Manuscript

Click here to access/download;Manuscript;aDNA_Damage_v7.6.docx

Manuscript title:

Patterns of aDNA Damage Through Time and Environments – lessons from herbarium specimens

Authors:

Stefano Porrelli[1,*], ORCID: 0000-0003-3878-7745

Alice Fornasiero[2], ORCID: 0000-0001-6165-4233

Hong Phuong Le[1], ORCID: 0000-0002-2862-6502

Wenzhe Yin[3,4], ORCID: 0000-0002-4350-3578

Maria Navarrete Rodriguez[2], ORCID: 0000-0001-5052-3713

Nahed Mohammed[2], ORCID: 0000-0002-8857-3246

Axel Himmelbach[5], ORCID: 0000-0001-7338-0946

Andrew C. Clarke[3], ORCID: 0000-0003-2293-1423

Nils Stein[5,6], ORCID: 0000-0003-3011-8731

Paul J. Kersey[1], ORCID: 0000-0002-7054-800X

Rod A. Wing[2,7] ORCID: 0000-0001-6633-6226

Rafal M. Gutaker[1,*], ORCID: 0000-0001-9226-879X

[1] Trait Diversity and Function, Royal Botanic Gardens, Kew, Richmond, TW9 3AE, United Kingdom.

[2] Plant Science Program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia.

[3] School of Biosciences, University of Nottingham, Sutton Bonington, LE12 5RD, United Kingdom.

[4] State Key Laboratory of Emerging Infectious Diseases, School of Public Health, The University of Hong Kong, Hong Kong, SAR, China.

[5] Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, 06466, Germany.

[6] Institute of Agricultural and Nutritional Sciences, Martin Luther University of Halle-Wittenberg, Halle (Saale), Germany.

[7] Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, AZ, USA.

* To whom correspondence should be addressed: R.Gutaker@kew.org; S.Porrelli2@kew.org

Abstract:

Herbarium collections are a vast but underutilized resource for ancient DNA research, containing over 400 million specimens with detailed metadata and spanning centuries of global biodiversity. Understanding patterns of DNA preservation in natural collections is crucial for optimizing ancient DNA studies and informing future curation practices. We analysed genomic data for 573 herbarium specimens from six plant species from the genera *Hordeum* and *Oryza* collected from the Americas and Eurasia over 220 years. Using standardized laboratory protocols and shotgun sequencing, we quantified DNA degradation and elucidated factors that accelerate it. We find significant age-dependent DNA fragmentation rates, indicating temporal degradation processes not detected in prehistoric samples. In our analysis, DNA decay rates in herbarium specimens were almost eight times faster than in moa bones, reflecting fundamental differences in tissue composition and preservation environments. Environmental conditions at the time of specimen collection emerged as the major determinants of post-mortem damage rates, with the interaction term between temperature and genus being the dominant driver of

cytosine deamination. We find no effect of sample storage on DNA damage and degradation. These findings provide insights into how climatic origin, preservation environment, taxonomic identity and age influence DNA preservation while highlighting opportunities for improving institutional preservation practices. Due to standardised preservation conditions, museum collections can provide better insights into DNA damage and degradation over time than archaeological and paleontological samples.

Introduction:

Understanding preservation of DNA from old biological samples, commonly referred to as ancient DNA (aDNA), has been at the core of the genomic revolution in archaeological research [1,2]. There is a consensus among researchers that aDNA is defined by its degradation and not by its age [3,4], though the two are correlated [5]. The two most prominent *post-mortem* reactions associated with DNA damage and degradation are deamination and depurination, which occur spontaneously in the absence of enzymatic repair machinery [6,7]. Deamination of cytosines into uracils leads to a characteristic pattern of 'C>T' misincorporations, with increased frequency at fragment termini [8], whilst depurination causes "nicks" (breakage) in the phosphodiester bonds and subsequent hydrolysis of DNA backbone, resulting in DNA fragmentation towards very small molecules (~30-100 bp). Both patterns are common in historical, archaeological and sedimentary samples [2,5,9] and make bioinformatic processing and downstream analyses challenging [10]. Although 5' C>T misincorporations (and the complementary 3' G>A misincorporations in double-stranded libraries [3]) can be used to authenticate genuine aDNA sequences, they can bias variant calling and phylogenetic analyses, potentially leading to incorrect inferences [3]. Similarly, extensive DNA fragmentation reduces mapping efficiency and increases the likelihood of spurious alignments [11,12]. These challenges are particularly acute for ancient samples from hot and humid environments, where

accelerated degradation processes further compromise DNA integrity [13,14]. As a result, tools and approaches have been developed to improve quality control, reads processing, mapping and downstream analyses in a quest to utilize even highly degraded samples [15–17]. In addition to chemical degradation, DNA can also be altered through biological processes such as microbial colonization. In effect, the *bona fide* DNA from the target species can be depleted and replaced with post-mortem microbial DNA [18].

While the vast majority of published aDNA research focuses on human skeletal remains [19,20], there are an increasing number of studies in other mammals [21], arthropods [9,22], and in plants [23–25]. Archaeobotanical materials, primarily seeds, are often used as sources of prehistoric DNA. While not older than 500 years, herbarium specimens (figure 1) number over 400 million [26], have solid species identification and, frequently, collection metadata. They comprise a vast but underutilised resource for studying evolutionary and ecological changes in the last couple of centuries [27–30]. The leaf tissue is most often the target in DNA isolation protocols but does not possess the same structural isolation from the environment as bones or seeds. This is the most likely reason why DNA in herbarium samples decays at a rate roughly six times faster than in ancient Moa bones [25], but also more than twice as slow when compared to dry-pinned arthropod museum specimens [9]. On the other hand, herbarium specimens are commonly dried upon collection and are generally stored in standardized conditions that are favourable for DNA preservation, such as stable temperatures and low air humidity. Availability of materials, good geolocation information and relatively stable preservation condition *ex situ* make herbarium specimens a perfect system to study the effects of the environment at the point of collection on the subsequent preservation of DNA in biological samples.

Figure 1: Representative herbarium specimens for the six species analysed in this study. Photos courtesy of Royal Botanic Gardens, Kew (RBGK) and Missouri Botanic Gardens (MBG). (a) *Hordeum vulgare* sample HV0080, collected in 1849 (RBGK); (b) *Hordeum spontaneum* sample HS0156, collected in 1909 (RBGK); (c) *Oryza alta* sample HRG0858, collected in 1913 (MBG); (d) *Oryza rufipogon* sample HRG0849, collected in 1926 (RBGK); (e) *Oryza latifolia* sample HRG0287, collected in 1920 (RBGK); (f) *Oryza grandiglumis* sample HRG0580, collected in 1989 (MBG).

One of the major challenges in studies trying to investigate the dynamics of DNA damage and fragmentation is lack of large datasets with a global distribution and consistent sampling and

laboratory processing procedures. Previous studies were limited to one or two species representing narrow geographical range [13,25]. Here, we present new sequencing data for a total of 573 herbarium samples from six plant species, spanning the Americas and Eurasia, processed with the same laboratory protocol in dedicated aDNA laboratory facilities, and sequenced using a whole-genome shotgun approach. Our main aim is to understand which environmental factors influence the rates of DNA damage and decay. The outcome of this investigation brings new insights into the fundamental processes of DNA degradation that can inform future conservation of museum specimens, as well as help researchers to prioritize materials for their studies.

Methods

Herbarium samples

Historical herbarium samples (N = 573) were obtained from seven herbaria in Europe and North America (table 1): Royal Botanic Gardens, Kew (UK), Smithsonian Institute Herbarium (US), Missouri Botanic Gardens (US), New York Botanic Gardens (US), National History Museum (France), Harvard University Herbarium (US), and National History Museum (UK). Passport information for each sample is included in the supplementary material (supplementary table S1). Although we aimed at representing the majority of environments for each species in our dataset, some parts of distribution might have been overlooked due to unavailability of samples in our collections. The *Hordeum* samples were predominantly collected from temperate Europe to semi-arid regions of the Middle East, with highest counts from countries such as Iraq, Israel and Iran. The *Oryza* samples were predominantly collected from tropical and subtropical regions of the Americas and Southeast Asia, particularly Brazil, Costa Rica, Colombia, Mexico and Thailand (figure 2).

Table 1: Number of historical herbarium samples (N = 573) included in this study collected from seven herbaria. RBGK: Royal Botanic Gardens Kew, UK; SIHUS: Smithsonian Institute Herbarium, US; MBG: Missouri Botanic Gardens, US; NYBGUS: New York Botanic Gardens, US; NHMFR: National History Museum, France; HUHUS: Harvard University Herbarium, US; NHMUK: National History Museum, UK.

Species	**RBGK**	**SIHUS**	**MBG**	**NYBGUS**	**NHMFR**	**HUHUS**	**NHMUK**	**Total**	**Age range**
Hordeum spontaneum	146	0	0	0	0	0	0	146	1822-2017
Hordeum vulgare	82	0	0	0	0	0	0	82	1819-2015
Oryza alta	0	2	9	0	2	0	0	13	1913-2002
Oryza grandiglumis	1	30	1	0	6	0	0	38	1930-2002
Oryza latifolia	22	97	36	30	20	16	10	231	1797-2011
Oryza rufipogon	29	31	3	0	0	0	0	63	1909-1998
Total	280	160	49	30	28	16	10	573	1797-2017

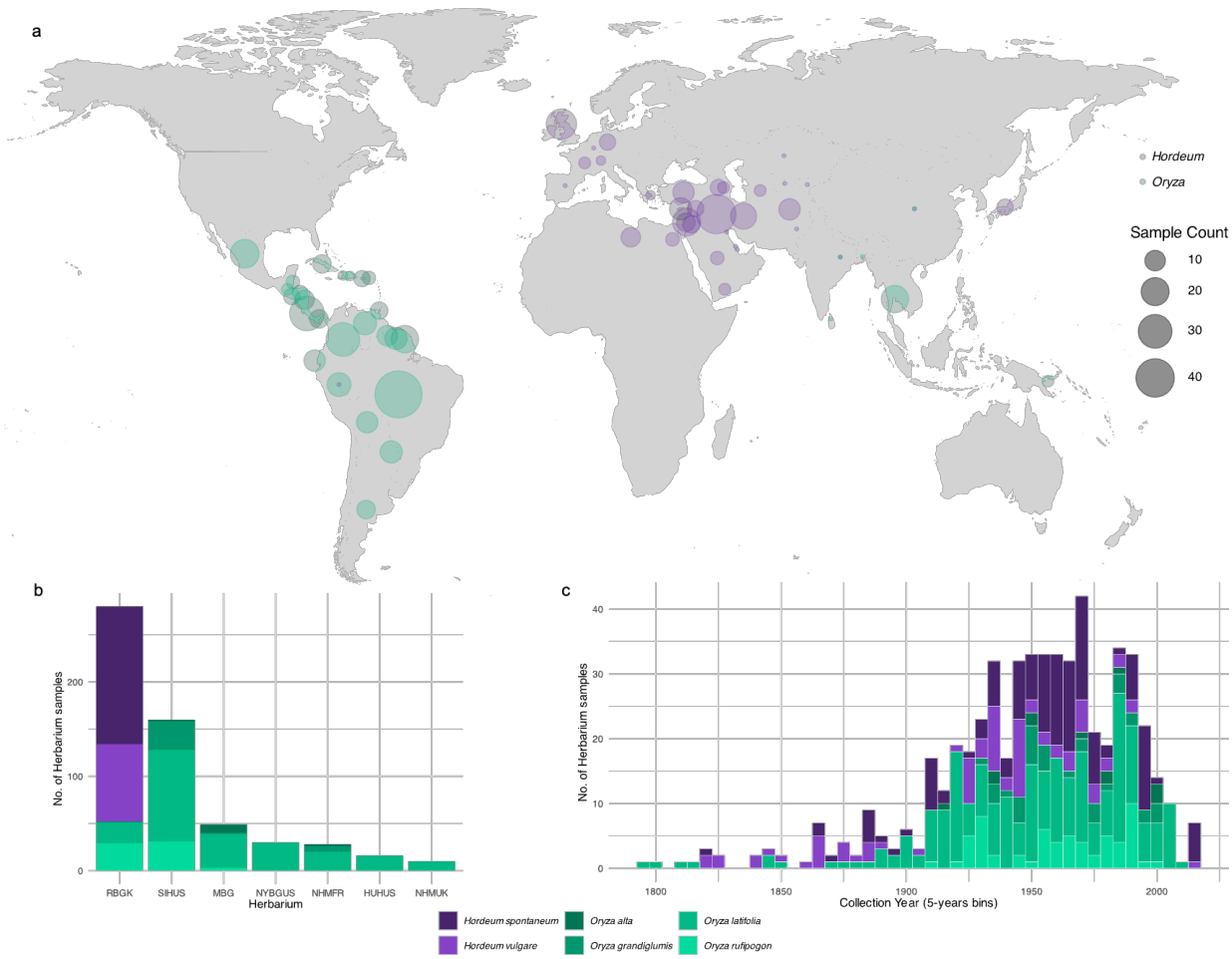


Figure 2: Overview of the herbarium samples of barley (*Hordeum*) and rice (*Oryza*) species: (a) geographical distribution based on country of collection; (b) source of herbarium of *Hordeum* and *Oryza* species (RBGK: Royal Botanic Gardens Kew, UK; SIHUS: Smithsonian Institute Herbarium, US; MBG: Missouri Botanic Gardens, US; NYBGUS: New York Botanic Gardens, US; NHMFR: National History Museum, France; HUHUS: Harvard University Herbarium, US; NHMUK: National History Museum, UK); (c) temporal distribution for each species.

DNA extraction, library preparation and low throughput sequencing:

Laboratory steps for DNA extraction and library preparation for all 228 *Hordeum* (146 *H. spontaneum* and 82 *H. vulgare)* and 58 *Oryza* (30 *O. latifolia,* 18 *O. rufipogon* and 10 *O. grandiglumis*) (supplementary table S1) samples were carried out in a clean room facility at

155 Royal Botanic Gardens, Kew (RBGK, UK) following previously published best practices and
156 aDNA protocol described in [31] with adjusted volumes. A further 287 *Oryza* samples (201 *O.*
157 *latifolia*, 45 *O. rufipogon*, 28 *O. grandiglumis* and 13 *O. alta*) (supplementary table S1) were
158 processed in a clean room facility at the Ancient and Environmental DNA Laboratory (ÆDNA)
159 at the University of Nottingham (UoN, UK), following the same protocol and volume
160 adjustments. Following best practices for aDNA research [3,31], negative controls were
161 processed alongside samples to monitor for environmental and reagent contamination. These
162 included extraction blanks and library preparation controls.
163
164 Leaf tissue samples (5-10 mg) were grinded in 2mL PowerBead metal tubes (QIAGEN, 13117-
165 50) with a bead mill homogeniser (Precellys Evolution, P002511-PEVT0-A.0). 1mL of PTB-
166 based mix was added to the homogenized samples, which were then incubated on a rotor
167 overnight at 37°C. Genomic DNA (gDNA) was isolated and half the volume (0.5 mL) was
168 purified using the DNeasy Plant Mini Kit (QIAGEN, 69106) with modifications described in
169 [31], and quantified with Quantus™ fluorometer (Promega, E6150). A subset of the gDNA
170 isolates (*N*=40, 10 *H. spontaneum*, 10 *H. vulgare*, 10 *O. rufipogon*, and 10 *O. latifolia*;
171 supplementary table S2) with concentrations > 5 ng/µL were selected for measurements of
172 gDNA fragment size distributions with capillary electrophoresis. These samples were analysed
173 on a TapeStation 4200 system (Agilent, G2991BA) using a D1000 ScreenTape (Agilent, 5067-
174 5582).
175
176 For all 573 samples, double stranded and doubled indexed genomic libraries were constructed
177 by blunt-end DNA ligation [32,33] following [31]. Briefly, genomic DNA was blunt-end
178 repaired, and universal Illumina double-stranded adapters were ligated. This was followed by
179 fill-in of adapters, indexing with unique combinatorial Illumina indexes, and amplification.

180 Between each step, samples were purified with the MiniElute® PCR purification kit
181 (QIAGEN, 28006). To quantify the excess of 5' C>T misincorporations at reads termini caused
182 by spontaneous deamination of cytosines [34], and to validate aDNA authenticity [25,35], we
183 did not perform enzymatic removal of aDNA-associated DNA misincorporation [36]. Indexed
184 and amplified libraries were quantified with Quantus™ fluorometer (Promega, E6150).
185 Fragment size of each library were estimated with a 4200 TapeStation System (Agilent,
186 G2991BA) using a D1000 ScreenTape (Agilent, 5067-5582).
187
188 Libraries were pooled equimolarly and sequenced in paired-end mode. All 228 *Hordeum*
189 libraries processed at RBGK were sequenced on a NovaSeq 6000 or NovaSeq X+ device (2 x
190 62 bp) according to manufacturer's instructions (Illumina, San Diego, CA, USA) at the Leibniz
191 Institute of Plant Genetics and Crop Plant Research (IPK, Gatersleben, Germany). The 58
192 *Oryza* libraries processed at RBGK were sequenced at Macrogen (Macrogen, Europe), on an
193 Illumina Novaseq X+ system (2 x 150 bp). The 287 *Oryza* samples processed at the Ancient
194 and Environmental DNA Laboratory (ÆDNA) at the University of Nottingham (UoN, UK)
195 were sequenced on an Illumina Miseq platform (2 x 150 bp) according to manufacturer's
196 instructions (Illumina, San Diego, CA, USA) at the Deep Seq facility at UoN.
197
198 **Bioinformatic screening and validation of aDNA authenticity**
199 Bioinformatics protocols for read processing, quality assessment, screening and authentication
200 of aDNA-derived libraries follow [31]. Demultiplexed raw Illumina paired-end reads were
201 adapter trimmed and merged with AdapterRemoval v2.3.4 [37] with a minimum overlap of 11
202 bases, and quality-checked with FastQC v0.11.8
203 (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc). Reference genome assemblies were
204 retrieved from the National Center for Biotechnology Information (NCBI) database (table 2)

205 and indexed with BWA v0.7.19 [38] and SAMtools v1.21 [39]. For each species, quality- and
206 adapter-trimmed merged reads were mapped to their respective reference genome with BWA-
207 aln using aDNA-specific parameters ("-l 1024"). PCR optical duplicate reads were removed
208 with DeDup [17]. Base frequencies at DNA break points and nucleotide misincorporations
209 were estimated for each library using MapDamage2 v2.0.6 [40]. Mapping statistics of the
210 additional 287 *Oryza* samples were obtained by mapping sequencing reads to the following
211 reference genomes: *O. alta* (PRJNA1039467), *O. grandiglumis* (PRJNA737282), and *O.*
212 *latifolia* (PRJNA737486) [41]. The *O. rufipogon* samples originally collected in the American
213 continent were aligned to the *O. glumipatula* genome reference (PRJNA48429). Negative
214 laboratory controls (extraction and library blanks) were included in the bioinformatic screening
215 to monitor/assess possible contamination, but they were excluded from further analyses.
216
217 Table 2: Reference genome assemblies used in this study with associated BioProject,
218 BioSample and Accession IDs.

Reference	**BioProject**	**BioSample**	**Assembly Accession**
Hordeum spontaneum	PRJEB57567	SAMEA112465237	GCA_949783245.1
Hordeum vulgare	PRJEB40589	SAMEA7384724	GCF_904849725.1
Oryza alta	PRJNA1039467	SAMN38217704	GCA_047899615.1
Oryza grandiglumis	PRJNA737282	SAMN19687255	GCA_048188845.1
Oryza latifolia	PRJNA737486	SAMN19696891	GCA_048174585.1
Oryza rufipogon	PRJNA1029807	SAMN40302812	GCA_037997075.1
Oryza glumipatula	PRJNA48429	SAMN02981440	GCA_000576495.2

219
220 **Ancient DNA damage metrics and regression analyses**
221 Four metrics were selected to quantify patterns of aDNA damage: (i) the proportion of
222 endogenous DNA content, (ii) the fragment length distribution, (iii) the damage fraction per
223 site (λ), and (iv) the frequencies of 5' C>T misincorporations at the first base. The four metrics

224 were analysed in linear models as a function of collection year and sample age using the '*lm*'
225 function in R [42]. Model assumptions were verified using standard diagnostic procedures. For
226 all linear regressions, we examined residual plots to assess the assumptions of normality and
227 homoscedasticity.
228
229 *Endogenous content*
230 The proportion of endogenous DNA was calculated for each sample as the fraction of quality-
231 and adapter-trimmed merged reads that mapped to their respective reference genome divided
232 by the total number of reads retained after adapter-trimming and quality filtering. Only primary
233 alignments were considered, with mapped reads representing unique mapping decision for each
234 query sequence. Since this metric can be impacted by increasing evolutionary distance between
235 target and reference [9], we used the most closely related reference genome available for each
236 of the species included in our dataset (table 2). Endogenous content was calculated with
237 SAMtools "*flagstat*" tool [39] and plotted as a function of collection year.
238
239 *Fragment length*
240 DNA fragmentation can be quantified by agarose gel electrophoresis, automated
241 electrophoresis or by the *in-silico* generation of fragments by merging overlapping paired reads
242 of Illumina libraries [43,44]. To validate that the fragment sizes of the merged reads reflects
243 the original molecule length, we examined the relationship between the fragment size
244 distribution of isolated gDNA and that of the amplified libraries for a subset of the samples
245 (N= 40; 10 *H. spontaneum*, 10 *H. vulgare*, 10 *O. rufipogon*, and 10 *O. latifolia*; supplementary
246 table S2) using TapeStation profiles. Library and gDNA peak size measurements from the
247 TapeStation profiles were compared with each other and with bioinformatically-derived
248 median fragment lengths of merged reads to assess the concordance between direct physical

measurements and computational estimates. In addition, we further analysed the correlation between peak sizes of the gDNA isolates and age of the sample. Having established a relationship between fragment size distributions of gDNA and bioinformatically-derived merged reads for a representative subset (N=40), we examined the relationship between fragment size of merged reads and age for the whole dataset (N= 573). We fitted the fragment length distribution of the mapped merged reads to a lognormal distribution using the 'fitdistr' function from the MASS package [45] in R. We used the mean of this distribution (log-mean) to summarise fragment length for each library and carried out regressions on the relationship between the log-mean of fragment lengths and collection year. As suggested in [25], we used the median plotted on a log-scaled y-axis for visualisation, as the latter is more intuitive to understand than the log-mean.

Damage fraction per site (λ) and DNA decay rate (k)

The damage fraction per site (λ) was calculated for each sample using previously described methods [13]. We fitted an exponential decay model to the empirical fragment length distribution of the mapped reads and derived λ as the negative of the slope coefficient of the linear regression. To identify samples deviating from the assumption of exponential decay of fragment length, we assessed goodness of fit and statistical significance on a per-sample basis. Model fit was considered satisfactory when $R^2 > 0.95$ and $p < 0.05$.

Having determined the damage fraction per site (λ), we calculated genus-specific DNA decay rates (k) by plotting λ as a function of sample age and extrapolating k from the slope of the linear regression, according to the linear relationship $k = \lambda/age$ [46]. We also calculated the overall DNA decay rate (k) of all herbarium samples irrespective of genus.

Nucleotide misincorporations

274 The frequency of 5' C>T misincorporations at first position was used as a proxy for 5' damage.
275 Although the depth of sequencing needed for reliable estimation of C>T frequencies is sample-
276 dependent, an order of magnitude of thousands of reads is considered sufficient [47]. Therefore,
277 only samples with >5,000 merged paired reads were retained for further analyses. Furthermore,
278 the exponential increase of C>T misincorporations at reads termini is routinely used as a metric
279 to validate authenticity of aDNA [5]. We scored this pattern for each sample by fitting an
280 exponential model to the C>T frequencies for the first 20 bases at 5' terminus. We evaluated
281 the goodness of fit with a one-sided *t*-test on a per-sample basis, as described in [48], and only
282 retained samples that fitted the exponential model ($R^2 > 0.5$, $p < 0.05$). Finally, we carried out
283 a regression analysis of 5' C>T damage as a function of sample age.
284 For comparison, and to demonstrate that both 5' and 3' ends in double-stranded libraries show
285 concordant deamination patterns, we also analysed the complementary 3' G>A
286 misincorporations. To demonstrate the patterns observed are specific to deamination, we
287 computed the average frequency of non-deamination substitutions for all samples following
288 the approach described in [12], and plotted this as a function of sample age. To account for
289 increased rates of baseline substitutions arising from evolutionary divergence between samples
290 and reference genomes, we calculated mean divergence from reference on a species-basis. We
291 also determined divergence-corrected deamination rates by subtracting baseline substitution
292 frequencies from total deamination frequencies and compared the results against the non-
293 corrected deamination rates.
294
295 **Analysis of covariance**
296 To investigate differences in aDNA damage patterns between *Hordeum* and *Oryza*, we
297 performed Analysis of Covariance (ANCOVA) using the 'anova' function in R. For each aDNA
298 damage metric, sample age was used as the covariate and the genus as the factor using the

model "*y ~ covariate* x *factor",* which also tested for possible interactions between sample age and genus (i.e., differences in the slope of regression are dependent on genus). We tested this model against a model of type "*y ~ covariate + factor"* to assess whether the removal of the interaction influenced model fit. When the interaction was not significant ($p > 0.05$) we accepted the simpler model without interaction and concluded that regression slopes did not differ between genera, though intercepts might [25].

Climate analyses

We obtained climate data from the CHELSA V2.1 climate dataset [49] to investigate the relationship between specimen preservation conditions, aDNA damage patterns, and climatic variables. Four primary bioclimatic variables were extracted for each sample location based on latitude and longitude metadata: (i) annual mean temperature (bio1), (ii) temperature seasonality (bio4), (iii) annual precipitation (bio12), and (iv) precipitation seasonality (bio15), henceforth referred as "annual climate". Additionally, we retrieved monthly temperature (tas_01 - tas_12) and monthly precipitation (pr_01 - pr_12) means for year 1981-2010 (CHELSA V2.1) corresponding to each sample's recorded month of collection and geographical location. We used this data to infer climatic conditions at the time of specimen collection, henceforth referred as "collection climate".

To quantify the unique and shared contributions of multiple explanatory variables to the total variance in aDNA damage metrics, we performed a variance partitioning analysis using the 'varpart' function implemented in the VEGAN package [50]. For each aDNA metric, we applied a "collection climate" model, where monthly climatic variables (temperature and precipitation) were assigned to samples based on their recorded location and month of collection, and an "annual climate" model, where annual mean temperature, mean precipitation and their seasonality were assigned to samples based on their geographical location. The

"collection climate" model aimed at capturing DNA damage occurring during the initial post-collection period (field handling, drying, and early preservation), when specimens are exposed to ambient environmental conditions, rather than during subsequent long-term storage under standardised herbarium conditions [2,25,51]. The "annual climate" model aimed at capturing the general climatic regime of the collection locality, thus providing a comparison to assess whether month-specific climate data improves explanatory power over annual averages. Annual climate variables may also better represent cumulative environmental exposure if pre-collection conditions (e.g., growing season climate) influence tissue properties relevant to DNA preservation.

As a possible confounding effect, we also included genus as a variable, as well as herbarium, which could be interpreted as different long-term institutional storage conditions. Since 'varpart' function accommodates only up to four variables, for some analyses, temperature and precipitation were merged into single climatic variable. For each response (endogenous DNA content, fragment size, lambda and 5' C>T damage), statistical significance of individual fractions and combinations were tested using redundancy analysis with the 'rda' function implemented in VEGAN, with 999 permutation followed by ANOVA.

Divergence-corrected nucleotide misincorporations analyses

To assess whether reads-reference divergence influenced our main findings, we repeated all primary analyses using divergence-corrected deamination frequencies. For each sample, corrected 5' C>T frequencies were calculated by subtracting the mean baseline substitution rates from the observed 5' C>T frequencies. Specifically, we re-ran the regression analysis of corrected deamination frequencies against sample age, the variance partitioning analyses to quantify the relative contributions of temperature, precipitation, age, and genus to variation in corrected deamination, and the regressions between corrected deamination and temperature.

349

350 **Results:**

351 A total of 573 herbarium specimens were sequenced, generating libraries sequenced to variable
352 read counts and depth (supplementary table S1). *Hordeum* samples (*N* = 228) processed at the
353 Royal Botanic Gardens, Kew (RBGK) generated an average of 12.08 ± 5.94 million reads per
354 library, with 86.2 ± 24.3% of reads mapping to the relevant reference genome. *Oryza* samples
355 (*N*= 58) processed at the RBGK yielded an average of 1.41 ± 0.93 million reads per library,
356 with 86.5 ± 14.6% of reads mapping to the relevant refence genome. The remaining *Oryza*
357 samples processed at the University of Nottingham (*N*= 287) generated an average of 0.06 ±
358 0.09 million reads per library, with 78.0 ± 26.2% of reads mapping to the relevant reference
359 genome.

360 Baseline substitution rates, calculated as the mean frequency non-deamination substitution
361 types, served as a proxy for evolutionary distance between samples and their respective
362 reference genomes. Overall, sample-reference divergence rates were low across all species,
363 ranging from 0.53% to 0.88% (supplementary figure S1). *Hordeum* species showed the lowest
364 divergence from their respective reference genome, with *H. vulgare* at 0.53% ± 0.52% and *H.*
365 *spontaneum* at 0.57% ± 0.27%. Among *Oryza* species, *O. grandiglumis* showed the lowest
366 divergence (0.68% ± 0.16%), followed *by O. rufipogon* (0.76% ± 0.31%), *O. latifolia* (0.80%
367 ± 0.34%), and *O. alta* (0.88% ± 0.08%).

368

369 Damage pattern analysis revealed characteristic aDNA signatures. *Hordeum* samples exhibited
370 an average 5' C>T misincorporation frequency of 1.09 ± 0.44% at the first base of sequenced
371 molecule, whilst *Oryza* samples exhibited an average of 2.13 ± 0.70%. These damage
372 signatures, combined with fragment length distributions, were used to authenticate samples and
373 filter out those inconsistent with genuine aDNA characteristics. Our filtering strategy identified

374 samples deviating from the assumption of exponential decay of fragment length, as well as
375 samples that did not display an exponential increase of C>T misincorporations at read termini
376 diagnostic of aDNA. In total, 117 (20%) samples were removed, leaving a final dataset
377 containing 456 samples.
378
379 **Regression analyses**
380 *Endogenous fraction*
381 We analysed the proportion of reads mapping to the reference genome as a proxy for
382 endogenous DNA content. The regression analyses revealed no statistically significant
383 relationship between the proportion of endogenous DNA and the sample collection year in
384 *Hordeum* (R^2 = 0.003, p = 0.451, N = 211), but a very weak yet significant relationship was
385 observed in *Oryza* (R^2 = 0.04, p = 0.00167, N= 245; figure 3a). As we aimed at investigating
386 the effect of genus on the rates of aDNA damage, we also carried the regression for all samples
387 irrespective of genera, which provided comparable results (R^2 = 0.012, p = 0.0215, N= 456,
388 supplementary figure S2a).
389
390 *Fragment length*
391 We measured DNA fragment size using two complementary approaches. First, we validated
392 our bioinformatic estimates of fragment size distributions using TapeStation profiles of both
393 gDNA and amplified libraries for a subset of samples (N= 40). Peak size of amplified libraries
394 strongly correlated with the bioinformatically-derived median fragment size of the merged
395 library reads (R^2 = 0.614, p =2.22 x 10^{-9}, supplementary figure S3a), compared to the weaker
396 correlation between the TapeStation peaks of libraries and gDNA origin, (R^2 = 0.129, p =0.029,
397 supplementary figure S3b). A significant, albeit weaker correlation was also observed between
398 the gDNA peak size and the median fragment size of merged reads (R^2 = 0.287, p =6.35 x 10^{-}

399 ^{4}, supplementary figure S3c), indicating that merging of overlapping reads of short insert
400 libraries reflects, at least in part, the original molecule length [32]. The weaker correlation was
401 expected, as library preparation and sequencing involve processing and purification steps that
402 can impact the fragment size distribution. Furthermore, we observed a strong and significant
403 relationship between gDNA peak size and collection year (R^2 = 0.61, p =3.04 x 10^{-8}, N= 40,
404 figure 4), but only after removing the two *Hordeum* outliers HV0061 and HV0081, which were
405 particularly old (collection years: 1842 and 1867, respectively) but showed gDNA size
406 distribution patterns inconsistent with their age, possibly due to modern contamination of
407 exogenous DNA. With the inclusion of these two outliers, we still observed a significant albeit
408 weaker relationship (R^2 = 0.263, p =1.19 x 10^{-3}, supplementary figure S4).
409 Second, we extended this analysis to the full dataset and examined the relationship between
410 bioinformatically-derived fragment size and age. We observed a statistically significant
411 relationship between the log-mean fragment length and the sample collection year for both
412 genera (figure 3b), with a stronger relationship for *Hordeum* (R^2 = 0.27, p =5.33 x 10^{-16}, N=
413 211) than *Oryza* (R^2 = 0.112, p = 8.47 x 10^{-8}, N= 245). A statistically significant relationship
414 was also observed when analysing all samples irrespective of genera (R^2 = 0.171, p = 3.16 x
415 10^{-20}, N= 456; supplementary figure S2b).
416
417 *Damage fraction per site (λ) and DNA decay rate (k)*
418 The slope of log-transformed exponential decline of fragment length frequencies in aDNA (λ)
419 describes the probability of bond breaking in DNA backbone [46]. We estimated the DNA
420 decay rate per year (k) for *Hordeum* and *Oryza* from the slope of the linear relationship between
421 λ and sample age (figure 3c). We observed a per nucleotide decay rate of k= 2.64 x 10^{-4} per
422 year for *Hordeum* (R^2 = 0.208, p =3.27 x 10^{-12}, N= 211), which was 1.5 times faster than the
423 decay rate of *Oryza* of k= 1.79 x 10^{-4} per year (R^2 = 0.101, p = 3.65 x 10^{-7}, N= 245). The overall

424 decay rate for all herbarium samples was k= 2.08 x 10^{-4} per year (R^2 = 0.129, p = 2.52 x 10^{-15},
425 N= 456, supplementary figure S2c), which is slightly faster than the decay rate of k= 1.66 x 10^{-}
426 ^{4} per year observed in *Arabidopsis* and *Solanum* herbarium specimens [25], approximately 2.2
427 times slower than the k = 4.6 × 10^{-4} per year observed in dry-pinned arthropod museum
428 specimens [9], and nearly eight times faster than the rate of k= 2.71 x 10^{-5} observed for ancient
429 moa bones [13].
430
431 *Nucleotide misincorporations*
432 Deamination frequencies at 5' (C>T) and 3'(G>A) were highly correlated (R^2 = 0.951, p = 7.14
433 x 10^{-161}, N= 245 for *Oryza*, and R^2 = 0.989, p =1.18 x 10^{-12}, N= 211 for *Hordeum*;
434 supplementary figure S5a). This significant correlation remained largely unchanged even after
435 correcting deamination rates by subtracting baseline substitutions means (supplementary figure
436 S5b). All samples that passed filtering showed the expected "mirrored" deamination patterns
437 characteristic of aDNA (supplementary figure S5).
438 Both genera displayed statistically significant increases in the frequencies of 5' C>T
439 misincorporations at first position correlating with the age of the sample (figure 3d), with *Oryza*
440 starting from a higher baseline of damage when compared to *Hordeum* and displaying a
441 stronger relationship (R^2 = 0.303, p = 8.62 x 10^{-21}, N= 245 for *Oryza*, and R^2 = 0.207, p =3.63
442 x 10^{-12}, N= 211 for *Hordeum,* respectively). A slightly weaker yet significant relationship
443 between 5'C>T substitutions and sample age was also observed when analysing all samples
444 together (R^2 = 0.106, p = 1.11 x 10^{-12}, N= 456; supplementary figure S1d). In stark contrast to
445 the significant relationship between 5' C>T misincorporation and age, non-deamination
446 substitution rates showed no significant relationship with sample age (supplementary figure
447 S6), indicating that the patterns observed are specific to *post-mortem* deamination rates.
448

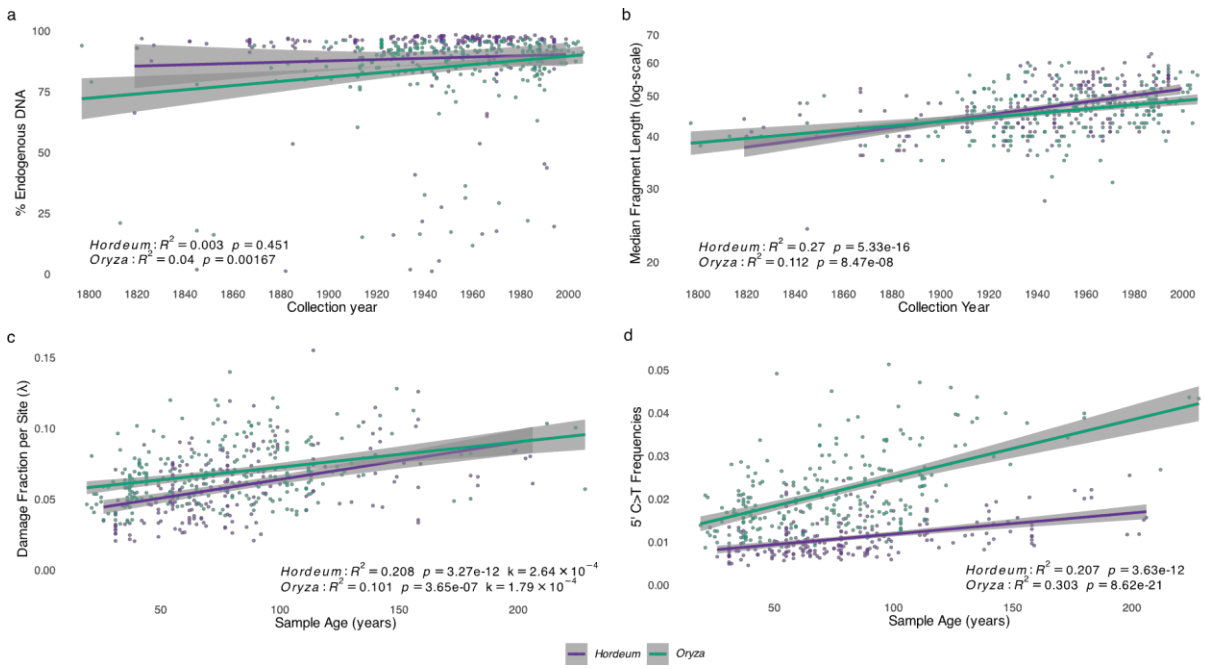


Figure 3: Regression analyses of aDNA damage metrics for *Hordeum* and *Oryza*: (a) Endogenous DNA content as a function of collection year. (b) Median fragment length of merged reads as a function of collection year, with log-scaled y-axis to show exponential relationship. (c) Damage fraction per site (λ) as a function of sample age, with the slope of regression corresponding to the DNA decay rate per base per year for *Hordeum* ($k = 2.64 \times 10^{-4}$) and *Oryza* ($k = 1.79 \times 10^{-4}$). (d) Frequencies of C>T misincorporations at first base (5' -end) as a function of sample age. Insets show regression statistics for each aDNA damage metric for each genus.

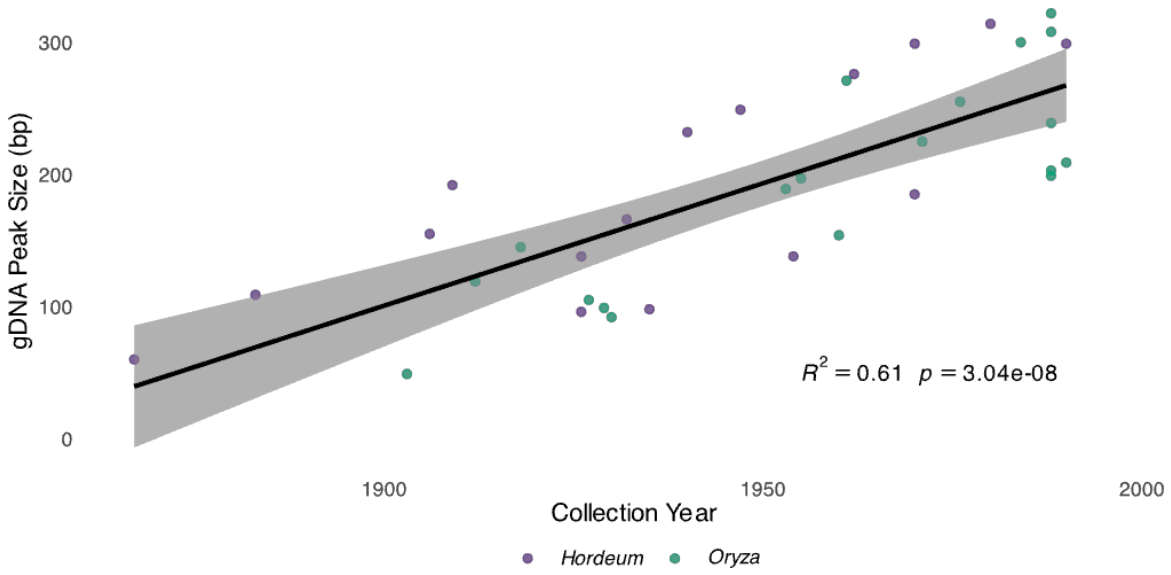


Figure 4: Regression between peaks of genomic DNA (gDNA) fragment size distribution obtained from high resolution fractionation electrophoresis (TapeStation) and collection year for a subset of samples ($N = 40$; 10 *H. spontaneum*, 10 *H. vulgare*, 10 *O. rufipogon*, and 10 *O. latifolia*;) after outlier removal. Inset shows regression statistics.

Differences in rates of damage for *Hordeum* and *Oryza*

We compared differences in the regression slopes and intercepts for all aDNA damage metrics between the genera *Hordeum* and *Oryza* with an analysis of covariance (ANCOVA) and visualised this comparison with boxplots.

Endogenous content

The analysis of covariance revealed significant effects of both sample age (*Pr*(Sample age) = 0.0114, $N = 456$) and genus (*Pr*(Genus) = 0.0204, $N = 456$) on the fraction of endogenous DNA content (figure 5a). No significant interaction was observed between sample age and genus (*Pr*(Sample age: Genus) = 0.176, $N = 456$), indicating that the rate of exogenous DNA colonisation over time does not differ significantly between the two genera and is not time dependent. This was further supported by ANOVA between the model including the interaction

and the model with no interaction, which confirmed that adding the interaction did not improve model fit ($F = 1.841$, $Pr = 0.175$). The model with no interaction displayed a significant fit ($F = 5.395$, $Pr = 0.004835$) but explained only a small proportion of the variance ($R^2 = 0.02327$), suggesting that factors beyond genus and sample age can substantially influence endogenous DNA content.

Fragment length

The analysis of covariance revealed significant effects of sample age and genus on fragment length (figure 5b) (Pr(Sample age) $< 3.45 \times 10^{-21}$; Pr(Genus) $= 0.0103$; $N = 456$). The interaction between sample age and genus was not significant (Pr(Sample age: Genus) $= 0.0893$; $N = 456$) and its inclusion did not significantly affect model fit ($F = 2.8993$, $Pr = 0.0904$). The model with no interaction explained 18% of the variance in fragment length ($R^2 = 0.1827$) and displayed a significant fit ($F = 50.65$, $Pr < 2.2 \times 10^{-16}$), indicating that the genera differ in their baseline fragment lengths, but not in their rate of fragmentation over time.

Damage fraction per site (λ) and DNA decay rate (k)

We observed a slower DNA decay rate of *Oryza* ($k = 1.79 \times 10^{-4}$ per nucleotide per year) than that of *Hordeum* ($k = 2.64 \times 10^{-4}$ per nucleotide per year). The analysis of covariance revealed significant effects of both sample age (Pr(Sample age) $= 7.48 \times 10^{-18}$; $N = 456$) and genus (Pr(Genus) $= 5.18 \times 10^{-8}$; $N = 456$) on the rates of bond breaking (figure 5c). However, no interaction between sample age and genus was observed (Pr(Sample age: Genus) $= 0.0884$; $N = 456$) and including the interaction did not significantly improve model fit ($F = 2.9152$, $Pr = 0.08843$). The model with no interaction explained 18% of the variance in lambda values ($R^2 = 0.1842$) and displayed a highly significant fit ($F = 51.14$, $Pr < 2.2 \times 10^{-16}$). Therefore, the

DNA decay rates (k) of *Hordeum* and *Oryza*, which correspond to the slopes of the regression between damage fraction per site (λ) and age, are not significantly different.

Nucleotide misincorporations

We observed a significant effect of both sample age (Pr(Sample age) = 3.66×10^{-6}; $N = 456$) and genus (Pr(Genus) = 3.14×10^{-4}; $N = 456$) on the rates of 5' C>T misincorporations (figure 5d). The interaction between sample age and genus was significant (Pr(Sample age: Genus) = 2.16×10^{-8}; $N = 456$), indicating that the rate of cytosine deamination over time differs substantially between the two genera. The model with interaction explained almost 55% of the variance in 5'C>T damage ($R^2 = 0.5448$) and displayed a significant fit ($F = 180$, $Pr < 2.2 \times 10^{-16}$). This was further corroborated by the ANOVA analysis, which confirmed that including the interaction significantly improved model fit ($F = 32.49$, $Pr = 2.165 \times 10^{-8}$). Therefore, not only do *Hordeum* and *Oryza* samples differ in their baseline levels of cytosine deamination, but they also accumulate this type of damage at significantly different rates over time, with *Oryza* showing a steeper increase in deamination with age.

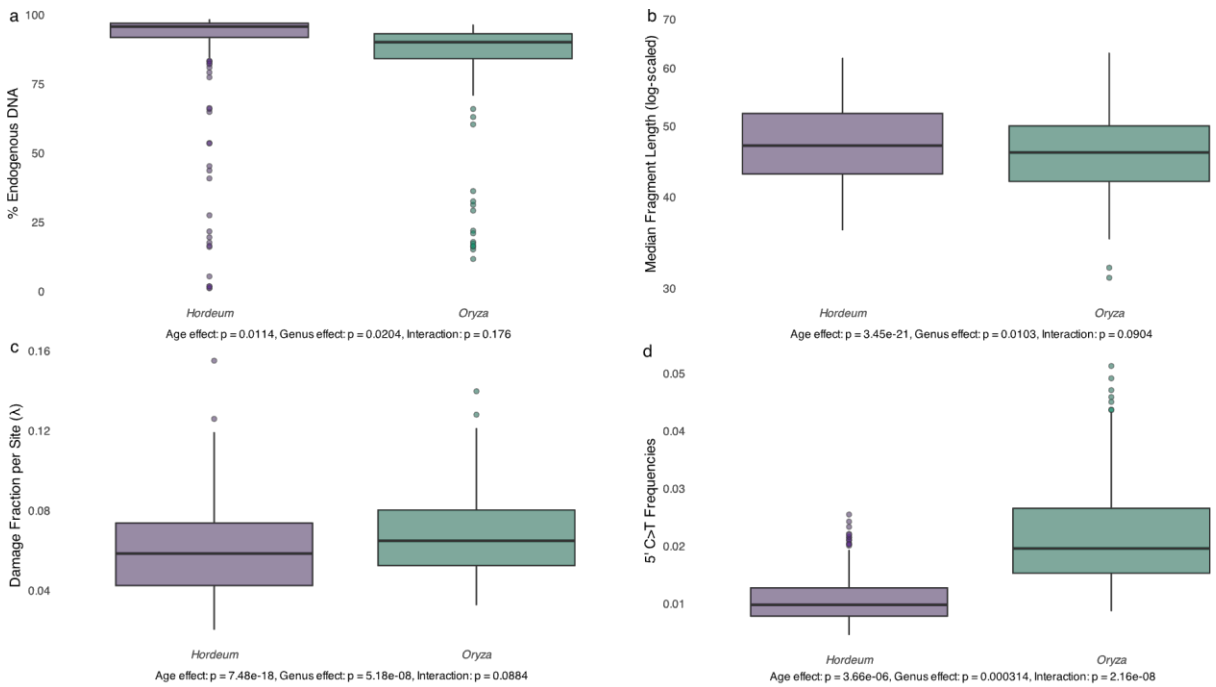


Figure 5: Analysis of Covariance (ANCOVA) of aDNA damage metrics for *Hordeum* and *Oryza*, with sample age as covariate and genus as factor. (a) Endogenous DNA content; (b) Fragment length (c) Damage fraction per site (λ); (d) 5' C>T damage frequencies. Significance values reported below each boxplot indicate the effects of age, genus and their interaction upon the analysed aDNA damage metric.

Effects of climatic variables upon rates of aDNA damage

We examined climatic effects using two complementary approaches: a "collection climate" model that assigned monthly temperature and precipitation values based on each sample's recorded collection location and month, and an "annual climate" model that used annual means and seasonality measures based solely on geographical origin. Both models aimed at capturing climate-driven DNA damage occurring during the initial post-collection period (field handling, drying, and early preservation), rather than during subsequent long-term storage. This assumption is supported by two lines of evidence: first, the rapid initial phase of DNA degradation driven by endogenous nucleases and proteases and exogenous microbial digestion [51,52], which occurs immediately after host death, is known to dominate post-mortem DNA damage [2,6,7,25]. Second, herbarium specimens are subsequently stored under standardised conditions that minimise further environmental variation [2,25]. We further tested this by including herbarium identity (as a proxy for institutional storage conditions) as a factor, and indeed found negligible explanatory power for herbarium identity, supporting the premise that collection-time climate, rather than storage climate, is the relevant environmental predictor of DNA damage patterns. Variance partitioning analysis revealed distinct patterns of environmental and temporal control over different aspects of DNA preservation (figure 6). The relative importance of factors varied substantially among damage metrics, with some showing strong environmental sensitivity while others were primarily controlled by age-related

processes. In addition to environmental factors, genus emerged as having significant explanatory power, possibly as a variable confounded with climate due to distinct distributions of the *Hordeum* and *Oryza* genera from temperate and tropical climates respectively (figure 2 and 6).

Endogenous content

Endogenous DNA content showed little association with the tested variables in the variance partitioning analysis (figure 6a). Only the collection climate model displayed statistical significance, but the fraction of variance explained by all analysed predictors was negligible (Adj. R^2 = 0.024, p = 0.007). The unique contributions of temperature and genus and the combined contributions of temperature, precipitation, age and genus showed marginal statistical significance ($p \leq 0.05$, supplementary table S3a). However, their poor performance in explaining the fraction of variance (Adj. R^2 ranging from 0.001 to 0.01) indicates that endogenous DNA fraction is largely determined by factors not captured by climatic variable at the time of specimen collection nor age. Removing genus from the variance partitioning analysis did not significantly alter the results (supplementary figure S7a, supplementary table S4a). Similarly, inclusion of herbarium as a proxy of institutional preservation practices and post-collection long term storage conditions also did not significantly alter the results (supplementary figure S8a, supplementary table S5a).

Fragment length

Our models had more explanatory power when it came to fragment length, with both collection and annual climate factors being statistically significant ($p \leq 0.001$) and performing similarly (figure 6b, Adj. R^2 = 0.227 and 0.228, respectively). Sample age was revealed as the largest contributing factor in both collection and annual models (Adj. R^2 = 0.199, 0.193 respectively),

consistent with temporal degradation processes. The shared variance between climatic variables and age were small, indicating that age effects on fragment size are largely independent of environmental variables. Indeed, whilst climatic variables were significant in both models (supplementary table S3b), their unique and combined variance explained were negligible (Adj. $R^2 \leq 0.01$). Removal of genus from the analysis marginally improved the unique and combined fraction of variance explained by temperature and precipitation in the collection and annual model respectively (supplementary figure S7b, supplementary table S4b), implying collinearity between genus and climatic variables, whereby further inclusion of herbarium had minimal effects (supplementary figure S8b, supplementary table S5b).

Damage fraction per site (λ)

Models predicting rates of DNA bond breaking (damage fraction per site, λ) from climatic and temporal variables explained up to 23% of the total variance in λ (figure 5c; Adj. $R^2 = 0.231$, 0.225 in the collection and annual models, respectively). Sample age contributed the largest unique fraction in both models (Adj. R^2 =: 0.170, 0.163), reflecting the fundamental relationship between specimen age and DNA backbone degradation. Temperature and precipitation showed minimal unique contributions and minimal shared variance with genus in the collection climate model (Adj. R^2 ranging from 0.01 to 0.03), and moderate shared variance with genus in the annual climate model (Adj. R^2 ranging from 0.03 to 0.04), suggesting indirect effects through correlations between genus and climatic variables. Indeed, removal of genus from the analysis did not strongly affect the overall variance explained by each model but inflated the unique and shared variance explained by temperature and precipitation (supplementary figure S7c).

Nucleotide misincorporations

The 5' C>T damage metric showed the highest explained variance among all damage metrics, with both collection and annual climate models being significant ($p \leq 0.001$) and explaining approximately 49% of the total variance (figure 6d). Whilst the unique contribution of age emerged as the strongest predictor of 5' damage (Adj. R^2 = 0.120, 0.119 in the collection and annual models, respectively), most explanatory power derived from shared effects among predictors rather than unique contributions. The combined effects of temperature and genus explained a substantial proportion of the total variance (Adj. R^2 = 0.135, 0.097). A similar pattern was observed for the combined effect of precipitation and genus (Adj. R^2 = 0.047, 0.008) and for the combined effects of temperature, precipitation and genus (Adj. R^2 = 0.096, 0.213), suggesting a strong correlation between climatic variables and genus. Removal of genus from the analysis (supplementary figure S7d) revealed temperature as the largest unique contributing factor in the collection model (Adj. R^2 = 0.134), followed by sample age (Adj. R^2 = 0.121) and precipitation (Adj. R^2 = 0.047), whilst the combined effect of temperature and precipitation explained almost 10% of the variance (R^2 = 0.096). Notably, the variance explained by the combined effect of temperature and precipitation was inflated in the annual climate model (Adj. R^2 = 0.214, supplementary figure S7d) and emerged as the largest contributor, followed by unique contribution of sample age (Adj. R^2 = 0.109) and temperature (Adj. R^2 = 0.096). The inclusion of herbarium did not improve the total fraction of variance explained in neither model (supplementary figure S8d), with the unique and shared fraction of explained variance by herbarium and other variables (age, climate and genus) being negligible (Adj. R^2 ranging from 0.0003 to 0.01).

Temperature effects on 5' C>T deamination frequencies

Given the predictive power of temperature in explaining 5' C>T damage patterns, we conducted post-hoc regression analyses to examine the relationships between temperature and cytosine

deamination. We observed a strong positive relationship between temperature and 5' C>T misincorporation frequencies (figure 7). The annual temperature model explained 18.4% of variance (R^2 = 0.184, p = 1.93 x 10^{-18}, N= 456, figure 7a), while the collection temperature model showed a slightly higher explanatory power (R^2 = 0.201, p = 3.74 x 10^{-20}, N= 456, figure 7b). However, this relationship was only observed when analysing the genera together and disappeared when analysing the genera on an individual basis (figure 7c, d).

Furthermore, whilst the relationship between non-deamination substitutions and temperature showed some significance, the explanatory power was negligible (R^2 = 0.007 - 0.07, supplementary figure S9), indicating that the temperature-damage relationship we reported are specific to deamination.

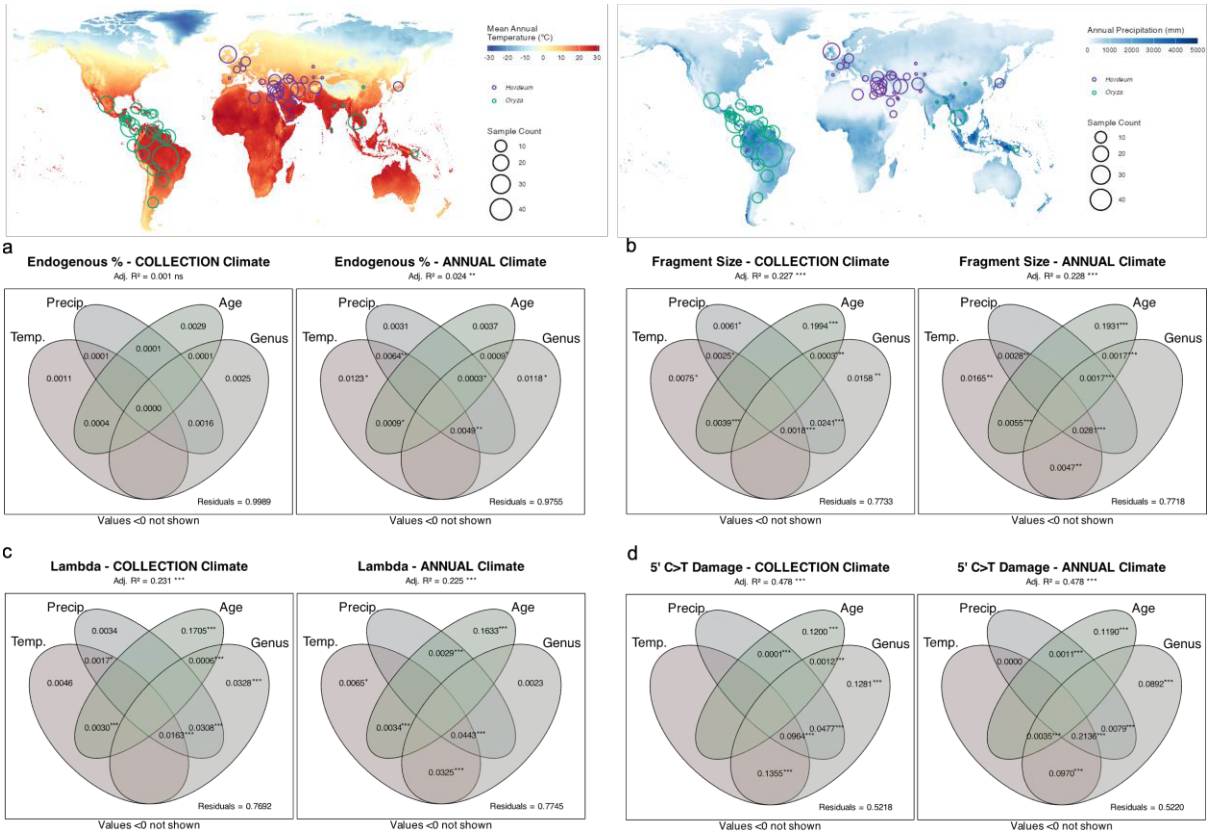
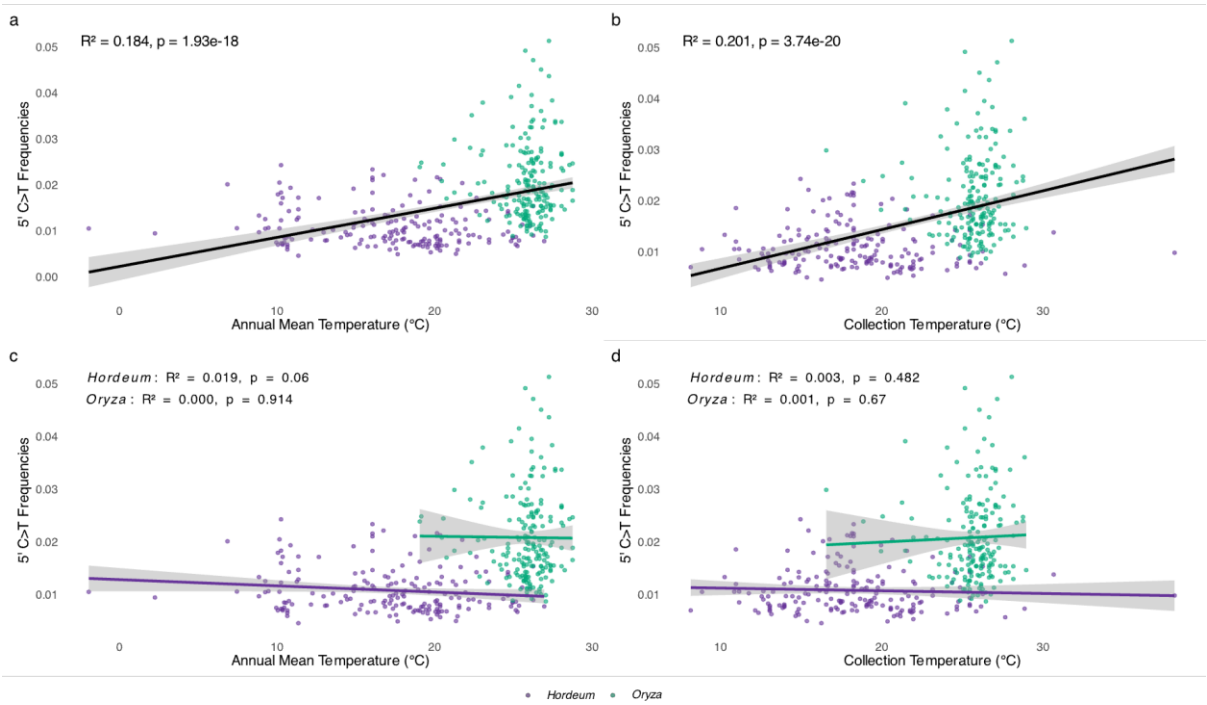


Figure 6: Climate influences on aDNA damage metrics in herbarium specimens. Maps (top panels) show the distribution of sampling locations for *Hordeum* and *Oryza* specimens overlaid on mean annual temperature (left) and annual precipitation (right) from the WorldClim climate

633 dataset [53]. Venn diagrams (bottom panels) display the unique and shared contributions of the
634 explanatory variables to the total variance in aDNA damage metrics.: (a) endogenous DNA
635 fraction, (b) fragment size, (c) damage fraction per site (lambda), and (d) 5' C>T substitution
636 frequencies at first base. Each metric is analysed using two models: collection climate (left)
637 and annual climate (right). Adjusted R^2 values for each model are shown above the plots.
638 Asterisks indicate statistical significance for the overall models and for each unique predictor
639 and combination of predictors (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$).
640



641

642 Figure 7: Relationship between temperature and deamination rates (5' C>T misincorporation
643 frequencies) in herbarium specimens for (a) annual mean temperature model, (b) collection
644 temperature model, (c) annual mean temperature model for *Hordeum* and *Oryza*, and (d)
645 collection temperature model for *Hordeum* and *Oryza*. Insets show regression statistics. The
646 significant relationships observed when analysing all samples together (a, b) are largely driven
647 by the contrasting climatic origins and damage levels between genera; these relationships are

not significant when analysing each genus separately (c, d), indicating that temperature and genus effects are confounded in this dataset.

Divergence-corrected nucleotide misincorporations analyses

All analyses of 5' C>T misincorporations were repeated using deamination frequencies corrected for reads-reference divergence. After correction, mean 5' C>T frequencies decreased from 1.74 ± 0.55% to 1.39 ± 0.50% across all samples (supplementary table S6). However, despite these corrections, all major patterns observed in the primary analyses remained largely unchanged. Corrected deamination frequencies still showed strong positive relationships with sample age in both genera (R^2 = 0.318, p = 6.17 x 10^{-22}, N= 245 for *Oryza*, and R^2 = 0.205, p =4.61 x 10^{-12}, N= 211 for *Hordeum,* respectively, supplementary figure S10), which were comparable to the uncorrected damage-age relationship observed in the primary analysis (R^2 = 0.303, p = 8.62 x 10^{-21}, N= 245 for *Oryza*, and R^2 = 0.207, p =3.63 x 10^{-12}, N= 211 for *Hordeum,* respectively, figure 3d). Variance partitioning analyses revealed age as the dominant factor explaining variation in corrected deamination rates (Adj. R^2 = 0.122, 0.156 in the collection and annual models, respectively), with the combined effects of temperature and genus contributing comparable fractions (Adj. R^2 = 0.110, 0.113, supplementary figure S11) to those obtained in the primary analyses (figure 6d). Similarly, the significant correlation between corrected deamination frequencies and temperature persisted for both annual (R^2 = 0.171, p = 2.59 x 10^{-18}, N= 456, supplementary figure S12a) and collection month models (R^2 = 0.173, p = 2.74 x 10^{-17}, N= 456, supplementary figure S12b) but disappeared when analysing the genera on an individual basis (supplementary figure S12c,d), which was also comparable to the results obtained in the primary analyses (figure 7).

Discussion:

Genomic inferences from preserved samples allows scientists to directly add temporal scales to the studies of evolutionary and ecological histories of species, and of fundamental questions about the mechanisms and tempo of evolution. The advent of high-throughput sequencing technologies fuelled the revolutionary growth of the field of archaeo- and paleo-genomics. In 2022, the number of ancient human genomes passed 10,000 with over 200 papers published [20]. The discovery of extremely well-preserved samples [54] and development of new approaches [55–58] have pushed the limits of DNA recovery and sequencing. For plants, owing to the great diversity of wild and cultivated taxa, it is not possible to know the exact number of sequenced preserved specimens. In recent publication of 'Plant Tree of Life' alone, genome-wide enrichment sequences have been generated for over 2,500 plant genera from herbarium specimens [59]. Current projects focusing on crops and their wild relatives routinely sequence whole genomes for hundreds of historical specimens [60–62]. RGB Kew is currently aiming to sequence genomes of 7,000 preserved specimens of fungi from its fungarium collection [63]. With increasing scale of aDNA research across all kingdoms, it is fundamentally important to improve our understanding of DNA preservation in historical and archaeological samples.

The focus of aDNA preservation research has been the impact of time on DNA degradation. Thanks to two well-researched examples, we indeed know that DNA deamination and depurination is correlated with sample age [13,25]. However, it has been noted that environmental conditions pre- and post-mortem should modulate the pace at which DNA is degraded [2]. Emphasis has been put on temperature and humidity, two parameters that have been suggested as reasons for difficulties in sequencing ancient genomes from the tropics [2]. Our work on herbarium specimens has allowed us to quantify the impact of age, environment and storage conditions on DNA preservation, revealing complex genus-specific differences extending beyond age-related degradation. Our regression analyses showed relationships

698 between sample age and different damage metrics, suggesting that while some aspects of aDNA
699 degradation follow predictable temporal trajectories, others are more heavily influenced by
700 environmental and biological factors.
701
702 **Age-dependent aDNA degradation patterns**
703 We further confirmed the highly fragmented nature of aDNA retrieved from herbarium
704 specimens, with the median fragment size of all analysed samples averaging 46.24 bp (SD =
705 6.97). This is consistent with previous studies on herbarium samples of a similar age range to
706 that of the current study [25] and on dry-pinned arthropods museum collections [9], but also
707 comparable to fragment sizes observed for animal remains that are several orders of magnitude
708 older, from a few hundred up to thousands of years old [5,13]. Possibly owing to lower levels
709 of environmental variation experienced by herbarium samples [25], we were able to detect a
710 weak yet significant relationship between median fragment length and collection year (figure
711 3b). For a subset of samples ($N = 40$), we directly measured gDNA fragment size distributions
712 prior to library preparation with high-resolution capillary electrophoresis (TapeStation),
713 allowing us to assess DNA fragmentation independent of any library construction artifacts or
714 bias. When examining the relationship between gDNA fragment size (TapeStation) and
715 collection year, we found an even stronger correlation with age (figure 4). This contrasts with
716 findings from animal bone studies, where no correlation between DNA fragmentation and
717 sample age has been observed [5,13,64], even when controlling for environmental variables
718 [2]. This fundamental difference can be explained by the two-step process of DNA decay and
719 degradation: a first rapid phase of enzymatic decomposition occurring immediately after host
720 death, driven by endogenous nucleases and proteases [51], and microbial digestion [52]; and a
721 second phase of chemical decomposition driven by hydrolytic and oxidative reactions,
722 occurring at much lower rates [6,7]. While the environment during sampling of herbarium

723 specimens is variable, the standardized preparation and storage procedures used in herbaria
724 reduce environmental variation during storage compared to the highly variable burial
725 conditions experienced by animal bones. Indeed, the source herbaria (a proxy for storage
726 conditions) have very limited explanatory power for DNA degradation (figure 6). The lower
727 levels of environmental variation experienced by herbarium specimens allows detection of the
728 underlying temporal fragmentation process that occurs during the second phase of chemical
729 degradation [25,51]. In archaeological contexts, the influence of variable environmental,
730 physical and chemical conditions, as well as tissue types and sample excavation and storage
731 [2,13], may mask more subtle age-related fragmentation patterns that become detectable under
732 the controlled conditions of herbarium storage [25].
733
734 The distinct preservation pathways of herbarium specimens compared to bones and museum
735 specimens also likely reflect different balances between hydrolytic and oxidative DNA
736 degradation processes. Herbarium specimens are often dried with heat upon collection [43,65].
737 Whilst this rapid desiccation can curtail hydrolytic damage, it also increases rates of oxidative
738 processes [6,43,65]. In contrast, animal bones or museum specimens such as pinned arthropods
739 dry more slowly, either through gradual burial diagenesis or museum preparation. These
740 mechanistic differences can contribute to the variation in DNA decay rates observed across
741 different biological materials: our herbarium samples exhibited decay rates ($k = 2.08 \times 10^{-4}$)
742 nearly eight times faster than that observed in bones ($k = 2.71 \times 10^{-5}$;[13]), suggesting a higher
743 susceptibility of *post-mortem* enzymatic and chemical DNA damage in herbarium samples
744 compared to bones [25]. Interestingly, arthropod museum specimens show even faster decay
745 ($k = 4.6 \times 10^{-4}$,[9]), approximately twice the rate observed in our herbarium samples. This
746 variation likely reflects differences in tissue composition (cellulose/lignin in plants,
747 chitin/protein in arthropods and hydroxyapatite/collagen in bones), storage methods (pressed

748 herbarium sheets, pinned insects, and buried remains), and the distinct degradation dynamics
749 imposed by rapid versus gradual desiccation. The intermediate position of herbarium
750 specimens between museum arthropods collections and ancient bones in terms of decay rate
751 suggests that DNA fragmentation is influenced by both intrinsic tissue properties and storage
752 conditions.
753 Furthermore, the 5' C>T damage patterns also showed significant correlations with age in both
754 genera (figure 3d). However, while both genera show similar rates of DNA fragmentation, their
755 divergent responses to chemical modifications suggest different susceptibilities to oxidative
756 damage and deamination. The higher baseline damage and steeper accumulation rates in *Oryza*
757 species indicate that the genus is more susceptible to *post-mortem* deamination, possibly due
758 them growing in different environments.
759 While aDNA damage metrics such as fragmentation and deamination were observed to be time-
760 dependent processes, the fraction of endogenous DNA content showed no correlation with age
761 (figure 3a), suggesting that microbial colonization and subsequent displacement of endogenous
762 DNA is largely independent of specimen age. Biological degradation processes are thus
763 primarily driven by factors other than time, such as *ante-* and *post-mortem* microbial
764 colonisation, tissue characteristics, and individual handling and storage conditions [66].
765 However, whilst we used species-specific reference genomes, increasing evolutionary distance
766 between query and reference might have impacted this metric [9]. We therefore suggest caution
767 when interpreting such results.
768
769 **Environmental controls on aDNA preservation**
770 Our variance partitioning analysis shows that climatic effects on aDNA damage are largely
771 mediated through complex interactions among predictors rather than independent
772 contributions. The dominance of shared variance fractions indicates that temperature,

773 precipitation and genus effects are highly correlated in our dataset, reflecting the geographic
774 and temporal sampling patterns of herbarium collections.
775
776 For metrics reflecting DNA backbone integrity (fragment size and lambda), age consistently
777 contributed the largest unique variance fractions, confirming that temporal degradation
778 processes are the primary drivers of DNA fragmentation. The relatively large unique age
779 contributions suggest that these physical degradation processes proceed independently of
780 environmental and taxonomic factors once specimens enter standardized herbarium
781 preparation methods storage practices. This age-dependent fragmentation appears to be a
782 distinctive feature of museum specimens (such as herbarium specimens [25] and dry-pinned
783 arthropods collections [9]) that contrasts with the patterns observed in archaeological animal
784 bones [5,13]. For the latter, variable burial or storage conditions introduce environmental
785 heterogeneity that can mask the underlying temporal degradation signal, whereas the
786 standardised preparation and storage practices in museum specimens reduce such variation,
787 allowing age-related fragmentation patterns to emerge.
788
789 The strong correlation between temperature and 5' C>T damage provides mechanistic insight
790 into why *Oryza* specimens, predominantly from tropical and sub-tropical regions, consistently
791 show higher damage levels compared to temperate *Hordeum* specimens (figure 7). Elevated
792 temperatures accelerate the spontaneous deamination of cytosine residues, explaining both the
793 higher baseline damage observed in *Oryza* and the overall temperature-damage relationship
794 observed across all herbarium samples. Precipitation variables as a proxy of humidity also
795 showed significant effects, indicating that deamination is a time-dependent process modulated
796 by temperature and humidity [2]. Overall, the complex interplay between temperature,
797 humidity and their seasonality and damage metrics suggests that specimens from regions with

less seasonal climates may experience different preservation trajectories than those from highly seasonal climates. The substantial unique genus contributions to 5' C>T damage confirm differential susceptibility to cytosine deamination between *Hordeum* and *Oryza*. However, the large shared variance fractions between genus and temperature reflect confounding effect of the contrasting geographic origins of these genera (temperate vs. tropical; figure 6). This geographic segregation limits our ability to fully disentangle genus-specific susceptibility from temperature-driven effects upon DNA damage. When examining temperature-damage relationships within each genus separately (figure 7c, d), the correlations become non-significant. It is possible that genus captures other environmental factors that were not included in temperature and precipitation. Precipitation is a decent proxy for humidity, but there are multiple other factors influencing it, and high humidity is expected to accelerate the deamination process. Alternatively, the differences observed due to tropical and temperate samples (captured in genus) could be explained by different sample processing approaches in the two areas, with tropical specimens being more often oven-dried/baked. Indeed, baking has been shown to substantially affect DNA degradation [65]. Additionally, in the tropics, alcohol treatment used to be common to prevent moulding, and this has been shown to limit the success in DNA amplification, presumably due to rapid degradation [67]. Future studies incorporating additional plant families with overlapping geographic distributions would be necessary to clearly separate genus-specific susceptibility from environmental temperature effects. Nevertheless, the consistent genus-level difference we observe remains informative about the challenges of DNA preservation across different biogeographic contexts.

The negligible variance explained for endogenous fraction across all models supports our hypothesis that microbial colonization is largely independent of the measured environmental and biological predictors. This finding suggests that endogenous DNA loss is driven by

822 stochastic factors such as initial contamination loads, handling procedures, or unmeasured
823 specimen-specific characteristics.
824
825 **Conclusions and implications for herbarium curation and ancient DNA research**
826 We show that ancient DNA damage patterns result from complex interactions between temporal
827 degradation processes, environmental conditions during sampling, biological properties and
828 storage condition. Herbarium specimens exhibit age-dependent DNA fragmentation patterns
829 that are not observed in animal bone studies, indicating that standardized preservation
830 conditions can reveal underlying temporal degradation processes masked by environmental
831 variation in archaeological contexts. We identified significant genus-specific differences in
832 aDNA damage susceptibility and established temperature as the dominant environmental driver
833 of cytosine deamination.
834 Environmental effects on DNA damage operate primarily through complex interactions with
835 taxonomic and temporal factors rather than direct independent contributions. Age-related
836 degradation dominates physical DNA breakdown, while genus-specific differences in cytosine
837 deamination susceptibility reflect both environmental adaptation and intrinsic biochemical
838 properties. The comparison across biological materials reveals that tissue composition
839 fundamentally constrains preservation potential, while storage conditions modulate decay
840 rates.
841 These findings highlight the importance of considering specimen age, climatic origin and
842 storage conditions when selecting herbarium specimens for ancient DNA analyses. The
843 mechanisms underlying genus-specific differences in aDNA damage accumulation warrant
844 detailed biochemical investigation. The significant, although small effect of herbarium-specific
845 factors on damage metrics indicates that institutional preservation practices vary and represent
846 an underexplored factor in ancient DNA research. Future studies should provide insights into

847 the biological processes driving DNA degradation, including detailed microbiome analyses.
848 Understanding which microbial taxa are most problematic for DNA preservation and how their
849 colonization is influenced by specimen characteristics could inform targeted preservation
850 strategies. As the scale of ancient DNA research continues to expand, such findings will be
851 essential for maximizing the scientific value of the world's natural collections and informing
852 evidence-based approaches to specimen sampling and preservation.
853
854 **Availability of code and requirements:**
855 Project name: Herbaria aDNA Damage
856 Project home page: https://github.com/Stefano-Porrelli/Herbaria_aDNA_Damage
857 Operating system(s): Linux/Unix (tested on SLURM-based HPC systems)
858 Programming language: Bash, R (version ≥4.0)
859 Other requirements:
860 - Conda/Miniconda,
861 - AdapterRemoval (≥2.0), BWA (≥0.7.17), FastQC (≥0.11), SAMtools (≥1.10), DeDup
862 (≥0.12), mapDamage2 (≥2.0), MultiQC (≥1.8), Preseq (≥2.0), AMBER.
863 - R packages: dplyr, tidyr, purrr, stringr, readr, MASS, vegan, car, ggplot2, colorspace,
864 viridis, ggrepel, ggtext, ggpubr, gridExtra, cowplot, sf, geodata, terra, maps
865 - External data: CHELSA V2.1 climate data (https://chelsa-climate.org/)
866 License: MIT licence
867 Any restrictions to use by non-academics: None
868
869 **Data availability:**
870 Raw FASTQ DNA sequences of all samples processed at the Royal Botanic Gardens, Kew
871 (UK) are deposited on the sequence reads archive (SRA) on NCBI: *Hordeum vulgare*

sequences are deposited under BioProject PRJNA1288534. *Hordeum spontaneum* sequences are deposited under BioProject PRJNA1289164. *Oryza rufipogon* sequences are deposited under BioProject PRJNA1288425. *Oryza grandiglumis* sequences are deposited under BioProject PRJNA1288424. *Oryza latifolia* sequences are deposited under BioProject PRJNA1288423. All remaining raw FASTQ DNA sequences of *Oryza* are deposited on the SRA under BioProject PRJNA1302186. Scripts and dataset to reproduce the analyses are available at https://github.com/Stefano-Porrelli/Herbaria_aDNA_Damage.

Authors' contributions:

S.P. and R.M.G. conceived the project and designed the study; R.M.G. and S.P. designed the sampling strategy, with contributions from A.F. and P.H.L.; P.J.K and N.S. overseen and supervised the generation of sequencing data for *Hordeum* samples; S.P. sampled herbaria and performed the aDNA laboratory work and bioinformatic screening for the *Hordeum* samples processed at RBGK; A.H. sequenced the *Hordeum* libraries at IPK; P.H.L. sampled herbarium samples and performed the aDNA laboratory work and bioinformatic screening for the *Oryza* samples processed at RBGK; A.F. selected herbarium specimens, A.F., M.N.R., and R.A.W. performed the sampling, A.F., W.Y., N.M., M.N.R., and A.C.C. isolated aDNA and prepared aDNA libraries, and A.F. performed screening and validation for the *Oryza* samples processed at KAUST/UoN; S.P. analysed the historical data and interpreted the results with the contribution of R.M.G. and P.H.L; S.P. and R.M.G. wrote the manuscript with contributions from all authors.

Acknowledgements:

We thank curators for providing herbarium specimens: Anna Haigh and Sue Zmartzy (Royal Botanic Gardens, Kew); Paul Peterson and Robert Soreng (Smithsonian Institute Herbarium);

Jordan K. Teisher (Missouri Botanic Gardens); Miriam Gaudeul (National History Museum, France); Michaela Schmull and Anthony R. Brach (Harvard Herbarium); Matthew C. Pace (New York Botanic Gardens). We acknowledge the expert technical assistance of Ines Walde and Jacqueline Pohl during DNA sequencing of *Hordeum* samples at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK). We also thank Anne Fiebig for her expert technical assistance with data submission of *Hordeum* low-throughput SRA sequences at IPK.

Funding:

This work was supported by the European Union Horizon 2020 research and innovation programme under grant agreement No. 862613 (AGENT - Activated GEnebank NeTwork), which funded the processing and low-coverage sequencing of *Hordeum* samples at the Royal Botanic Gardens, Kew and the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK). UK Research and Innovation grant EP/X022404/1 funded the processing and low-coverage sequencing of *Oryza* samples at the Royal Botanic Gardens, Kew, while King Abdullah University of Science and Technology (KAUST) grant ORA-CRG10-2021-4734 to R.A.W. supported the processing and low-coverage sequencing of *Oryza* samples at the University of Nottingham.

Ethics and permissions:

Historical herbarium specimens were obtained from established herbaria following institutional specimen access protocols. All sampling was conducted under standard museum loan agreements and followed institutional guidelines for destructive sampling of herbarium material. No additional ethical permissions were required as the study involved only preserved specimens collected under historical botanical collecting practices prior to CBD, but authors adhere to the ethical principles for Access and Benefit Sharing.

Conflict of interest:

The authors declare no conflicts of interest.

References:

1. Kristiansen K. Towards a new paradigm? The third science revolution and its possible consequences in archaeology. *Curr Swed Archaeol*. Svenska Arkeologiska Samfundet; 2021; doi: 10.37718/csa.2014.01.
2. Kistler L, Ware R, Smith O, Collins M, Allaby RG. A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic Acids Res*. 2017; doi: 10.1093/nar/gkx361.
3. Orlando L, Allaby R, Skoglund P, Der Sarkissian C, Stockhammer PW, Ávila-Arcos MC, et al.. Ancient DNA analysis. *Nat Rev Methods Primers*. Springer Science and Business Media LLC; 2021; doi: 10.1038/s43586-020-00011-0.
4. Dabney J, Meyer M, Pääbo S. Ancient DNA damage. *Cold Spring Harb Perspect Biol*. Cold Spring Harbor Laboratory; 2013; doi: 10.1101/cshperspect.a012567.
5. Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One*. Public Library of Science (PLoS); 2012; doi: 10.1371/journal.pone.0034131.
6. Lindahl T, Nyberg B. Rate of depurination of native deoxyribonucleic acid. *Biochemistry*. American Chemical Society (ACS); 1972; doi: 10.1021/bi00769a018.
7. Lindahl T. Instability and decay of the primary structure of DNA. *Nature*. Springer Science and Business Media LLC; 1993; doi: 10.1038/362709a0.
8. Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüfer K, et al.. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A*. Proceedings of the National Academy of Sciences; 2007; doi: 10.1073/pnas.0704665104.

9. Mullin VE, Stephen W, Arce AN, Nash W, Raine C, Notton DG, et al.. First large-scale quantification study of DNA preservation in insects from natural history collections using genome-wide sequencing. *Methods Ecol Evol*. Wiley; 2023; doi: 10.1111/2041-210x.13945.

10. Lan T, Lindqvist C. Technical advances and challenges in genome-scale analysis of ancient DNA. *Population Genomics*. Cham: Springer International Publishing;

11. Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-Rasheid KAS, Willerslev E, et al.. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics*. Springer Science and Business Media LLC; 2012; doi: 10.1186/1471-2164-13-178.

12. Dolenz S, van der Valk T, Jin C, Oppenheimer J, Sharif MB, Orlando L, et al.. Unravelling reference bias in ancient DNA datasets. *Bioinformatics*. Oxford University Press (OUP); 2024; doi: 10.1093/bioinformatics/btae436.

13. Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, et al.. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc Biol Sci*. The Royal Society; 2012; doi: 10.1098/rspb.2012.1745.

14. Hofreiter M, Paijmans JLA, Goodchild H, Speller CF, Barlow A, Fortes GG, et al.. The future of ancient DNA: Technical advances and conceptual shifts. *Bioessays*. Wiley; 2015; doi: 10.1002/bies.201400160.

15. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*. Springer Nature; 2014; doi: 10.1186/s12859-014-0356-4.

16. Schubert M, Ermini L, Der Sarkissian C, Jónsson H, Ginolhac A, Schaefer R, et al.. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc*. Springer Science and Business Media LLC; 2014; doi: 10.1038/nprot.2014.063.

17. Peltzer A, Jäger G, Herbig A, Seitz A, Kniep C, Krause J, et al.. EAGER: efficient ancient genome reconstruction. *Genome Biol*. Springer Science and Business Media LLC; 2016; doi: 10.1186/s13059-016-0918-z.

18. Willerslev E, Cooper A. Ancient DNA. *Proc Biol Sci*. The Royal Society; 2005; doi: 10.1098/rspb.2004.2813.

19. Liu Y, Bennett EA, Fu Q. Evolving ancient DNA techniques and the future of human history. *Cell*. Elsevier BV; 2022; doi: 10.1016/j.cell.2022.06.009.

20. Mallick S, Micco A, Mah M, Ringbauer H, Lazaridis I, Olalde I, et al.. The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Sci Data*. Springer Science and Business Media LLC; 2024; doi: 10.1038/s41597-024-03031-7.

21. Frantz LAF, Bradley DG, Larson G, Orlando L. Animal domestication in the era of ancient genomics. *Nat Rev Genet*. Springer Science and Business Media LLC; 2020; doi: 10.1038/s41576-020-0225-0.

22. Peris D, Janssen K, Barthel HJ, Bierbaum G, Delclòs X, Peñalver E, et al.. DNA from resin-embedded organisms: Past, present and future. *PLoS One*. Public Library of Science (PLoS); 2020; doi: 10.1371/journal.pone.0239521.

23. Kistler L, Bieker VC, Martin MD, Pedersen MW, Ramos Madrigal J, Wales N. Ancient plant genomics in archaeology, herbaria, and the environment. *Annu Rev Plant Biol*. Annual Reviews; 2020; doi: 10.1146/annurev-arplant-081519-035837.

24. Estrada O, Breen J, Richards SM, Cooper A. Ancient plant DNA in the genomic era. *Nat Plants*. Springer Science and Business Media LLC; 2018; doi: 10.1038/s41477-018-0187-9.

25. Weiß CL, Schuenemann VJ, Devos J, Shirsekar G, Reiter E, Gould BA, et al.. Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. *R Soc Open Sci*. The Royal Society; 2016; doi: 10.1098/rsos.160239.

996 26. Thiers B: Index Herbariorum: a global directory of public herbaria and associated staff. The
997 William & Lynda Steere Herbarium, New York Botanical Gardens.
998 http://sweetgum.nybg.org/ih/ Accessed 2025 Oct 23.
999 27. Burbano HA, Gutaker RM. Ancient DNA genomics and the renaissance of herbaria.
1000 *Science*. 2023; doi: 10.1126/science.adi1180.
1001 28. Eckert L, Eckert I, Rahn O, So CP, Barrett RDH. Using herbarium collections to study
1002 genetic responses to global change. *New Phytol*. 2025; doi: 10.1111/nph.70454.
1003 29. Bieker VC, Martin MD. Implications and future prospects for evolutionary analyses of
1004 DNA in historical herbarium collections. *Bot Lett*. Informa UK Limited; 2018; doi:
1005 10.1080/23818107.2018.1458651.
1006 30. Bieker VC, Battlay P, Petersen B, Sun X, Wilson J, Brealey JC, et al.. Uncovering the
1007 genomic basis of an extraordinary plant invasion. *Sci Adv*. American Association for the
1008 Advancement of Science (AAAS); 2022; doi: 10.1126/sciadv.abo5115.
1009 31. Latorre SM, Lang PLM, Burbano HA, Gutaker RM. Isolation, library preparation, and
1010 bioinformatic analysis of historical and ancient plant DNA. *Curr Protoc Plant Biol*. Wiley;
1011 2020; doi: 10.1002/cppb.20121.
1012 32. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex
1013 sequencing on the Illumina platform. *Nucleic Acids Res*. Oxford University Press (OUP);
1014 2012; doi: 10.1093/nar/gkr771.
1015 33. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target
1016 capture and sequencing. *Cold Spring Harb Protoc*. Cold Spring Harbor Laboratory; 2010;
1017 doi: 10.1101/pdb.prot5448.
1018 34. Hofreiter M, Jaenicke V, Serre D, Haeseler A, Pääbo S. DNA sequences from multiple
1019 amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic*
1020 *Acids Res*. 2001; doi: 10.1093/NAR/29.23.4793.

1021 35. Gutaker RM, Reiter E, Furtwängler A, Schuenemann VJ, Burbano HA. Extraction of
1022 ultrashort DNA molecules from herbarium specimens. *Biotechniques*. 2017; doi:
1023 10.2144/000114517.
1024 36. Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Pääbo S. Removal of deaminated
1025 cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res*. Oxford
1026 University Press (OUP); 2010; doi: 10.1093/nar/gkp1163.
1027 37. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming,
1028 identification, and read merging. *BMC Res Notes*. Springer Nature; 2016; doi:
1029 10.1186/s13104-016-1900-2.
1030 38. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
1031 arXiv [q-bio.GN].
1032 39. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al.. Twelve years of
1033 SAMtools and BCFtools. *Gigascience*. Oxford University Press (OUP); 2021; doi:
1034 10.1093/gigascience/giab008.
1035 40. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. mapDamage2.0: fast
1036 approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*.
1037 Oxford University Press (OUP); 2013; doi: 10.1093/bioinformatics/btt193.
1038 41. Fornasiero A, Feng T, Al-Bader N, Alsantely A, Mussurova S, Hoang NV, et al.. Oryza
1039 genome evolution through a tetraploid lens. *Nat Genet*. Springer Science and Business
1040 Media LLC; 2025; doi: 10.1038/s41588-025-02183-5.
1041 42. . R: language environment statistical computing. *R Foundation Statistical Computing*.
1042 Vienna, Austria;
1043 43. Bakker FT. Herbarium Genomics: Plant Archival DNA Explored. *Paleogenomics,*
1044 *Population Genomics*. Cham: Springer International Publishing;

44. Turner FS. Assessment of insert sizes and adapter content in fastq data from NexteraXT libraries. *Front Genet*. Frontiers Media SA; 2014; doi: 10.3389/fgene.2014.00005.

45. Venables WN, Ripley BD, editors. Modern applied statistics with S-plus. New York, NY: Springer;

46. Deagle BE, Eveson JP, Jarman SN. Quantification of damage in DNA recovered from highly degraded samples--a case study on DNA in faeces. *Front Zool*. Springer Science and Business Media LLC; 2006; doi: 10.1186/1742-9994-3-11.

47. Gutaker RM, Burbano HA. Reinforcing plant evolutionary genomics using ancient DNA. *Curr Opin Plant Biol*. Elsevier BV; 2017; doi: 10.1016/j.pbi.2017.01.002.

48. Weiß CL, Dannemann M, Prüfer K, Burbano HA. Contesting the presence of wheat in the British Isles 8,000 years ago by assessing ancient DNA authenticity from low-coverage data. *Elife*. eLife Sciences Publications, Ltd; 2015; doi: 10.7554/eLife.10005.

49. Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, et al.. Climatologies at high resolution for the earth's land surface areas. *Sci Data*. Springer Science and Business Media LLC; 2017; doi: 10.1038/sdata.2017.122.

50. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al.. Vegan: community ecology package. *R package version 25-4*. 2019;

51. Molak M, Ho SYW. Evaluating the impact of post-mortem damage in ancient DNA: a theoretical approach. *J Mol Evol*. Springer Science and Business Media LLC; 2011; doi: 10.1007/s00239-011-9474-z.

52. Eglinton G, Logan GA. Molecular preservation. *Philos Trans R Soc Lond B Biol Sci*. The Royal Society; 1991; doi: 10.1098/rstb.1991.0081.

53. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol*. Wiley; 2017; doi: 10.1002/joc.5086.

1069 54. van der Valk T, Pečnerová P, Díez-Del-Molino D, Bergström A, Oppenheimer J, Hartmann
1070 S, et al.. Million-year-old DNA sheds light on the genomic history of mammoths. *Nature*.
1071 Springer Science and Business Media LLC; 2021; doi: 10.1038/s41586-021-03224-9.
1072 55. Dabney J, Knapp M, Glocke I, Gansauge M-T, Weihmann A, Nickel B, et al.. Complete
1073 mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from
1074 ultrashort DNA fragments. *Proc Natl Acad Sci U S A*. Proceedings of the National Academy
1075 of Sciences; 2013; doi: 10.1073/pnas.1314445110.
1076 56. Dabney J, Meyer M. Extraction of highly degraded DNA from ancient bones and teeth.
1077 *Methods in Molecular Biology*. New York, NY: Springer New York;
1078 57. Gansauge M-T, Meyer M. Single-stranded DNA library preparation for the sequencing of
1079 ancient or damaged DNA. *Nat Protoc*. Springer Science and Business Media LLC; 2013;
1080 doi: 10.1038/nprot.2013.038.
1081 58. Slon V, Hopfe C, Weiß CL, Mafessoni F, de la Rasilla M, Lalueza-Fox C, et al.. Neandertal
1082 and Denisovan DNA from Pleistocene sediments. *Science*. American Association for the
1083 Advancement of Science (AAAS); 2017; doi: 10.1126/science.aam9695.
1084 59. Zuntini AR, Carruthers T, Maurin O, Bailey PC, Leempoel K, Brewer GE, et al..
1085 Phylogenomics and the rise of the angiosperms. *Nature*. Springer Science and Business
1086 Media LLC; 2024; doi: 10.1038/s41586-024-07324-0.
1087 60. Kistler L, de Oliveira Freitas F, Gutaker RM, Maezumi SY, Ramos-Madrigal J, Simon MF,
1088 et al.. Historic manioc genomes illuminate maintenance of diversity under long-lived clonal
1089 cultivation. *Science*. American Association for the Advancement of Science (AAAS); 2025;
1090 doi: 10.1126/science.adq0018.
1091 61. Alsantely A, Gutaker R, Navarrete Rodríguez ME, Arrieta-Espinoza G, Fuchs EJ, Costa de
1092 Oliveira A, et al.. The International Oryza Map Alignment Project (IOMAP): the Americas-

1093 past achievements and future directions. *J Exp Bot*. Oxford University Press (OUP); 2023;
1094 doi: 10.1093/jxb/erac490.
1095 62. Kreiner JM, Latorre SM, Burbano HA, Stinchcombe JR, Otto SP, Weigel D, et al.. Rapid
1096 weed adaptation and range expansion in response to agriculture over the past two centuries.
1097 *Science*. American Association for the Advancement of Science (AAAS); 2022; doi:
1098 10.1126/science.abo7293.
1099 63. Varga T, Woods R, Pitsillides F, Hill R, Biketova AY, Llewellyn T, et al.. Whole genome
1100 sequencing of historical specimens from the world's largest fungal collection yields high-
1101 quality assemblies. *New Phytol*. Wiley; 2025; doi: 10.1111/nph.70472.
1102 64. Pääbo S. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic
1103 amplification. *Proc Natl Acad Sci U S A*. Proceedings of the National Academy of Sciences;
1104 1989; doi: 10.1073/pnas.86.6.1939.
1105 65. Staats M, Cuenca A, Richardson JE, Vrielink-van Ginkel R, Petersen G, Seberg O, et al..
1106 DNA damage in plant herbarium tissue. *PLoS One*. Public Library of Science (PLoS); 2011;
1107 doi: 10.1371/journal.pone.0028448.
1108 66. Bieker VC, Sánchez Barreiro F, Rasmussen JA, Brunier M, Wales N, Martin MD.
1109 Metagenomic analysis of historical herbarium specimens reveals a postmortem microbial
1110 community. *Mol Ecol Resour*. Wiley; 2020; doi: 10.1111/1755-0998.13174.
1111 67. Särkinen T, Staats M, Richardson JE, Cowan RS, Bakker FT. How to open the treasure
1112 chest? Optimising DNA extraction from herbarium specimens. *PLoS One*. Public Library of
1113 Science (PLoS); 2012; doi: 10.1371/journal.pone.0043808.