Herbarium DNA degradation – Falling to pieces non-randomly

Freek T. Bakker and Lia Hemerik[1]

Post-mortem damage in herbarium DNA, mostly from 18th and 19th century collections, and with specimens usually heat-treated for conservation, consists mainly of genome fragmentation (single- and double-stranded breaks) rather than miscoding lesions. With typical herbarium DNA fragment sizes encountered (20–200 bp) this easily leads to insert sizes in library construction being smaller than Illumina read lengths applied (i.e. 100–250 bp).

Using a previously-published series of 56 genome-skimmed herbarium DNA extracts representing 10 angiosperm families, overlapping read pairs were found to occur in roughly 80 % of all read pairs obtained. After merging such overlapping pairs, the resulting fragments and their length-distributions are considered to reflect actual DNA fragmentation. Similar to occurrence in ancient DNA, we found over-representation of purines at fragment-ends in herbarium material. Distributions of fragment lengths fit gamma rather than exponential distributions, without apparent correlation with specimen age. The observed gamma distributions would indicate higher-order degradation kinetics, implying multiple processes acting during degradation. Possibly, the genome skimming data used here, in which repetitive sequences or compartments are over-represented, has biased genomic fragment-length distributions and half-lives as compared to the non-repetitive fraction of plant genomes, but no data was available to test this hypothesis. Overall, our results imply that we cannot confirm whether a plant archival DNA half-life exists and what its rate would be.

Keywords
DNA degradation, Genomic fragmentation, Herbarium DNA, Plant aDNA

Address of the authors
Biosystematics Group and [1]Biometris, Department of Mathematical and Statistical Methods, Wageningen University & Research/The Netherlands

Contact
freek.bakker@wur.nl
lia.hemerik@wur.nl

Accepted
13. 5. 2023

Electronic supplementary material
The online version contains supplementary material available at https://eterna.unibas.ch/bauhinia/article/view/1378/1606

DOI
https://doi.org/10.12685/bauhinia.1378

Herbarium collections constitute an enormous repository of botanical (meta)data, centred around the specimens and often including a range of different kinds of evidence such as biosequences and genomes, chemical and isotope data, data on associated microbes, pathogens and, obviously, collection locality as well as data relevant to taxonomy. Herbarium collections can be considered important ‚instruments' for testing historical hypotheses (Bakker et al. 2020), such as species' response at the genetic level to global change (Lang et al. 2020), or reconstructing the domestication/evolutionary history of crops (e.g., Sebastian et al. 2010; Gutaker et al. 2019), or for making taxonomic decisions (e.g., Bebber et al. 2010; James et al. 2018), providing the specimens are well-accessible and post-mortem genomic damage (see below) can be overcome. There are an estimated 397 million herbarium specimens deposited in 3500 herbaria world-wide (Thiers 2022). Collectively, these represent a huge past collection effort, sometimes under severely sub-optimal conditions, and in many cases the actual localities may no longer exist.

Herbarium collections enable a time-series perspective in plant species' past ecology, phenotypes, pathogens, and demography (Bieker and Martin 2018). We refer to Bakker (2017, 2018) for an overview of studies including on organelle genomes from extinct species, historical pathogens, and shift to C4 photosyn-

thesis in grasses, and see Lang et al. (2020) for a great example in herbarium-based study of global change biology.

With regards to taxonomy, Bebber et al. (2010) estimated that around 70 000 new species are already in herbarium collections, „waiting to be described", which further indicates the relevance of herbarium genomics, as it is expected to expedite archival DNA barcoding (see Xu et al. 2015). In fact, we are currently at the dawn of a herbarium genomics era (Buerki and Baker 2015, Bieker and Martin 2018, Olofsson et al. 2016), and chances are high that a large body of plant archival genomic data is generated in the years to come. In museomics, plants take a special position as their cell walls probably offer increased protection against oxidative DNA damage (see Bakker 2018). Their nuclear genomes are usually much larger in size than those from animal or fungal genomes (Gregory et al. 2007) and contain many repeats, which can hamper genome sequence assembly. Plant genomics is therefore inherently challenging, be it from archival or fresh DNA, although plastomes have both a small size (around 160 kb) and exhibit extensive structural conservation across land plants (Wicke and Schneeweiss 2015), enabling straightforward (re)sequencing with genome skimming (Straub et al. 2012; Bakker et al. 2016). Nevertheless, getting full nuclear genomic sequences from herbarium DNA is still rare to date, although the approach by Hart et al. (2016) and Brewer et al. (2019), targeting 353 nuclear genes in DNA from a range of herbarium specimens, provides a highly promising alternative.

For a summary of published studies on extraction of herbarium DNA (e.g., Särkinen et al. 2012, Gutaker et al. 2017, Xu et al. 2015) and the occurrence of possible post-mortem damage (Staats et al. 2011, 2013) we refer to Bakker (2017, 2018). Post-mortem damage in herbarium DNA consists mainly of genome fragmentation (single- and double-stranded breaks) caused during specimen fixation by heating, and herbarium specimen age since fixation does not appear to play a role (Staats et al. 2011, based on comparison of living trees and their historic herbarium vouchers; Staats et al. 2013). Herbarium specimens are often dried with heat, which can typically be 60–70°C, causing living cells in the specimen to rupture quickly, releasing nucleases and other cellular enzymes (Gill and Tuteja 2010), as well as reactive oxygen species (ROS). Such physiological conditions resemble necrosis, and this cellular stress typically causes DNA to degrade randomly into smaller fragments, running as a smear on agarose gels (Reape et al. 2008; McCabe et al. 1997). Indeed, herbarium DNA is typically highly-degraded into low molecular weight fragments (Doyle and Dickson 1987; Pyle and Adams 1989; Harris 1993) and this genomic fractionation causes the number of PCR amplifiable template molecules to be reduced (Staats et al. 2011). Heating is known to cause de-purination and subsequent hydrolysis of the DNA sugar-phosphate backbone (Lindahl and Andersson 1972). This process is therefore expected to result in an excess of AG purines just before fragment break-points. In addition, de-

amination of C to U, which is read as T by the polymerases during sequencing, results in C→T transitions. Excess of AG purines has been observed to occur just before the ends of fragments in studies on ancient DNA (aDNA) (Briggs et al. 2007), suggesting they are the cause of breaks. The same study showed the occurrence of over-representation of CT at fragment ends, which can probably be attributed to oxidation of ‚loose' single strands at fragment endings. These typical aDNA patterns have also been observed in historic (heated) herbarium DNA (Weiss et al. 2016), suggesting that aDNA and (heated) herbarium DNA share the same sequence damage-characteristics. Therefore, the conclusion is probably fair that historic herbarium DNA from heated specimens looks rather similar to (non-heated) ancient DNA.

To what extent the C→T/G→A transitions at the ends of the reads drives post-mortem transitions in herbarium DNA sequences the authors do not mention, but possibly the post-mortem transitions reported by Staats et al. (2013) correspond to these.

Herbarium DNA fragmentation can occur sometimes to the extent that the efficiency of paired-end sequencing using Illumina HiSeq (and hence subsequent sequence assembly) is affected. When template insert sizes are shorter than twice the Illumina read lengths applied, the actual sequencing reads will ‚meet in the middle' of the insert and start to overlap (Fig. 1). However, when template insert sizes are smaller than the Illumina read length applied, this will result in the presence of adapter sequence at the end of the read (Turner 2014). In both scenarios of read-overlap, the two reads can be merged into a single, longer read. In a previous study we used a series of 93 herbarium DNA samples (some of which 146 years old), representing 10 angiosperm families (Bakker et al. 2016). Overlapping read pairs were found to occur in roughly 80 % of all read pairs obtained. After merging such overlapping pairs the resulting fragments and their distribution can be considered to reflect (the ongoing process of) genome fragmentation up to the moment of DNA extraction. Merging reads enables assessing the distribution of genomic fragment lengths

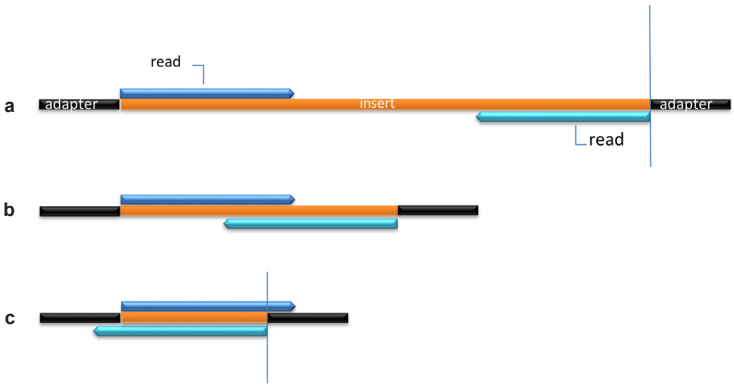

Fig. 1. Herbarium DNA fragments (orange) may be longer **a** or shorter (**b** and **c**) than Illumina read lengths, in which case reads are overlapping and could be merged. From Bakker (2018), with permission from Springer.

(fragment length distribution or FLD) in a herbarium DNA extract, as was carried out in Weiss et al. (2016). Here the authors inferred FLDs for a series of non-heated herbarium specimens of up to 300 years old, by merging overlapping reads as outlined above. By assuming a log-normal FLD the authors claimed they were able to deduce decay rates for their genomic extracts, based on the slope of „the exponential part" of the FLD. Weiss et al. (2016) conclude that the herbarium decay rate is „six times the rate of bone DNA decay".

Here, we infer the pattern of fragmentation and the FLD in degraded herbarium DNA, as measured from genome skimming data. With our data we confirm that the duration of being in the herbarium affects the average fragment lengths but that the FLDs appear not correlated with specimen age. We explore to what extent the data can be explained by de-purination and de-amination known to occur upon heating (and applied to the majority of specimens in herbarium collections today). We elucidate whether typical plant genomic fragment-length distributions can tell us whether genome degradation is a first- or higher-order kinetic process, i.e. whether a single or more processes are involved in fragmentation. A better understanding of such processes could help in assessing whether **1** a general herbarium DNA decay rate actually exists, and **2** what drives fragmentation in repetitive genomic compartments in plants.

Material and Methods

Herbarium genome fragment-merging

In order to explore the distribution of short-sized fragments in herbarium DNA and the extent to which read-overlap was occurring among them we re-analysed a subset of 56 Illumina genome skimming historic herbarium samples from Bakker et al. (2016) which included species of *Lactuca, Karelinia* and *Nicolasia* (Asteraceae), *Polyscias* (Araliaceae), *Pelargonium* (Geraniaceae), *Aethionema* and *Tarenaya* (Brassicaceae), *Anthochortus*, *Dovea* and *Hypodiscus* (Restionaceae), *Anaxagorea*, *Desmopsis* and *Monanthotaxis* (Annonaceae), *Hymenostegia* and *Duparquetia* (Fabaceae), *Begonia* (Begoniaceae), *Paphiopedilum* (Orchidaceae), and *Rinorea* (Violaceae). Most of these specimens were expected to have been fixated after collection, by heat treatment. This could be in ovens at around and perhaps up to 70°C in (local) herbarium facilities, or in situ using hot air from hair driers or camping cookers (Jan Wieringa, pers. comm). In addition, other techniques have been used for field drying, including kerosene stoves, 100-watt lightbulbs, and air-drying on a moving vehicle (Staats et al. 2011). Some of our historic specimens, especially from the wet-tropics, may have been subjected to ‚Schweinfurt' treatment (i.e. the temporary fixation with methylated spirits or 30 % formaldehyde) to prevent specimens from moulding, whilst underway to local herbarium facilities. As discussed in Bakker et al. (2016), clear

documentation for the fixing method actually applied to each herbarium specimen is usually lacking, and therefore no specific hypotheses regarding preservation method effects can be tested. Wet-tropical specimens gave lower N50 values (N50 is similar to a mean or median of assembled contig lengths, but with greater weight given to the longer contigs) and concomitantly higher number of contigs in plastome assemblies (Bakker et al. 2016).

We used BBMerge from the BBTools package (http://jgi.doe.gov/data-and-tools/bbtools) in order to check whether overlap exists between read-pairs (Fig. 1) and in case it was, reads were subsequently merged. When insert size is shorter than the read length (in this case 100 bp) reads will have adapter sequence at the tail end, which was removed by using BBMerge after merger (Brian Bushnell, pers. comm.). When insert size is the length of two read lengths, i.e. 200bp, reads cannot be merged anymore because there is no overlap. Using the default mode in BBMerge, the proportion of overlapping reads, as well as the average fragment length and its standard deviation were recorded. The FLD that resulted from merging the overlapping reads was plotted for each accession and in order to make the FLDs comparable we compared relative frequencies of read pairs. Fragment lengths were between 26 and 184 bp, which reflects the minimum and maximum fragment length given the adapters used.

In order to investigate a time-series of specimens, and hence whether older specimens yield higher fragmentation, we compared length distributions for two series of accessions from the Bakker et al. (2016) data, one for species of *Aethionema* (Brassicaceae), used and further described in Mohammadin et al. (2017) for phylogenetic analysis, and for *Lactuca* (Asteraceae), used and further described for the same purpose in Wei et al. (2017). The *Aethionema* series included both silica gel-dried and historic herbarium specimens of 23, 33, 40, 44, twice 50, 66 and 146 years old. The *Lactuca* series included silica gel-dried and historic herbarium specimens of 7, 36, 42, 43, twice 49, 54 and 64 years old. By comparing these congenerics it can be assumed that genome size, GC contents, specimen tissue characteristics, and specimen fixation histories (in most cases) are comparable too. Differences in FLD should therefore be due to specimen age, different specimen fixation (if applicable), herbarium collection locality or perhaps even stochasticity. Reads were merged using BBMerge as described above and fragment lengths between 26 and 184 bp plotted and their relative frequencies compared.

We used MapDamage 2.0 (Jónsson et al. 2013) in order to investigate over-representation of purines (A and G) at fragment endings, by mapping reads against a set of assembled contigs, enabling assessment of nucleotide positions around fragment-ends, summarised across all reads.

Fitting models to distributions of genomic fragment-lengths

If breakpoints in DNA fragments are randomly-distributed, one would in principle expect an exponential distribution of resulting fragments (simulation data, not shown). In order to investigate what model fits our observed genomic FLDs best we fitted an exponential distribution and a gamma distribution (Bolker 2008) to each of the 56 data sets. We estimated the rate parameter of the exponential, and the shape and scale parameters of the gamma distribution. It should be noted that the exponential function is a special case of the gamma distribution (i.e. when the shape parameter is 1). The AIC criterion (Bolker 2008) was used to select for each of the 56 datasets which distribution fitted best. The fitting procedure and model selection was performed in R version 4.2.1 (R Core Team 2022) using the fitdistr function of the R-package MASS (Venables and Ripley 2002).

Results

Herbarium DNA reads

Across the 56 samples, the average fragment length appeared to be negatively correlated with specimen age ($R^2 = 0.29$), which confirmed earlier studies indicating that older herbarium DNA extracts contain smaller fragments (Fig. 2a). We found the standard deviation of the average fragment lengths to increase with longer fragments (Fig. 2b; polynomial regression, $R^2 = 0.94$), i.e. short fragments were less length-variable and occurred in ‚peaks' within a fragment length distribution. In contrast, longer fragments occurred across broader size ranges. Apparently, genomic fragments ‚end up' in increasingly small, uniform, sizes, but in the same time, the smaller sizes are correlated with higher specimen age as seen above. The percentage of read pairs that can be merged appeared to be fairly independent of specimen age (Fig. 2c). The actual numbers of reads was lower in older specimens, but these yield lower amounts of reads in the first place (Bakker et al. 2016), therefore also lower amounts of read pairs that can be merged.

Fitting models to distributions of genomic fragment lengths

Fragment length distributions (FLDs) for a subset of all 56 accessions, representing the two time-series *(Aethionema* and *Lactuca),* are given in Fig. 3. (In addition, FLDs for another subset of 36 of the 56 accessions, are given in Suppelementary Fig. 1, with specimen age indicated by colour-coding.) There does not appear to be a correlation between FLD and specimen age. As outlined above, we fitted both exponential and gamma distributions to the 56 FLDs contained in our data. All distributions appeared to fit well to a gamma distribution: either there are many short fragments

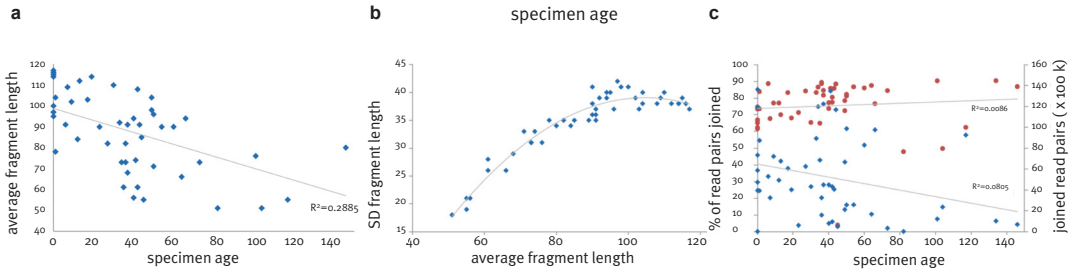

Fig. 2 a-c: Overlapping Illumina HiSeq reads from herbarium DNA extracts; the average fragment length after merging reads **a** plotted against specimen age; the SD of average fragment length **b** plotted against average fragment length; and **c** the percentage of total reads that could be merged (red dots) and the actual number of merged read pairs (blue diamonds). From Bakker (2018), with permission from Springer.

and few longer ones, or there is a gradual increase in longer fragments (Fig. 3 and Suppelementary Fig. 1; and see Suppelementary Appendix 1 for histograms of the 56 datasets with the fitted exponential and gamma distributions plotted). The values for rate, shape and scale parameters showed an average value of 0.0122, 8.1267 and 13.1067 resp., with associated standard deviation values of 0.0046, 6.0394 and 4.2897 (see also Suppelementary Appendix 2 for histograms of all values for these parameters in the 56 analysed datasets). For all 56 datasets, the gamma distribution had the minimum AIC value and Δ-AIC was on average 5 313 812 in a range of (11 267 to 15 712 447) as is shown in Suppelementary Appendix 3. The high standard deviations reflect the range of fragment-length distributions among the accessions included, and indicates that fragmentation dynamics differs across all accessions. It should be noted that some of the datasets are bimodal and that although the gamma distribution fits best, when choosing between exponential and gamma only, the gamma distribution is a bad choice for these datasets (see X002Paustr, X27mult, X21abro in Suppelementary Appendix 1). In our taxonomic sampling 14 accessions (i.e. from Annonaceae, Araliaceae, Fabaceae, Begoniaceae, Orchidaceae and Violaceae) had been collected in the wet-tropics. As outlined above and based on previous studies (Bakker et al. 2016), wet-tropical origin appeared to be the main factor correlating with plastome-assembly success, possibly due to difference in underlying genomic fragmentation patterns. These 14 accessions however did not seem to differ in rate, shape and scale parameters for their FLDs (indicated in green in Suppelementary Appendix 3).

Purine over-representation at fragment endings

Weiss et al. (2016) found over-representation in A and G (purines) towards fragment ends, a pattern that reflects what is encountered in ancient DNA (Briggs et al. 2007). Depurination, or loss of A and G bases, is known to be a first step towards dou-

ble-stranded breaks (Lindahl and Andersson, 1972). Therefore, it is expected for purines to be overrepresented towards fragment ends and some of the samples analysed here with Map Damage 2.0 indeed did show this pattern (Suppelementary Fig. 2), but predominantly in herbarium (not fresh) accessions.

Discussion

Herbarium genomics has seen great opportunities and development over the past decade, mainly driven by the ever-increasing availability of NGS technology. Especially when concerned with organelle genomes and other repetitive genomic compartments, approaches such as genome skimming appear effective in extracting DNA sequence data from large series of archival specimens (Straub et al. 2012; Bakker 2017). As a general feature of herbarium DNA, genomic fragment size can be small (25–300bp). Overlapping read pairs are the result of template insert size being smaller than twice the read length applied (or even smaller than the read length itself). Using a genomic skimming series of 56 herbarium DNA samples, representing 10 angiosperm families, overlapping read pairs were found to occur in roughly 80 % of all read pairs obtained for most samples. Fragmentation is therefore confirmed to occur across families, and insert sizes can be as small as <100 bp that still represent a majority of fragments. As outlined above, the distribution of herbarium DNA fragment-lengths could in principle inform us about biases or trends that may exist in the actual process (or processes) by which herbarium genomes break down.

Intuitively one would expect older specimens to be more fragmented than younger ones, given that more post-mortem time has been available. On the other hand, the experimental herbarium results by Staats et al. (2011), comparing fresh and century-old DNA from the same individuals of trees, indicated that this does not need to be the case (see above, and Fig. 3). We compared genomic fragment-length distributions (FLDs) for two series of herbarium samples (included in our set of 56 accessions), each from the same genera *(Lactuca* and *Aethionema)* and each also including non-historic (i.e. silica gel-dried) samples for comparison. For the *Lactuca* series, the oldest sample was indeed the most highly fragmented (Fig. 3a). For the *Aethionema* series however, the older specimens did not appear to have highest proportion of small fragments, but specimens around 50 years did (Fig. 3b). For both series, we saw that the silicagel-dried samples showed a gradual increase in occurrence of longer fragment lengths that would probably have extended beyond 200 bp, had current Illumina read lengths of 150bp or more been used.

Following our FLD model-fitting analysis we found the best-fitting models to be gamma distributions, as indicated by the AIC criterion used. As indicated above, the high standard deviations for the shape and scale parameters for the gamma distribution probably reflect the range of FLDs among the accessions included, and indicates that fragmentation dynamics apparently differs

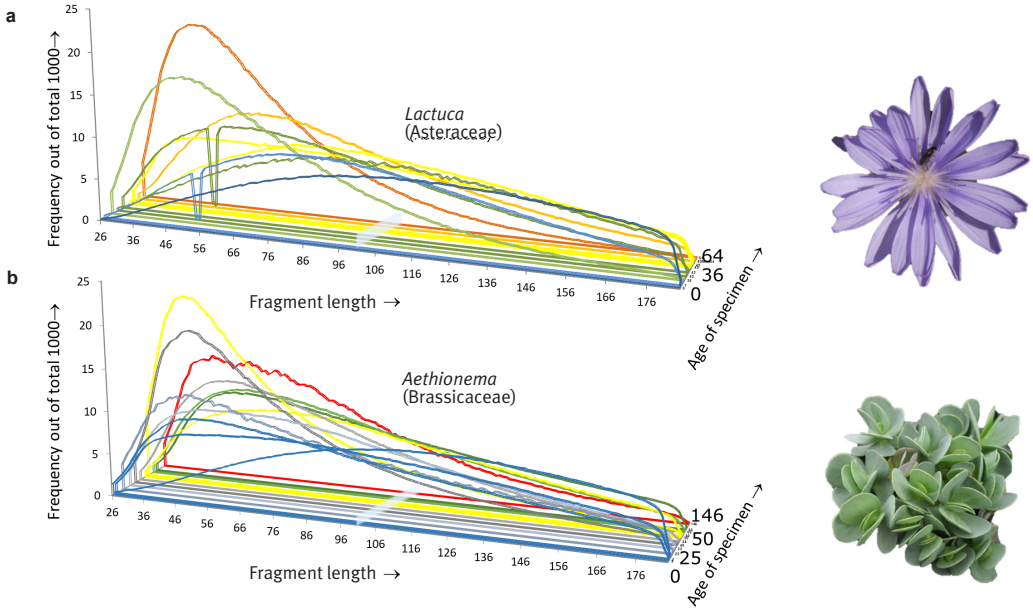

Fig. 3. Distributions of fragment lengths (in bp) from fresh and herbarium specimens of different ages of *Lactuca* a and *Aethionema* b, with distributions sorted by (increased) specimen age. The transparent bar indicates the read length used (100 bp). Fragments were produced after merging 100 bp Illumina reads, with reads up to 25 bp discarded, and reads with length <100 trimmed with regard to adapter sequences (see text). The distribution resembles a gamma function, with either a maximum of lengths around 30bp, or a wide length-range. From Bakker (2018), with permission from Springer.

across all accessions. Weiss et al. (2016) and Allentoft et al. (2012) suggested lognormal distributions of fragment lengths fit best in historic *Arabidopsis,* and bone DNA, respectively, and based this partly on the observed linear relation after loglog transformation. Yao et al. (2016) found the same for DNA degradation in human serum, urine, and saliva DNA. These distributions would be consistent with a first-order kinetics at which DNA degrades, i.e., DNA has a half-life and the rate of degradation is constant (Allentoft et al. 2012). However, Weiss et al. (2016; in their Fig. 2) appear to consider the „exponential decline" to start after the median of their genomic FLD. The first part of the distribution would then not be taken into account. Only considering the second half (> median) of the distribution leads indeed to an exponential distribution, as the authors emphasize, especially after log transforming the y-axis (Weiss et al. 2016). In contrast, we chose to include the entire fragment-length distribution and find that gamma distributions fit the distributions significantly better than exponential distributions. As the exponential is a special case of the gamma distribution this indicates a higher order kinetics underlying fragmentation in these data sets.

Looking directly into the herbarium break-points, by summarising nucleotide composition at fragment endings (Supplementary Fig. 2) indicates that there is over-representation of purines (A and G) in case of (heat-treated) herbarium DNA. This would imply that the distribution of purines in the herbarium genome would drive the FLDs observed, and hence that purines are gamma distributed in the genome (which would probably be unrealistic). However, our data was generated using genome skimming, which means that repetitive compartments and sequences are probably overrepresented in all samples. To what extent such regions are non-representative of general genome composition and complexity is difficult to say. Possibly the repeats themselves may contain purine biases but no published studies indicating this exist to date. If the herbarium DNA degradation investigated here indeed fits a gamma rather than a lognormal or exponential distribution, this would indicate either a non-constant rate of degradation, or decay consistent with a higher-order kinetics, differing from the usually-observed first-order genomic degradation kinetics. In the latter, break-points are randomly distributed in DNA sequences and therefore would be expected to yield exponential FLDs. Our data as used in this study is genome skimming data, derived from genomic repetitive regions. Possibly, degradation of repetitive genomic compartments occurs at higher-order kinetics, i.e., a different half-life is present compared with non-repetitive DNA. However, this would need to be tested with (ancient) genomic samples that are deep-sequenced rather than genome-skimmed.

Acknowledgements

We thank the organisers of Bauhin2022 Conference for inviting FTB to the meeting, and enabling contribution to these Proceedings.

References

Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, Campos PF, Samaniego JA, Gilbert MTP, Willerslev E, Zhang G, Scofield RP, Holdaway RN & Bunce M (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proceedings of the Royal Society B 279: 4724–4733 https://doi.org/10.1098/rspb.2012.1745

Bakker FT (2017) Herbarium genomics: skimming and plastomics from archival specimens. Webbia: Journal of Plant Taxonomy and Geography 72: 35–45 doi.org/10.1080/00837792.2017.1313383

Bakker FT (2018) Herbarium genomics: plant archival DNA explored. In: Lindqvist C & Rajora OP (eds) Paleogenomics, Population Genomics. Springer, Cham. https://doi.org/10.1007/13836_2018_40

Bakker FT, Antonelli A, Clarke JA, Cook JA, Edwards SV, Ericson PGP, Faurby S, Ferrand N, Gelang M, Gillespie RG, Irestedt M, Lundin K, Larsson E, Matos-Maraví P, Müller J, Von Proschwitz T, Roderick GK, Schliep A, Wahlberg N, Wiedenhoeft J & Källersjö M (2020) The Global Museum: natural history collections and the future of evolutionary science and public education. PeerJ 8: e8225 https://doi.org/10.7717/peerj.8225

Bakker FT, Lei D, Yu J, Mohammadin S, Wei Z, Van De Kerke S, Gravendeel B, Nieuwenhuis M, Staats M, Alquezar-Planas DE & Holmer R (2016) Herbarium genomics: Plastome sequence assembly from a range of herbarium specimens using an terative organelle genome assembly (IO-

GA) pipeline. Biological Journal of the Linnean Society 117: 33–43 http://dx.doi.org/10.1111/bij.12642

Bebber DP, Carine MA, Wood JRI, Wortley AH, Harris DJ, Prance GT, Davidse G, Paige J, Pennington TD, Robson NKB & Scotland RW (2010) Herbaria are a major frontier for species discovery. Proceedings of the National Academy of Sciences 107: 22169–22171

Bieker VC & Martin MD (2018) Implications and future prospects for evolutionary analyses of DNA in historical herbarium collections. Botany Letters, 165: 409–418 https://doi.org/10.1080/23818107.2018.1458651

Bolker BM (2008) Ecological Models and Data in R. Princeton University Press, Princeton, NJ, USA

Brewer GE, Clarkson JJ, Maurin O, Zuntini AR, Barber V, Bellot S, Biggs N, Cowan RS, Davies NMJ, Dodsworth S, Edwards SL, Eiserhardt WL, Epitawalage N, Frisby S, Grall A, Kersey PJ, Pokorny L, Leitch IJ, Forest F And Baker WJ (2019) Factors Affecting Targeted Sequencing of 353 Nuclear Genes From Herbarium Specimens Spanning the Diversity of Angiosperms. Frontiers in Plant Science 10: 1102 doi:10.3389/fpls.2019.01102

Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prufer K, Meyer M, Krause J, Ronan MT, Lachmann M, Pääbo S (2007) Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Academy of Sciences 104: 14616–14621 doi:10.1073/pnas.0704665104

Buerki S & Baker WJ (2015) Collections-based research in the genomic era. Biological Journal of the Linnean Society 117: 1–5 doi: 10.1111/bij.12721

Doyle JJ & Dickson EE (1987) Preservation of plant species for DNA restriction endonuclease analysis. Taxon 36: 715–722

Gill SS & Tuteja N (2010) Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. Plant Physiology and Biochemistry 48: 909–930

Gregory TR, Nicoll JA, Tamm H, Kullman B, Kullman K, Leitch IJ, Murray BG, Kapraun DF, Greilhuber J & Bennett MD (2007) Eukaryotic genome size databases. Nucleic Acids Research 35(Database issue): D332–D338 doi:10.1093/nar/gkl828

Gutaker RM, Reiter E, Furtwängler A, Schuenemann VJ & Burbano HA (2017) Extraction of ultrashort DNA molecules from herbarium specimens. BioTechniques 62: 76–79 doi 10.2144/000114517

Gutaker RM, Weiß CL, Ellis D, Anglin NL, Knapp S, Fernández-Alonso JL, Prat S & Burbano HA (2019) The origins and adaptation of European potatoes reconstructed from historical genomes. Nature Ecology & Evolution 3: 1093–1101, https://doi.org/10.1038/s41559-019-0921-3

Harris SA (1993) DNA analysis of tropical plant species: an assessment of different drying methods. Plant Systematics & Evolution 188: 57–64

Hart ML, Forrest LL, Nicholls JA & Kidner CA (2016) Retrieval of hundreds of nuclear loci from herbarium specimens. Taxon 65(5): 1081–1092

James SA, Soltis PS, Belbin L, Chapman AD, Nelson G, Paul DL & Collins M (2018) Herbarium data: Global biodiversity and societal botanical needs for novel research. Applications in Plant Sciences 6(2): e1024. John Wiley and Sons Inc. https://doi.org/10.1002/aps3.1024

Jónsson H, Ginolhac A, Schubert M, Johnson P & Orlando L (2013) MapDamage 2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29: 1682–1684 doi: 10.1093/bioinformatics/btt193

Lang PLM, Weiß CL, Kersten S, Latorre SM, Nagel S, Nickel B, Meyer M & Burbano HA (2020) Hybridization ddRAD-sequencing for population genomics of nonmodel plants using highly degraded historical specimen DNA. Mol Ecol Resour 20: 1228–1247 https://doi.org/10.1111/17550998.13168

Lindahl T & Andersson A (1972) Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid. Biochemistry 11: 3618–3623

McCabe PF, Levine A, Meijer PJ, Tapon NA & Pennell RI (1997) A programmed cell death pathway activated in carrot cells cultured at low cell density. Plant Journal 12: 267–280

Mohammadin S, Peterse K, Van De Kerke SJ, Chatrou LW, Dönmez AA, Mummenhoff K, Pires JC, Edger PP, Al-Shehbaz IA & Schranz ME (2017) Anatolian origins and diversifi-

cation of *Aethionema*, the sister lineage of the core Brassicaceae. American Journal of Botany 104: 1042–1054

Olofsson JK, Bianconi M, Besnard G, Dunning LT, Lundgren MR, Holota H, Vorontsova MS, Hidalgo O, Leitch IJ, Nosil P, Osborne CP & Christin PA (2016) Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex trait. Molecular Ecology 25: 6107–6123.

Pyle MM & Adams RP (1989) In situ preservation of DNA in plant specimens. Taxon 38: 576–581

Reape TJ, Molony EM & McCabe PF (2008) Programmed cell death in plants: distinguishing between different modes. Journal of Experimental Botany 59: 435–444

R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Särkinen T, Staats M, Richardson JE, Cowan RS & Bakker FT (2012) How to Open the Treasure Chest? Optimising DNA Extraction from Herbarium Specimens. PLoS ONE 7(8): e43808. https://doi.org/10.1371/journal.pone.0043808

Sebastian P, Schaefer H, Telford IRH & Renner SS (2010) Cucumber *(Cucumis sativus)* and melon *(C. melo)* have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. Proceedings of the National Academy of Sciences 107: 14269–14273

Staats M, Cuence A, Richardson JE, Vrielink-Van Ginkel R, Petersen G, Seberg O & Bakker FT (2011) DNA damage in plant herbarium tissue. PLoS ONE 6(12): e28448. https://doi.org/10.1371/journal.pone.0028448

Staats M, Erkens RHJ, Van De Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE & Bakker FT (2013) Genomic Treasure Troves: Complete Genome Sequencing of Herbarium and Insect Museum Specimens. PLoS ONE 8(7): e69189 doi:10.1371/journal.pone.0069189

Straub SCK, Parks M, Weitemeir K, Fishbein M, Cronn RC & Liston A (2012) Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. American Journal of Botany 99: 349–364

Thiers BM (2022) The World's Herbaria 2021: A summary report based on data from Index Herbariorum. https://sweetgum.nybg.org/science/ih/annual-report/

Turner FS (2014) Assessment of insert sizes and adapter content in fastq data from NexteraXT libraries. Frontiers in Genetics 5: 1–7, doi.org/10.3389/fgene.2014.00005

Venables WN & Ripley BD (2002) Modern Applied Statistics with S, Fourth edition. Springer, New York

Wei Z, Zhu S-X, Van den Berg RG, Bakker FT & Schranz ME (2017) Phylogenetic relationships within *Lactuca* L. (Asteraceae), including African species, based on chloroplast DNA sequence comparisons. Genetic Resources and Crop Evolution 64: 55–71

Weiss CL, Schuenemann VJ, Devos J, Shirsekar G, Reiter E, Gould BA, Stinchcombe JR, Krause J & Burbano HA (2016) Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. Royal Society Open Science 3: 160239 http://dx.doi.org/10.1098/rsos.160239

Wicke S & Schneeweiss GM (2015) Next-generation organellar genomics: potentials and pitfalls of high-throughput technologies for molecular evolutionary studies and plant systematics. In: Hörandl E & Appelhans MS (eds) Next generation sequencing in plant systematics, p 9–50. International Association for Plant Taxonomy (IAPT)

Xu C, Dong W, Shi S, Cheng T, Li C, Liu Y, Wu P, Wu H, Gao P & Zhou S (2015) Accelerating plant DNA barcode reference library construction using herbarium specimens: improved experimental techniques. Molecular Ecology Resources 15: 1366–1374 doi: 10.1111/1755-0998.12413

Yao W, Mei C, Nan X & Hui L (2016) Evaluation and comparison of in vitro degradation kinetics of DNA in serum, urine and saliva: A qualitative study. Gene 590: 142–148 https://doi.org/10.1016/j.gene.2016.06.033