA has limited chemical stability and decays without the enzymatic repair mechanisms of living cells [18]. Following cell death, nucleases start to cleave the DNA into fragments [19] and during decomposition, the DNA is digested by micro-organisms [18,20]. In determining long-term DNA decay, it is believed that hydrolysis of amino groups accelerates the loss of purine residues (depurination), resulting in strand cleavage [21,22]. This random DNA fragmentation generates a characteristic negative exponential correlation between DNA fragment length and number of molecules (figure 1\(a)\) [23,24,26,27].

Although other types of DNA damage occur, such as cross-links [28,29], the role of depurination in DNA fragmentation is well documented, and damage patterns from high-throughput sequencing data show that purines are overrepresented adjacent to strand breaks in ancient samples [30–33]. However, it is not known whether the rate of fragmentation can be regarded as constant through time, nor to what extent it varies between specimens from similar depositional environments. Indeed, a constant rate would imply that DNA decay follows first-order kinetics.
cannot be predicted \[42–44\]. In the light of many inconclusive sample, to overcome the ‘noise’ introduced by the afore-
it takes a large, homogenous, and accurately dated specimens, leading to potential problems of mixed taphonomies. The absence of clear temporal trends in such as taphonomy, fossil storage, oxygenation, microbial diagenesis, pH and ionic strength, and the presence of cations, humics and humates. Moreover, most studies are limited by small sample sizes and lack of individually dated specimens, leading to potential problems of mixed taphonomies. The absence of clear temporal trends in previous work suggests that the rate of DNA decay either fluctuates widely (hence not a rate per se), or that it takes a large, homogenous, and accurately dated sample, to overcome the ‘noise’ introduced by the aforementioned factors. In the light of many inconclusive results, a common refrain is that DNA decay rates cannot be predicted \[42–44\].

A few studies, however, have indicated a relationship between DNA preservation and post-mortem age. Bar et al. \[45\] showed that the length of amplifiable DNA fragments from human tissues (blood, muscle and spleen) decreased exponentially during the 20 days after a person’s death, and Campos et al. \[46\] showed a rapid decrease in DNA content in cow bone after burial. Moreover, Adler et al. \[26\] showed a weak exponential correlation between DNA preservation and age, on the basis of 37 ancient human samples (mainly teeth). These samples had not all been individually dated, but the result is encouraging and provides impetus for more detailed investigations of long-term DNA fragmentation.

Establishing an association between age and preservation is not the only challenge. The rate of depurination is influenced by temperature, among other factors \[22\], which explains why the most extreme survival of DNA was documented in approximately 450–800 kyr ice cores \[47\]. Smith et al. \[11,12\] argued for the possibility of using the temperature-dependence of DNA fragmentation to normalize the samples and predict DNA survival. Such a relationship has been demonstrated for collagen, the most abundant protein in bone \[48\]. However, to establish a thermal age model for DNA, the first step is to confirm that long-term in situ DNA degradation can be described by a rate kinetic.

In an attempt to document a correlation between sample age and DNA preservation, we use a quantitative real-time PCR (qPCR) design to measure relative copy numbers of mitochondrial DNA (mtDNA) fragments from bones of the extinct New Zealand moa (Aves: Dinornithiformes). Aside from a large dataset (qPCR data from 158 radiocarbon-dated bones), our study differs from previous attempts by being highly localized. Using Holocene fossil tibiotarsi recovered from anoxic, limestone-buffered sediments at three adjacent (less than 5 km apart) fossil deposits (figure 2), we endeavour to provide a high level of sample homogeneity and minimize variables of climate and taphonomy. Also, by sampling only natural bone accumulations, we expect to minimize other complicating factors (e.g. butchering and cooking) that could impact on taphonomic processes \[27,49\].

Because the three fossil sites were excavated at different times (figure 2), the influence of storage times on DNA recovery can also be assessed. Pruvost et al. \[50\] argued that DNA degradation intensifies when a bone is removed from its deposition environment. The moa fossils analysed here were recovered at well-documented dates over an approximately 70 year period, allowing us to test the importance of post-excavation storage time.

The three main objectives of this study were: (i) to test whether long-term DNA decay follows first-order kinetics, thereby confirming the foundation for a predictive model; (ii) to estimate the long-term decay rate in bone at a given burial temperature and compare this rate with the predicted depurination rate from DNA in solution \[21,22\]; and (iii) to estimate the relative importance of storage time on DNA preservation in bone.

Lastly, following Deagle et al.’s \[23\] demonstration that frequency distributions of DNA fragment lengths from a sample of known age can be used to estimate the DNA decay rate, we argue that histograms based on the thousands-to-millions of reads generated from high-throughput sequencing platforms could represent a ‘by-product’ to investigate DNA fragmentation. We therefore apply the method to Illumina HiSeq 2000 data

\[f(T) = f(T_0)e^{-k/T}\]

\[f(L) = f(L_0)e^{-L/50} \]

\[\lambda = 0.02\] per site

\[k = 0.02\] per site per year

\[T = (100 - \text{copy number (relative)} \times 100) / 10^5\]

\[L = \text{fragment length (bp)}\]

\[f(T) = f(T_0)e^{-k/T}\]

\[f(L) = f(L_0)e^{-L/50}\]

\[\lambda = 0.02\] per site

\[k = 0.02\] per site per year

\[T = (100 - \text{copy number (relative)} \times 100) / 10^5\]

\[L = \text{fragment length (bp)}\]
generated from two moa specimens, in order to achieve independent DNA decay-rate estimates, comparable with the qPCR results.

2. MATERIAL AND METHODS

(a) Sampling, dating and storage
Sampling of 158 moa left tibiotarsi was conducted as described previously [51]. The material covered three adjacent sites, located within a 5 km radius in North Canterbury, South Island, New Zealand; Pyramid Valley (PV; \( n = 103 \)), Bell Hill Vineyard (BHV; \( n = 47 \)) and Rosslea (\( n = 8 \)) (figure 2). Further information on samples and sites has been given previously [52, 53] and is also listed in electronic supplementary material, table S1. All DNA extractions were undertaken from 200 mg uniformly sampled bone. The DNA isolation protocol (molecular extraction and qPCR) was performed using SYBR-green detection chemistry (without a spiked-in control) and stability kinetic analysis was conducted on the entire dataset (\( n = 158 \)) and exclusively on 103 elements from PV that were excavated over a 35-year time frame, with excavation dates recorded for each bone.

Figure 2. Study site. The three fossil deposits, PV (42°58′22.0″ S, 172°35′49.0″ E), BHV (42°58′19.36″ S, 172°39′56.15″ E) and Rosslea (42°57′53.83″ S, 172°39′22.39″ E), from which 158 radiocarbon-dated moa fossils were characterized for DNA decay kinetics. The sites in North Canterbury, South Island, New Zealand are located within a 5 km radius in the eastern rain shadows of the Southern Alps. Most of the area is more than 200 m a.s.l. and consists of flat alluvial plains and rolling downlands. Information on calibrated radiocarbon ages and DNA preservation (\( C_T \) values) are shown for each site. \( m \), mean age.

(b) DNA extraction and qPCR
DNA was isolated from all samples at dedicated aDNA facilities (Murdoch University, Perth, Australia), using a standardized method described previously [52] and detailed in the electronic supplementary material, text S1. All DNA extractions were undertaken from 200 mg uniformly sampled and powdered bone. The DNA isolation protocol (molecular weight cut-off (MWCO) and Silica purification) was selected owing to its ability to generate DNA largely free of inhibitors. DNA amplifications of the 262F/441R fragment (mtDNA control region, 242 bp including primers—see Bunce et al. [55]) were performed using SYBR-green detection chemistry on a BioRad My-Q thermocycler. With a standardized level of detection, the recorded cycle threshold (\( C_T \) values) represented the relative DNA preservation in each extract. To minimize effects of post-extraction storage time, qPCR was conducted immediately following DNA extraction. The quantitative response and general applicability of our qPCR set-up has been validated before based partly on the same extracts [51, 56–58]. In addition, we conducted melt curve analyses for all PCR products, (see the electronic supplementary material, figure S1 and Table S1), as well as ‘spike-in’ experiments (see the electronic supplementary material, figure S3 and Table S3) and PCR efficiency (see the electronic supplementary material, figure S3) experiments. This was done to ensure that amplification of the correct target was being
monitored and to test for PCR inhibition (see the electronic supplementary material, text S1). In most instances (greater than 70%), the actual qPCR products were directly sequenced, confirming their species origin (data not shown) and specificity of the amplicons. Extracts displaying deviating melt curve diagnostics (owing to non-specific amplifications and/or dimer accumulation) were omitted from the analysis (ca 20 samples). We acknowledge that the relatively large (242 bp) amplicons, with primers that had to work on multiple species and across a large dynamic range of DNA yields, is not ideal in a qPCR set-up. All of our tests indicated, however, that the assay is sufficiently robust and reproducible for the intended purpose.

Stochastic variation between qPCR reactions is expected, in particular when amplifying low copy number DNA at Ct values greater than 35. Results from 50 extracts in repeated qPCRs (same extract, but independent PCR mastermix and thermocycling run) were used to isolate this effect and estimate the variation introduced by stochastic qPCR effects. To test for consistency across DNA extracts, one moa bone was sampled and extracted three times, showing only minor deviations in Ct (less than 1 cycle) between extracts at multiple levels of dilution (see the electronic supplementary material, table S4).

Preliminary data suggested that DNA from one of the four North Canterbury moa species (Pachyornis elephantopus) amplified less efficiently in the qPCR assay. We suspect that primer-binding-site mutations are the cause, and to prevent a systematic bias, data from this species were excluded.

(c) Estimating the rate of decay from qPCR data

The number of DNA copies in a PCR reaction follows an exponential increase per cycle. If one DNA extract produces a Ct three cycles higher than another, then the former has eight times (2³) less DNA (in an ideal situation with 100% amplification efficiency). To test for an association between DNA preservation and age, we calculated the relative number of template molecules in each sample according to the ‘best’ sample (Dinornis robustus, S39955, Ct = 24.8), which was set to 100 per cent. Although some previous studies convert Ct values into absolute starting copy numbers [24,59], others do not [46] as the accuracy and meaningfulness can sometimes be questioned [58]. For our purpose, the use of relative Ct values was sufficient, as we were interested only in rates and not in absolute numbers.

Following Adler et al. [26], we tested (in SPSS v. 19) whether the observed DNA degradation through time would be better described by a linear, inverse, S-shaped or exponential decay model. An exponential decay relation could be better described by a linear, inverse, S-shaped or exponential decay relation. Whether the observed DNA degradation through time would be better described by a linear, inverse, S-shaped or exponential decay model. An exponential decay relation could be estimated as: \( N_t = N_0 \times e^{-k_t/T} \) (\( N_t \) and \( N_0 \) being the quantity at time \( t \) and 0, respectively, and \( k_{242} \) the decay constant for the entire 242 bp fragment). The average molecular half-life (\( 1/t_{1/2} = (\ln 2)/k_{242} \)) could be calculated from the entire dataset and for each fossil site, respectively. A model of random fragmentation predicts that the relationship between decay rate and fragmentation length is a Poisson distribution (figure 1) [23].

The average per nucleotide rate of decay is therefore calculated by isolating \( k \) in: \( k_{242} = 1 - e^{-k_{242}} \).

(d) Estimating the rate of decay from Illumina HiSeq data

DNA extracts of two South Island giant moa (D. robustus) fossils were selected for Illumina HiSeq 2000 sequencing (see the electronic supplementary material, text S1 for details). Only a fraction of ‘shot-gun’ sequencing data from an ancient bone is expected to represent DNA of the target species. Microbial DNA will constitute a proportion of the reads [60,61]. To retrieve sequences of endogenous moa DNA, only reads that mapped against an unpublished draft of the ostrich (Struthio camelus) genome (source tissue provided by San Diego Zoo and advance access to genomic data provided courtesy of Oliver Ryder, San Diego Zoo; Genome 10K; BGI, China) and the mtDNA genome of a D. robustus moa (GenBank: AY016013.1) [62] were used in the downstream analyses. The software BWA [63] was used in the mapping, using default settings. The resulting fragment length histograms of all mapped reads were used to assess DNA decay rates. According to a model of random fragmentation, the amount of amplifiable template should decline exponentially with increasing fragment size [23] (figure 1a). Log-transformed copy numbers therefore have a linear relationship with amplicon length, with the slope of the decline (\( \lambda \)) describing the probability of a bond in the DNA backbone being cleaved [23]. Lambda can be converted to a damage rate when sample age is known (figure 1b).

3. RESULTS AND DISCUSSION

(a) Estimating the half-life of mtDNA in bone

The calibrated radiocarbon ages of the 158 fossils documented a Mid- to Late Holocene accumulation, ranging from 602 to 7839 BP, calibrated calendar years before present (‘present’ defined as CE 1950; figure 2; electronic supplementary material, table S1). When correlating relative DNA copy number with age, a linear (\( R^2 = 0.002, p = 0.604 \)), inverse (\( R^2 = 0.002, p = 0.598 \)) and S-shaped (\( R^2 = 0.229, p < 0.001 \)) model proved inferior to an exponential relationship (\( R^2 = 0.386, p < 0.001 \); figure 3). Hence, the qPCR results demonstrated that the number of template mtDNA molecules in the fossils was best described with first-order decay kinetics (see the electronic supplementary material, table S4 for individual \( C_T \) values). The equation for the fitted regression was:

\[ N_t = 3.61 \times e^{-0.0013t} \]

with \( N_t \) being the proportion of molecules left at the time \( t \) (years BP), compared with the best sample. Solving the equation yielded a molecular half-life of 521 years for the targeted 242 bp mtDNA fragment, and assuming a Poisson distribution [23], the per site decay rate \( k \) was 5.50 \( \times 10^{-6} \) per year.

As noted, several studies have attempted to investigate DNA degradation kinetics using samples of different ages. Here, we show that by using a controlled experimental set-up, with samples normalized for ambient temperature, an exponential relationship between sample age and DNA preservation can be demonstrated. This supports the view that random DNA fragmentation, likely caused by depurination, applies not only to short-term degradation of DNA in a solution [21,22] but also to ancient DNA in fossil bone.

(b) Comparison with Lindahl & Nyberg’s [22] results

The rate constant \( k \) depends on absolute temperature \( T \) according to the relationship:

\[ \ln k = \ln(A) - \frac{E_a}{R \times (1/T)} \]

where $A$ is the pre-exponential factor, $E_a$ is the activation energy and $R$ the gas constant. On the basis of the Arrhenius plot in Lindahl & Nyberg [22], showing the temperature-dependency of depurination at pH 5, we estimated $k$ to $2.11 \times 10^{-3}$ per site per year (see the electronic supplementary material, text S2), which is 384 times faster than the rate estimated from the moa data (figure 4).

Assuming that long-term DNA fragmentation happens primarily as a result of depurination, this discrepancy is probably related to pH and access to free water. Even if the burial environment is slightly acidic, sacrificial dissolution of microcrystalline carbonated bioapatite will likely act as buffer within the bone. On the basis of the measurements in Berna et al. [64], we used pH 7.5 as an appropriate value for bone (see the electronic supplementary material, Text S2), but still only roughly twice as fast, as the rate expected from Lindahl & Nyberg’s [22] in vitro results is perhaps only roughly twice as fast, as the rate we observe for the moa qPCR estimate.

From equation (2.2), it is clear that both the pre-exponential factor ($A$) and the activation energy ($E_a$) affect the decay rate ($k$), but Lindahl & Nyberg [22] showed similar values of $E_a$ at pH 5 and 6. We therefore assume that the estimated $E_a$ of 127 kJ mol$^{-1}$ also applies to pH 7.5. With this $E_a$ value, and an observed average $k$ for moa mtDNA of $5.5 \times 10^{-6}$ per site per year, and an estimated fossil burial temperature $T = 13.1^\circ$C or 286.25 K (see the electronic supplementary material, text S3), the pre-exponential factor $A$ is determined to $e^{41.2}$. The average decay rate (per site per year) for mtDNA in moa bone is therefore related to temperature as follows (figure 4):

$$
\ln k = 41.2 - 15267.6 \times \frac{1}{T}.
$$

(3.3)

This is arguably the best available approximation for the temperature-dependency of long-term decay of mtDNA in bone, and it should be used as a guide for future work on DNA decay kinetics. We note that Lindahl & Nyberg [22] tested only the temperature dependency of depurination between 45°C and 80°C (318.15–353.15 K), and it remains to be confirmed that the relationship in their Arrhenius plot can be projected to include more typical burial temperatures, as seen in table 1. Furthermore, we note that our model of mtDNA decay in bone does not account for a potential initial post-mortem phase of faster DNA decay, facilitated by nuclease activity rather than depurination [23,31,65]. With these caveats in place, table 1 shows our predicted half-lives, and average fragment length of mtDNA in bone at various temperatures.

(c) Variance in DNA preservation

Only 38.6 per cent (figure 3) of the variation in DNA preservation could be explained by the age of the fossils. Despite efforts to minimize variation (see §2), part of this variance can be ascribed to qPCR stochasticity. Results from qPCR repeats of 50 DNA extractions showed an average of 0.9 cycle shift in $C_T$ value (see the electronic supplementary material, table S5). In 36 of the 50 repeats, the two qPCRs deviated with less than one cycle, confirming the general consistency in the set-
Figure 4. Observed and predicted rates of DNA decay. The predicted survival of DNA in bone through time, measured as intact phosphodiester bonds in the DNA backbone \((a)\), and survival of a 242 bp fragment \((b)\). The depicted survival rates are based on: (i) the average mtDNA decay rate measured directly from qPCR of 158 moa bones; (ii) the rate of depurination measured from DNA in solution at pH 5 in Lindahl & Nyberg [22] but adjusted to 13.1°C to allow comparison with the moa data; (iii) the same rate adjusted further to pH 7.5, as expected inside a bone; (iv) mtDNA and nuDNA decay rates calculated based on Illumina HiSeq data from two moa samples (HiSeq 1 from sample S40114, HiSeq 2 from sample S39946-3). The estimated decay rate \((k\) per site per year) is listed for each of the seven datasets.

Table 1. Predictions of decay rates \((k)\) of mtDNA in bone at various temperatures (based on equation (3.3)). Estimates of mtDNA half-lives for three fragment lengths are indicated as well as the expected average read length \((1/\lambda\) where \(\lambda\) is damage fraction) after 10 000 years. The decay rates do not account for the potential initial post-mortem phase of rapid DNA decay governed by nucleases. Still, the results indicate that under the right conditions of preservation, short fragments of DNA should be retrievable from very old bone (e.g. greater than 1 Myr). However, even under the best preservation conditions at \(-5°C\), our model predicts that no intact bonds (average length \(= 1\) bp) will remain in the DNA ‘strand’ after 6.8 Myr. This displays the extreme improbability of being able to amplify a 174 bp DNA fragment from an 80–85 Myr old Cretaceous bone [1].

<table>
<thead>
<tr>
<th>temperature</th>
<th>(k) per site per year</th>
<th>half-life (years), 30 bp</th>
<th>half-life (years), 100 bp</th>
<th>half-life (years), 500 bp</th>
<th>average length at 10 kyr</th>
<th>time (years) until average length (= 1) bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>25°C</td>
<td>(4.5 \times 10^{-5})</td>
<td>500</td>
<td>150</td>
<td>30</td>
<td>2 bp</td>
<td>22 000</td>
</tr>
<tr>
<td>15°C</td>
<td>(7.6 \times 10^{-6})</td>
<td>3000</td>
<td>900</td>
<td>180</td>
<td>13 bp</td>
<td>131 000</td>
</tr>
<tr>
<td>5°C</td>
<td>(1.1 \times 10^{-6})</td>
<td>20 000</td>
<td>6000</td>
<td>1200</td>
<td>88 bp</td>
<td>882 000</td>
</tr>
<tr>
<td>(-5°C)</td>
<td>(1.5 \times 10^{-7})</td>
<td>158 000</td>
<td>47 000</td>
<td>9500</td>
<td>683 bp</td>
<td>6 830 000</td>
</tr>
</tbody>
</table>
The half-life of DNA

up. The 50 qPCR repeats spanned a total of 14.8 cycles in \( C_E \) implying that roughly 6 per cent (0.9/14.8) of the total variation in \( C_E \) could be ascribed to qPCR stochasticity. This value is the procedurally introduced background noise in our estimated decay rate. Lastly, with regard to accuracy, it is possible that low-template aDNA extracts (typically with \( C_E > 35 \)) could include an amplification signal of template fragments forming as a result of incomplete extension (and where fragments eventually forming a full length product in latter PCR cycles). Such fragmentary chimaeric templates forming as a qPCR artefact might subtly impact our estimates of decay kinetics on the most poorly preserved bones.

The mean \( C_E \) value for the Rosslea samples (39.6) was significantly higher than those from BHV (31.7; \( t \)-test, unequal variances, \( t_{d.f.} = 8 = 4.4, p < 0.004 \) and PV (34.1; \( t \)-test, unequal variances, \( t_{d.f.} = 7 = 2.9, p = 0.022 \). When converted to relative DNA copy numbers, PV fossils contained on average 19.6 per cent of the template molecules present in BHV fossils, and Rosslea contained just 0.3 per cent, corresponding to approximately 330 times fewer DNA templates. A ‘spike-in’ experiment showed that the poor DNA preservation at Rosslea did not result from PCR inhibition (see the electronic supplementary material, table S2). Whereas poorer DNA preservation at Rosslea may be explained by its greater age (figure 2), the difference between the largely contemporaneous PV and BHV sites is more intriguing. This difference was significant, not only when including data from the entire combined PV-BHV time frame (\( t \)-test, equal variances, \( t_{d.f.} = 148 = 4.5, p < 0.001 \), but also within the 1882 years of temporal overlap (915–2797 BP, \( n = 72 \)). Eliminating age as a contributing factor, PV fossils still displayed significantly higher \( C_E \) (i.e. poorer DNA preservation) than those from BHV (\( t \)-test, unequal variances, \( t_{d.f.} = 72 = 3.8, p < 0.001 \).

We tested whether these site-specific differences in DNA preservation could be explained by differing decay rates. After log-transformation, PV (\( R^2 = 0.359 \), ANOVA \( \text{df} = 102; F = 56.53, p < 0.001 \)) and Rosslea (\( R^2 = 0.749 \), ANOVA \( \text{df} = 7; F = 17.91, p = 0.005 \)) both displayed highly significant regressions, whereas the same decay signal was not supported for BHV (\( R^2 = 0.058 \), ANOVA \( \text{df} = 46; F = 2.79, p = 0.10 \); figure 3). The average molecular half-lives for DNA in bones from PV and Rosslea were 506 years and 389 years, respectively. However, an analysis of covariance (based on standard error of difference between slopes) showed that the decay rate was not significantly higher at Rosslea than in PV (\( t_{d.f.} = 107 = 0.92, p = 0.36 \)) or the whole dataset (\( t_{d.f.} = 162 = 0.70, p = 0.48 \)). It was therefore justifiable to pool the data from the three sites to achieve an average DNA decay rate for DNA in a North Canterbury bone (figure 3). We could not detect an exponential decay process in the BHV material; so the decay rate was not estimated for this site alone. Notably, whereas 74.9 per cent of the variance at Rosslea was explicable by geological age, only 5.8 per cent was explicable by age at BHV (figure 3). The discrepancy probably reflects differences in the periods of deposition at the sites (shorter at BHV), allowing geological age to dominate the relationship at Rosslea but not at BHV.

We then examined the effect of storage time. The difference in average \( C_E \) between PV and BHV could be an effect of the seasonal water saturation at the PV site, but it could also be a storage effect, given that the material was collected from BHV in 2001, whereas most of the PV material had been at Canterbury Museum for more than 50 years (figure 2). However, we note that the DNA in bones from Rosslea was more degraded, despite the bones having been sampled shortly after excavation. Moreover, the range in \( C_E \) (figure 2) was highest at BHV although the 47 bones from this site had identical storage times (7 years between excavation and sampling). A previous investigation showed that storage conditions can be a major contributor to variation in DNA preservation [50]. Our results nuance this observation by showing that storage time in itself does not seem to have a large effect. A multiple regression analysis showed that both variables (age and storage) had a significant effect (\( p < 0.001 \) for both), but whereas in combination, they explained 45.2 per cent of the variation in DNA preservation, only 7.7 per cent of this could be ascribed solely to storage time. Therefore, in our experimental setup, the geological age of a sample was a far better predictor of its DNA preservation than storage time.

(d) Decay rates from Illumina data: mtDNA versus nuDNA

Illumina sequencing of two DNA extracts yielded 10.1 and 24.8 million reads, respectively. When filtered and mapped against a draft of the ostrich genome (see §2), the return was 3128 and 34 495 unique, non-duplicated reads (see the electronic supplementary material, table S6). These should represent endogenous fragmented moa DNA and hence suitable for calculations of DNA decay rates. It was first confirmed that the number of fragments declined exponentially with fragment length (see the electronic supplementary material, table S6 and figure S3), thereby conforming to the random fragmentation model [23]. The DNA decay rate (\( k \)) was calculated as \( 2.13 \times 10^{-5} \) and \( 2.71 \times 10^{-5} \) per site per year, for each sample, respectively (figure 4). The majority of DNA within most vertebrate cells is derived from the nucleus, thus it was a priori hypothesized that the majority of the mapped ‘shotgun’ reads would represent nuclear DNA fragments. If there is a difference between post-mortem DNA decay rates in the nuclear and mtDNA genomes, it could explain why the rate is four to five times faster than estimated from our qPCR data (5.5 \( \times 10^{-6} \) per site per year), which represent fragmentation of mtDNA.

To investigate this further, the shotgun sequences were then mapped against the moa (\( D. \text{robustus} \)) mitochondrial genome, returning 878 and 4032 reads, respectively. When the decay rate was calculated based only on these mtDNA sequences, it was found to be 2 and 2.5 times slower, respectively, than that estimated from the entire dataset (figure 4, electronic supplementary material, table S6). These numbers suggest that mtDNA degrades at a slower rate than nuDNA, consistent with observations reported in the study of Schwarz et al. [24]. This could perhaps be explained by the circular structure of mtDNA, making it less accessible to exonuclease activity.

The two decay rates of mtDNA measured from Illumina data appear slightly faster than the rate based on qPCR (figure 4). This could be sampling error because these are just two individual data points compared with the qPCR average of 158 samples. However, it could also be explained by the fact that the qPCR decay rate
is based on differences in DNA preservation between bones that have all been in the ground for more than 600 years. This rate does not therefore incorporate the presumed faster DNA degradation occurring in the initial post-mortem phase. It has also been demonstrated that the library preparation for high-throughout sequencing may skew the fragment length distribution towards shorter average reads [66], and such bias could potentially explain the observed rate difference between our qPCR and HiSeq data. However, it is not clear whether such bias would affect only the average read length, or whether it could also alter the slope of the frequency distribution, which we use to estimate $k$. Further data are required to clarify these potential biases.

4. CONCLUDING REMARKS

We have demonstrated that in situ DNA decay is described by first-order kinetics, confirming that long-term post-mortem DNA fragmentation can be treated as a rate process. This closes the gap to theoretical in vitro observations made four decades earlier. We argue that equation (3.3) represents the best available approximation of the rate of mtDNA decay in fossil bone. Empirical data from a range of different depositional environments, also extending back into the Pleistocene, are now needed to further test and refine this model.

Importantly, we show how high-throughput sequencing data can be used to estimate DNA decay rates, and thus the vast amount of such data currently being generated worldwide may provide an excellent resource by which to study DNA decay. Factors such as bone thickness, burial depth, surrounding pH and water saturation can be tested and modelled in future research, whereas information on season of death, speed of cadaver incorporation into sediment and climatic fluctuations will probably be less accessible.

It is tempting to suggest that we can now predict the temporal limits of DNA survival, and finally refute the claims of authentic DNA from Cretaceous and Miocene specimens. This is, however, not straightforward. One needs information on the number of template molecules in living tissues, and estimates of post-mortem DNA decay rates for each tissue type. However, the half-life predictions (table 1) display the extreme improbability that an authentic 174 bp long mtDNA fragment of an 80–85 Myr old bone could have been amplified [1].

Our results indicate that short fragments of DNA could be present for a very long time; at ~5°C, the model predicts a half-life of 158 000 years for a 30 bp mtDNA fragment in bone (table 1). Even rough estimates such as this imply that sequenceable bone DNA fragments may still be present more than 1 Myr after deposition in deep frozen environments. It therefore seems reasonable to suggest that future research may identify authentic DNA that is significantly older than the current record of approximately 450–800 kyr from Greenlandic ice cores [47].

We are grateful to Malene Møhl for help with the sampling and to Emma McLay, Helen Hunt and Jayne Houston for technical assistance, and to Soren Overballe-Petersen for useful discussions. We thank the Hodgen (PV), Giesen (BHV) and Earl (Rosslea) families for their support of our research. Thanks to Oliver Ryder, San Diego Zoo for providing ostrich tissue. Financial support was provided by the Marsden Fund of the Royal Society of New Zealand (contract 06-PAL-001-EEB to Palaeol Research Ltd). M.C. and D.H. received support from the SYNTHESYS project (http://www.synthesys.info), financed by European Community Research Infrastructure Action under the FP7 ‘Capacities’ programme. M.T.P.G. and J.A.S. were supported by the Danish Council for Independent Research (Grant 10-081390). M.B. is funded by the Australian Research Council (Future Fellowship FT0991741).

REFERENCES


Lindahl, T. 1993 Instability and decay of the primary structure of DNA. Nature 362, 709–715. (doi:10.1038/362709a0)


Buckley, M. et al. 2008 Comment on ‘Protein sequences from mastodon and Tyrannosaurus rex revealed by mass...
spectrometry'. Science 319, 33; author reply. (doi:10.1126/science.1147046)


