1 Forum

2 The maturing relationship between Quaternary paleoecology and ancient 3 sedimentary DNA

4 Mary E Edwards

5 School of Geography and Environmental Sciences, University of Southampton,
6 Southampton, SO17 1BJ, UK and Alaska Quaternary Center, University of Alaska,
7 Fairbanks 99775, USA.

8 m.e.edwards@soton.ac.uk

9 Abstract

10 In the two decades or so since ancient sedimentary DNA (*sed*aDNA) took its place
11 as a new Quaternary paleo-proxy, there have been large advances in the scope of
12 its applications and its reliability. The two main approaches, metabarcoding and
13 shotgun sequencing, have contributed exciting insights into areas such as floristic
14 diversity change, plant-herbivore interactions, extinction, conservation baselines and
15 impacts of invasive species. Early doubts as to its potential to contribute novel
16 information have been dispelled; more is now understood about the passage of
17 *sed*aDNA from the original organism to a component of soil or sediment and about
18 the range of uncertainties that must be addressed in the interpretation of data. With
19 its move into the mainstream, it is now time to develop effective data archives for
20 *sed*aDNA, refine our understanding of central issues such as taphonomy, and further
21 expand the potential for describing, both qualitatively and quantitatively, the history
22 of past ecosystems.

Keywords: ancient sedimentary DNA, paleoecology, metabarcoding, proxy

Introduction

Few if any recent developments in Quaternary science have raised such levels of both interest and contention as ancient DNA (aDNA). Impressive advances have taken place, particularly in the study of mammalian paleoecology, where there have been insights into genetics, lineage development, and extinction (e.g., Shapiro *et al.*, 2004; Heintzman *et al.*, 2016) and also in the understanding of floristic change (e.g., Willerslev *et al.*, 2014). In this forum, I reflect on several decades of research that uses ancient DNA retrieved from sedimentary environments (*sed*aDNA), ancient DNA deposited in sedimentary environments and subsequently retrieved from deposits such as loess, paleosols, and lacustrine sediments. It is particularly useful in addressing a range of key questions about environmental change. The focus has been on plants to date, but the approach can address many other organismal groups.

Previous *sed*aDNA studies received some harsh criticism (e.g., Birks and Birks, 2016). The trajectory of ancient DNA publications by decade from the 1980s (Figure 1) shows a classic evolution. It starts as a topic that is novel and intriguing, then proceeds to the phase called by Jackson (2012) and Birks and Birks (2016) the "reification" of a new method, when its contribution is exaggerated and its shortcomings ignored or underplayed. Subsequently, we reach a more rigorous phase of critique, in which the community in general asks whether key issues have been checked and understood, and how various unusual or even inexplicable results are being dealt with. Today, we have come to a place where *sed*aDNA studies are recognised as a powerful new method in the Quaternary paleoecology repertoire, but

47 with a unique set of problems and pitfalls that require further work to address fully.
48 The contentious early phases of the field originated in no small part from the lack of
49 reciprocal knowledge among Quaternary paleoecologists on the one hand and
50 molecular biologists on the other. With more and more collaboration occurring across
51 this considerable disciplinary divide, *sed*aDNA studies are now contributing fully and
52 usefully to Quaternary paleoecology.

53 Here I attempt a simple overview of the approach, and in doing so address criticisms
54 levelled at it by some traditional paleoecologists with examples of recent studies. As
55 I came to this area from a Quaternary perspective myself, it is clear to me that the
56 fields need to work in partnership, and there is much to be learned on both sides
57 (see also Parducci, 2019). By using the forum, I invite further input from both
58 Quaternary and molecular scientists to update each other on key issues, so that we
59 can advance collaborations even further.

60 How *sed*aDNA is studied

61 Several authors have reviewed methods for environmental and ancient DNA analysis
62 and issues associated with these methods (Pedersen *et al.*, 2015, Taberlet *et al.*,
63 2018). Here I provide a brief overview.

64 *Retrieving DNA from sediments*

65 DNA may reside partly in fossil remains such as pieces of plant material but also as
66 extra-cellular DNA in soil solution or bound to mineral particles. The most promising
67 sediments are cold or frozen, anaerobic, around neutral pH and/or dry (as in cave
68 deposits). Clay-silt particles bind to DNA fragments and tend to reduce enzymatic
69 activity in the process, though recovery of *sed*aDNA may also be successful from

70 organic-rich sediments. Exactly why some sediments yield good results and others
71 not is still under investigation—for example, a recent study has demonstrated that
72 mineral composition and its relation to the charged nature of the sediment appears to
73 have an important effect on DNA retrieval (Freeman *et a*l., 2020), underlining the
74 need for a more detailed understanding of *sed*aDNA taphonomy (see below).

75 The environmental temperature influences preservation to some extent. Warmer
76 environments are expected to lead to shorter preservation times; in lake sediment,
77 for example, preservation appears to be limited to a few millennia in warm climate
78 (Epp *et al*., 2010; Boessenkool *et al*., 2012), compared with 10^4-10^5 ka in cold
79 climates. Fluctuating environmental/storage temperatures are also adverse
80 indicators for preservation. As the degree of DNA degradation and/or fragmentation
81 accumulates over time (with rates differing with environmental and storage
82 conditions), the DNA sequences that remain are often short and fragmented.

83 DNA is usually extracted from sediments in a dedicated clean lab but can also be
84 captured in the field in a sterile buffer solution, and then transferred to the lab (see
85 Taberlet *et al.,* 2018, p 36). After one of several processing pathways, captured
86 fragments are characterized via sequencing the base pairs of the DNA strand. With
87 the availability of high-throughput sequencing, multiple samples containing a range
88 of molecular taxononmic units (MOTUs) can be analysed simultaneously. MOTUs
89 are specific sequences matching information in available look-up databases that
90 could refer to a single species or be shared at the sub-genus, genus, or, in a few
91 cases, family level.

92 Extracts of DNA taken from sediments are subject to further treatment before
93 sequences can be identified. Most paleoecological *sed*aDNA studies to date use

94 metabarcoding, in which specific sequences are aligned with regional catalogues
95 (sequence databases), usually with a focus on a particular oganismal group. Another
96 approach is “shotgun” sequencing, in which all obtained sequences (subject to a
97 practical limit) are compared with a database containing whole genome or partial
98 genome sequences. Each method has its share of imperfections, but each has also
99 yielded exciting results over the past decade.

100

101 *Shotgun sequencing*

102 Shotgun sequencing has the potential to look for a multitude of taxa from different
103 organismal groups. It is less subject to bias introduced by laboratory processing than
104 metabarcoding (see below), but it generates huge challenges at the stage of
105 bioinformatic analysis. The sheer number of possibilities, combined with a highly
106 incomplete global register of genomic information, mean that many taxa go
107 unrecognised. The accidental sharing of a retrieved sequence by one or more
108 organisms that have no link to the system under study can lead to eyebrow-raising
109 results; these are, ideally, reported for full disclosure but then dismissed via a
110 biogeographic/paleoecological argument. Usefully, the ends of older sequences
111 retrieved using a shotgun approach will show deamination damage (tell-tale runs of
112 unlikely bases), which can confirm whether a sequence or set of sequences is
113 relatively ancient and not modern contamination.

114 The shotgun approach in paleoecology is exemplified by a study by Pedersen *et al.*
115 (2016) on the western North American ice-free corridor. Using a multi-proxy
116 approach that included *sed*aDNA, they assessed whether environmental conditions
117 could have allowed human access (via the corridor) prior to or coincident with the

118 first evidence of occupation of regions south of the ice sheet (the data suggest not).
119 This study to some extent suffered from an almost overwhelming signal of many taxa
120 not relevant to the question (e.g. bacterial sequences) that is a common feature of
121 shotgun sequencing. This feature can be circumvented, however, by targeted
122 capture of sequences via specific molecular probes that pick out the sequences of
123 taxa that are of ecological relevance (Taberlet *et al.* 2018).

124

125 *Metabarcoding*

126 The "barcode of life" represents a molecular approach to contemporary taxonomy
127 and identification (Ratnasingham and Herbert, 2007). In a similar vein,
128 paleoecologists identify ancient DNA sequences from Quaternary sediments by
129 matching them against short sequences of DNA taken from modern reference
130 organisms. While the majority of studies focus on plants, it is also possible to study
131 fungi, bacteria, mammals and other vertebrates (see below).

132 Taberlet *et al.* (2007; 2018) provide detailed information on the metabarcoding
133 technique for plants. The approach uses a short (~10-150 base pair) sequence of
134 DNA (metabarcode) from the plant chloroplast or nuclear genome that is highly
135 variable at the genus and species level as the basis for identification. Retrieved
136 sequences are matched against a regional catalogue of metabarcodes. Bearing in
137 mind that sequences longer than ~100-150 base pairs seldom survive long in
138 extracellular situations due to the action of hydrolyzing enzymes, this restricts the
139 useful areas of the plant genome to those showing high variability in a relatively short
140 sequence (Taberlet *et al.*, 2007). The use of high-throughput (next-generation)
141 sequencing allows multiple unidentified sequences derived from many samples to be

142 analysed together, generating huge numbers of sequence reads. While complete
143 laboratory and bioinformatic analyses of a set of samples usually takes several
144 weeks or months, the amount of information gained is highly time-effective, though
145 still costly, compared with a more conventional approach such as palynology.

146 Polymerase chain reaction (PCR) metabarcoding uses primer pairs to capture
147 certain key sequences that can differentiate taxa. As the relevant sequences are rare
148 among all those in the extraction, once captured their numbers are boosted via
149 adding DNA template and polymerase enzymes and multiplying-up the targeted
150 sequences. The total amplified sequence count is *likely* to reflect the original
151 abundance of different DNA sequences in the sample, but it can be unpredictably
152 biased in either direction (too many or two few), particularly for rare MOTUs (Figure
153 2).

154 PCRs can be repeated for each sample of extracted *sed*aDNA. Independent
155 amplifications (replicates) are often repeated as many as eight or 12 times. This
156 approach makes rare sequences more likely to be identified than if only one replicate
157 were used, since they are likely to be missed in a single PCR but should be
158 expected in one or more of the repeat PCRs. This is in contrast to taxa with
159 abundant biomass, which tend to appear in all repeats (see, for example, Alsos *et*
160 *al*., 2015; Clarke *et al*., 2019). This modification makes the estimation of relative
161 abundance of MOTUs less subject to random bias than the use of only one PCR per
162 sample (Figures 2 and 3).

163 Further variation in recovery patterns of sequences relates to the type of polymerase
164 enzyme used (Heintzman *et al*., 2018) and fine details of the PCR process, which
165 are learned and applied in individual labs through experience. These complications

166 increase uncertainty around the estimation of properties such as species richness
167 and may preclude detailed conclusions about relative abundances.

168 Applied first to modern environmental questions such as animal diet analysis, meta-
169 barcoding has subsequently been used in paleoecology to record floristic
170 composition of past vegetation communities, and much more. For example, an early
171 study on DNA retrieved from Quaternary sediments by Willerslev *et al.* (2003)
172 demonstrated that small fragments of DNA retrieved from frozen sediment in Siberia
173 contain information on major plant groupings present over late-Quaternary time.
174 While this information did not provide much paleoecological insight, its main
175 message was to demonstrate that ancient, extracellular DNA was preserved and that
176 it could be identified and attributed to taxonomic groups. A decade later the
177 technique had advanced to such a degree that for samples from late-Quaternary
178 sediments across Siberia, Willerslev *et al.* (2014) demonstrated the presence of
179 numerous MOTUS. Sites largely comprised frozen yedoma sections (loess-derived
180 deposits; see Kanevskiy *et al.*, 2011; Grosse *et al.*, 2013). The impressive advance
181 in taxononomic richness and resolution was aided by the development of large
182 (>2000 entries) metabarcoding species databases for northern regions (Sonstebø *et*
183 *al.*, 2010, Willerslev *et al.*, 2014). A major finding was that, according to the numbers
184 of DNA reads for different plant functional groups, the northern Siberian megafauna
185 may have had access to food sources that comprised forbs as much as they did
186 grasses (by biomass). The extent to which past taxon abundances or biomass might
187 be reflected by the number of occurrences of a particular sequence in a given
188 sample is uncertain and requires further work, however, and thus the quantitative
189 composition of the Pleistocene vegetation remains somewhat speculative.

190 **Taxonomy and taphonomy**

191 Critiques of *sed*aDNA studies mention the limited attainable level of taxonomic
192 resolution (compared with, say, pollen and plant macrofossils), the production of
193 biogeographically incongruous lists of taxa, and a lack of taphonomic understanding
194 (see Birks and Birks, 2016). There were indeed several problems with early studies:
195 they did not use the most complete molecular taxonomy available, and the DNA was
196 then compared with an arbitrary pollen dataset, which also was likely to be
197 taxonomically depauperate (e.g., Pedersen *et al.*, 2013). Some spectacular mixtures
198 of possible and improbable taxa were reported from assemblages, without clear
199 warnings about the vagaries of shotgun sequencing (Smith *et al.*, 2015). That some
200 publications tended to focus on how *sed*aDNA findings compared with, for example,
201 conventional pollen analysis, with little attention to taphonomy or the broader
202 Quaternary context, shows that in some cases data were generated faster than good
203 communication was developed between disciplines. Studies over the past decade
204 reflect greater clarity regarding their physical setting and interpretational constraints;
205 for example, they are placed in clear stratigraphic contexts (e.g., Jorgensen *et al.,*
206 2012; Alsos *et al.*, 2015), and several have now explicitly addressed issues of
207 modern representation, i.e., the calibration of modern DNA against modern
208 vegetation (see below). These studies provide an important foundation for ongoing
209 work, though there is much more still to be learned.

210 Several recent studies investigate the match between near-modern DNA and
211 ecological observations. Yoccoz *et al.* (2012) studied soil DNA in relation to extant
212 plant communities across a landscape in north Norway and found the different plant
213 communities to be reflected in modern soil DNA assemblages. Their data also

214 suggested differential scaling between measured biomass and amount of DNA
215 retrieved for different plant functional types. For example, forb DNA was recovered in
216 a greater proportion to measured forb biomass, whereas the DNA of dwarf shrubs
217 appeared under-represented in relation to biomass. In a study of the relationship of
218 DNA from soil with vegetation on Svalbard, Edwards *et al*. (2018) showed that 98%
219 of identified DNA taxa were recorded in exhaustive vegetation surveys carried out
220 over a 4.0-m radius around each sample location. Further, 50% of extant plant taxa
221 found within 1.0 m of the soil sampling point were represented in DNA. Thus the
222 DNA-vegetation relationship showed high floristic fidelity, but under-representation of
223 full species richness. Furthermore, identified DNA taxa were most likely to be
224 detected if they occurred within 0.5 m of the sampling point. Very few taxa found in
225 the DNA were located at a greater distance (> 1.0 m) from the sampling point (Figure
226 4). Even where different taxa grew on slopes above the sampling point, those taxa
227 did not appear in the samples, suggesting negligible local downslope transport.

228 These findings indicate a local signal should be expected from soil/paleosols, at least
229 in the northern high latitudes. Hence, the results from yedoma/paleosols (as used in
230 Willerselv *et al*., 2014; Zimmerman *et al*., 2017; and Zobel *et al*., 2018) almost
231 certainly represent a highly restricted sampling of the palaeoenvironment. It may also
232 explain the fact that only 100-200 MOTUs, of over 2000 possibilities in the
233 databases, appear in the Willerslev *et al*. (2014) dataset, as so little space on the
234 ground was actually sampled.

235 First steps have been made towards developing a taphonomic theory of *sed*aDNA.
236 Conceptual taphonomic models are provided by Pedersen *et al*. (2015) and Edwards
237 *et al*. (2018), among others. Much is still to be learned about the pathways leading

238 from the initial death assemblage, when DNA leaches from decomposing cells, to
239 what is eventually preserved in soils and other types of sediment, but the increasing
240 number of studies carried out in different types of sedimentary basins provide a
241 growing basis of empirical evidence. Several recent studies have focused on how
242 DNA enters lake sediments and what the DNA “catchment” is for lakes (e.g., via
243 overland flow, seepage, and groundwater, vs. stream/river input). Further questions
244 relate to the representation of vegetation in DNA from lake-surface sediments, coring
245 contamination, down-core leaching (i.e., is there temporal fidelity?) and whether
246 long-distance pollen contributes to *sed*aDNA (see, for example, Parducci *et al.*,
247 2015; Sjogren *et al.,* 2017; Alsos *et al.,* 2018).

248 To assess the spatial and temporal reliability of *sed*aDNA in representing vegetation
249 changes of known extent and date, Sjogren *et al*. (2017) studied the recent
250 *sed*aDNA records of small lakes in Galloway Forest in Scotland, where plantations of
251 non-native conifers were widely established in the 20th century. Soil erosion either
252 occurred prior to any planting or only much later with first harvesting, so reworking of
253 exotic conifer material should not have been a problem. Records from
254 radiometrically dated short cores revealed that the appearance of conifer DNA
255 reflects the time of exotic planting, so I, t is unlikely there was downward leaching of
256 DNA. A large and ecologically consistent turnover of taxa broadly coincident with
257 planting suggests that the record accurately shows the planting (not the first
258 reproductive activity, i.e., pollen production) of the conifers and a change in the field
259 layer of the forest. The DNA catchment appeared to be the hydrologic catchment,
260 and as there were inflowing streams, the plentiful amount of DNA retrieved may
261 reflect this geographically extended source. There was also high floristic fidelity (all

262 observed taxa occurred in vice-county botanical records), but under-sampling of full
263 floristic diversity.

264 The importance of inflowing streams and slope-wash is probably reflected in the
265 DNA record from a large lake in the Polar Urals, where the 215-km^2 hydrological
266 catchment included an inflowing river and spanned hundreds of meters in elevation
267 (Svendsen *et al.*, 2019). Over 100 vascular plant MOTUs were retrieved from
268 numerous sequences (Clarke *et al.*, 2019). In complete contrast, a suite of small
269 lakes lacking inflowing streams in north Norway had more variable DNA recovery,
270 and some records were dominated by in-lake taxa. There appeared to be a strong
271 link between the location of terrestrial species at or near the lake margin and their
272 likelihood of occurrence in lake surface-sediments (Alsos *et al.*, 2018). In a shotgun
273 approach to a lake-sediment study, Parducci *et al.* (2019) also found that in-lake
274 biomass tended to swamp the terrestrial signal. These observations suggest that the
275 conventions of pollen analysis as regards the best lakes for paleoecological studies
276 (i.e., relatively small, no inlet, often productive) do not necessarily hold for DNA
277 studies, and so the relationship of question to site type may need rethinking.

278 **Introduction of contamination**

279 Contamination is a problem for many proxy studies. It certainly can occur during
280 collection of sediment samples and cores, but it is now clear from numerous studies
281 in the early days of *sed*aDNA that draconian anti-contamination measures required
282 by critics and practitioners alike are excessive (see for example, measures taken in
283 Alsos *et al.*, 2015, compared with Clarke *et al.* 2019). A well-taken and properly
284 treated lake-sediment core, sampled for DNA from undisturbed inner sediments,
285 stands no more chance of contamination than any other proxy. Soil and terrestrial

286 sediment samples, however, do require bleached tools, gloves, and great care while
287 sampling, as here there is more chance of contamination.

288 There is a greater chance of contamination in the lab. In *sed*a*DNA* work, the
289 extremely small sample sizes mean that even reagents can and do contribute
290 contamination (typically, widely occurring foodstuffs). Other possible forms of lab
291 contamination are more subtle and include a still unresolved concern about residue
292 from paper and cardboard products, as *Picea* and/or *Pinus* is sometimes filtered out
293 as a potential contaminant (e.g., Alsos *et al.*, 2020). For these reasons, it is
294 important always to take measures to identify contamination sources (see below).

295 Sediment processes, particularly retransportation and reworking, are also of
296 concern. Contention over a *sed*aDNA study that included sediments in Andøya,
297 Norway (Birks *et al.*, 2012; Parducci *et al.*, 2012a,b) lay partly in the possibility that
298 *Pinus* and *Picea* DNA occurred because of unrecognized reworking of ancient
299 material into lake sediments. While it is generally assumed that DNA released to
300 aerobic and non-frozen soil will disappear rapidly if not incorporated into
301 frozen/anaerobic deposits, this is not widely demonstrated. Thus, the possibility
302 exists of "ghost" occurrences of DNA due to erosion of old material into a lake. Some
303 evidence that suggests this may not be a major issue comes from the lacustrine
304 *sed*aDNA record of Clarke *et al.* (2019), where boreal forest taxa are present in
305 samples representing the the Holocene thermal optimum but vanish completely from
306 late-Holocene samples, consistent with their absence from the modern vegetation.

307 Nevertheless, some contamination is almost inevitable in *sed*aDNA studies. An
308 important element during the processing of sequence data is the filtering out of
309 contaminants, both obvious and subtle. The desired result is to reduce false

310 positives while retaining as many true positives as possible (see Ficetola *et al.*,
311 2015). Blank samples at all stages of laboratory processing are used to detect
312 contamination. The total number of reads, or sums of reads across a sample set, or
313 the totals of replicates a sequence is found in, can be used to set reasonable
314 thresholds for positive identification. These criteria are fluid, and much thinking (and
315 some trial and error) is required to set them for a given study (see, for example,
316 Ficetola *et al.*, 2015; Alsos *et al.*, 2015; Clarke *et al.*, 2019). MOTUs representing
317 taxa that are ecologically and/or biogeographically out of place do pass filters, and
318 thus, as with any other paleoecological study, require assessment by a practioner
319 with good biogeographical and/or molecular knowledge. In speciose genera, a
320 single mistake in a sequence may shift an identification from biogeographically
321 acceptable to clearly wrong, leaving a decision to be made as to how to assign the
322 sequence in question.

323 **Data archiving**

324 DNA studies are unusual in the highly complex data that underlie interpretations and
325 conclusions. Making sense of these data and explaining them to audiences not well
326 versed in the details of molecular techniques is important. Furthermore, the
327 coordinated provision of archived data sets that could be used by others in future
328 studies remains an unresolved issue. The vast amounts of bioinformatically
329 processed data stored in genetic databases such as DRYAD (see
330 https://datadryad.org/stash/) are voluminous but may be of little use to many
331 interested in the paleoecological aspects of the study. Quaternary paleoecologists
332 are used to accessing data sorted by site and proxy, accompanied by appropriate
333 metadata such as radiocarbon dates. A given dataset records stratigraphic position,

334 age if available, plus occurrences of identified taxa (see for example, NEOTOMA,
335 https://www.neotomadb.org/). Here, to be widely relevant, the molecular community
336 must adapt to paleoecological norms of data archiving. This is gradually happening,
337 but as yet there is no formal system.

338 The challenge to those working with *sed*aDNA is to provide public access to data
339 that are, on the one hand, understandable in paleoecological terms; on the other
340 hand, the procedures and uncertainties in the dataset must be clear. Examples of
341 these procedures and uncertainties are the relationship of reads to replicates;
342 important details of the processing technique; databases consulted to derive MOTU
343 identity; and filtering protocols and contaminant controls. Indeed, given the flexibility
344 inherent in the filtering process, the initial raw data (all MOTU read numbers for all
345 levels and controls) should be available, if required, as well as the filtered dataset
346 used for paleoecological inference. A group being convened by the NEOTOMA
347 team will shortly address such issues, and input from the broader community is
348 welcome (contact the author).

349 New directions for Quaternary molecular paleoecology

350 Making use of as many proxies as possible is almost always the best strategy for
351 paleoecological studies. As Birks (2000) previously reminded us, pollen and
352 macrofossils together are better than either one alone. The same can be said of
353 *sed*aDNA: using multiple proxies in concert provides some of the strongest
354 inferences. As we understand more about similarities in source area (i.e., DNA more
355 similar to macrofossils than to pollen) and taxonomic resolution (overall pollen and
356 DNA taxonomic resolution is similar in northern regions but differs with respect to
357 taxa; Sonstebø *et al.*, 2010), we can use the strengths of each method to their best

358 advantage. I conclude this forum with a brief survey of recent studies that illustrate
359 the expanding scope of *sed*aDNA studies.

360 Several studies have now demonstrated the power of *sed*aDNA to evaluate changes
361 through time that feature multiple organismal groups, such as combined stratigraphic
362 records of mammals and plants. Examples include grazing patterns in relation the
363 vegetation cover and erosion in the Alps (Giguet-Covex *et al*., 2014), the demise of
364 the Holocene mammoths of St Paul Island (Graham *et al*., 2016), and insights into
365 paleo-migration routes (Pedersen *et al.,* 2016). The addition of *sed*aDNA analysis to
366 sites associated with archaeological finds should enhance information on features
367 such as the introduction of domestic animals, consequent ecosystem change, the
368 appearance of crop plants, and other aspects of natural resource use. Other
369 molecular proxies from lake-sediment records can reveal human colonization/activity
370 (e.g., coprostanol biomarkers; D'Anjou *et al*., 2012), and potentially this approach
371 can be combined with *sed*aDNA analysis (see Mackay *et al*., 2017).

372 Critical but cryptic (with regard to proxies) components of ecosystems can now be
373 addressed. For example, Wood *et al.* (2018) use DNA extracted from packrat
374 middens to track pathogen occurrence through changing climatic conditions in the
375 Atacama Desert. Zobel *et al.* (2018) report changing mycorrhizal mutualisms through
376 time based on the taxonomically rich Willserslev *et al.* (2014) dataset by assigning
377 mycorrhizal preferences to the MOTUs. Belle *et al.* (2014) record recent changes in
378 lacustrine carbon processing by bacterial methanotrophs. There is an exciting
379 potential to record genetic variation through time in plants. Recently, Lammers et al.
380 (ND), using shotgun sequencing, have demonstrated genomic variation at the

381 population level across time in a lacustrine alga, showing the potential for
382 paleoecology to move into paleogenomics.

383 Although this article has focused on terrestrial paleoecology, it should be noted that
384 *sed*aDNA is increasingly used in marine settings. For example, Palowska *et al*.
385 (2020) use diatom DNA in a reconstruction of the history of Stjorfjorden (Svalbard), a
386 critical site for deep-water production, and Giosan *et al*. (2018) similarly track
387 changes in the planktonic composition of the water column for an Indian Ocean core
388 in a reconstruction of winter monsoon dynamics.

389 Finally, given concerns about both future climate change and biodiversity loss,
390 *sed*aDNA can help develop baselines for conservation programmes by contributing
391 to knowledge about past ecosystem state. Clarke *et al*. (2019) use a highly diverse
392 arctic-alpine *sed*aDNA flora from the Polar Urals site to demonstrate the resilience of
393 mountain plant communities in the face of glacial-deglacial climate change.
394 Wilmshurst *et al*. (2014) combine pollen and *sed*aDNA records to establish a pre-
395 colonization baseline for ecological restoration on a New Zealand offshore island.
396 Tracking invasive species and documenting their effect is a critical area in
397 conservation biology, and, in an intriguing study on invasion, Ficetola *et al*. (2018)
398 use a 600-year record to document the before-and-after impact of the introduction of
399 rabbits to the sub-Antarctic Kergeulen Islands. As ecology and paleoecology merge
400 to document and understand effects of humans on many aspects of the environment,
401 *sed*aDNA studies will play an increasingly important role in documenting the extreme
402 and rapidly accelerating changes affecting our planet.

403 **Acknowledgements**

404 This article resulted from discussions at an international seminar held at the
405 Quaternary Research Center (November 2018); my thanks to the QRC for support to
406 attend the meeting. Thanks to Pierre Taberlet, Christian Brochmann, Eske Willerslev
407 and the members of EcoChange working group 2 (EU-FP6: GOCE-2006-036866) for
408 initially inspiring me about ancient DNA. Thanks also to Inger Alsos, Charlotte Clarke
409 and members of the After-Ice team (Norwegian Research Council grant No.
410 230617/E10) for an interesting and fruitful collaboration on *sed*aDNA in Norway.
411 Derek Booth, Charlotte Clarke and two reviewers made helpful suggestions about
412 the manuscript, and Charlotte Clarke kindly supplied Figure 3.

413

414 **References**

415 Alsos, I.G., Coissac, E., Edwards, M., Foreid, M.K., Gielly, L., Sjogren, P., Taberlet,
416 P., Yoccoz, N., 2015. Plant DNA in sediments: to which degree do they represent the
417 flora? *Genome* 58, 186-186.

418 Alsos, I.G., Lammers, Y., Yoccoz, N.G., Jorgensen, T., Sjogren, P., Gielly, L.,
419 Edwards, M.E., 2018. Plant DNA metabarcoding of lake sediments: How does it
420 represent the contemporary vegetation. *Plos One* 13, Aticle No. e0195403.

421 Alsos, I.G., Sjogren, P. Brown, A.G., Gielly, L., Foreid Merkell, M.K., Paus, A.
422 Lammers, Y., Edwards, M.E., Alm, T., Leng, M., Goslar, T., Langdon, C.T., Bakke,
423 J., van der Bilt, W.G.M. 2020. Last Glacial Maximum environmental conditions at
424 Andøya, northern Norway; evidence for a northern ice-edge ecological “hotspot”.
425 Quaternary Science Reviews (accepted).

426 Belle, S., Parent, C., Frossard, V., Verneaux, V., Millet, L., Chronopoulou, P.M.,
427 Sabatier, P., Magny, M., 2014. Temporal changes in the contribution of methane-
428 oxidizing bacteria to the biomass of chironomid larvae determined using stable
429 carbon isotopes and ancient DNA. *Journal of Paleolimnology* 52, 215-228.

430 Birks, H.H., Birks, H.J.B., 2000. Future uses of pollen analysis must include plant
431 macrofossils. *Journal of Biogeography* 27, 31-35.

432 Birks, H.H., Giesecke, T., Hewitt, G.M., Tzedakis, P.C., Bakke, J., Birks, H.J.B.,
433 2012. Comment on "Glacial Survival of Boreal Trees in Northern Scandinavia".
434 *Science* 338, 742.

435 Birks, H.J.B., Birks, H.H., 2016. How have studies of ancient DNA from sediments
436 contributed to the reconstruction of Quaternary floras? *New Phytologist* 209, 499-
437 506.

438 Boessenkool, S. , McGlynn, G. , Epp, L. S. , Taylor, D. , Pimentel, M. , Gizaw, A. ,
439 Nemomissa, S. , Brochmann, C. and Popp, M. (2014): Use of Ancient Sedimentary
440 DNA as a Novel Conservation Tool for High-Altitude Tropical Biodiversity.
441 *Conservation Biology* 28, 446-455.

442 Clarke, C.L., Edwards, M.E., Gielly, L., Ehrich, D., Hughes, P.D.M., Morozova, L.M.,
443 Haflidason, H., Mangerud, J., Svendsen, J.I., Alsos, I.G., 2019. Persistence of arctic-
444 alpine flora during 24,000 years of environmental change in the Polar Urals.
445 *Scientific Reports* 9, Article No. 19613.

446 Edwards, M.E., Alsos, I.G., Yoccoz, N., Coissac, E., Goslar, T., Gielly, L., Haile, J.,
447 Langdon, C.T., Tribsch, A., Binney, H.A., von Stedingk, H., Taberlet, P., 2018.

448 Metabarcoding of modern soil DNA gives a highly local vegetation signal in Svalbard
449 tundra. *Holocene* 28, 2006-2016.

450 Epp, L. S., K. R. Stoof, M. H. Trauth, and R. Tiedemann. 2010. Historical genetics on
451 a sediment core from a Kenyan lake: intraspecific genotype turnover in a tropical
452 rotifer is related to past environmental changes. *Journal of Paleolimnology* 43, 939–
453 954.

454 Freeman, Dieudonné, Collins and Sand (ND). Survival of environmental DNA in
455 natural environments: Surface charge and topography of minerals as driver for DNA
456 storage. Preprint available at https://doi.org/10.1101/2020.01.28.922997

457 Ficetola, G.F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M.,
458 Gielly, L., Lopes, C.M., Boyer, F., Pompanon, F., Raye, G., Taberlet, P., 2015.
459 Replication levels, false presences and the estimation of the presence/absence from
460 eDNA metabarcoding data. *Molecular Ecology Resources* 15, 543-556.

461 Ficetola, G.F., Poulenard, J., Sabatier, P., Messager, E., Gielly, L., Leloup, A.,
462 Etienne, D., Bakke, J., Malet, E., Fanget, B., Storen, E., Reyss, J.L., Taberlet, P.,
463 Arnaud, F., 2018. DNA from lake sediments reveals long-term ecosystem changes
464 after a biological invasion. *Science Advances* 4, Article No. 4292.

465 Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.J., Griggo, C., Gielly, L., Domaizon,
466 I., Coissac, E., David, F., Choler, P., Poulenard, J., Taberlet, P., 2014. Long livestock
467 farming history and human landscape shaping revealed by lake sediment DNA.
468 *Nature Commuications* 5, Article No. 3211.

469 Giosan, L., Orsi, W.D., Coolen, M., Wuchter, C., Dunlea, A.G., Thirumalai, K.,
470 Munoz, S.E., Clift, P.D., Donnelly, J.P. Galy, V., Fuller, D.Q. 2018. Neoglacial

471 climate anomalies and the Harappan metamorphosis. Climate of the Past 14, 1669-
472 1686.

473 Graham, R.W., Belmecheri, S., Choy, K., Culleton, B.J., Davies, L.J., Froese, D.,
474 Heintzman, P.D., Hritz, C., Kapp, J.D., Newsom, L.A., Rawcliffe, R., Saulnier-Talbot,
475 E., Shapiro, B., Wang, Y., Williams, J.W., Wooller, M.J., 2016. Timing and causes of
476 mid-Holocene mammoth extinction on St. Paul Island, Alaska*. Proceedings of the*
477 *National Academy of Sciences USA* 113, 9310-9314.

478 Grosse G, Robinson JE, Bryant R, Taylor MD, Harper W, DeMasi A, Kyker-
479 Snowman E, Veremeeva A, Schirrmeister L, Harden J. 2013. Distribution of late
480 Pleistocene ice-rich syngenetic permafrost of the Yedoma Suite in east and central
481 Siberia, Russia. US Geological Survey Open File Report 2013-1078.

482 Heintzman, P.D., Froese, D., Ives, J.W., Soares, A.E.R., Zazula, G.D., Letts, B.,
483 Andrews, T.D., Driver, J.C., Hall, E., Hare, P.G., Jass, C.N., MacKay, G., Southon,
484 J.R., Stiller, M., Woywitka, R., Suchard, M.A., Shapiro, B., 2016. Bison
485 phylogeography constrains dispersal and viability of the Ice Free Corridor in western
486 Canada. *Proceedings of the National Academy of Sciences USA* 113, 8057-8063.

487 Jackson, S.T., 2012. Representation of flora and vegetation in Quaternary fossil
488 assemblages: known and unknown knowns and unknowns. *Quaternary Science*
489 *Reviews* 49, 1-15.

490 Jorgensen, T., Haile, J., Moller, P., Andreev, A., Boessenkool, S., Rasmussen, M.,
491 Kienast, F., Coissac, E., Taberlet, P., Brochmann, C., Bigelow, N.H., Andersen, K.,
492 Orlando, L., Gilbert, M.T.P., Willerslev, E., 2012. A comparative study of ancient

493 sedimentary DNA, pollen and macrofossils from permafrost sediments of northern
494 Siberia reveals long-term vegetational stability. *Molecular Ecology* 21, 1989-2003.

495 Kanevskiy, M., Shur, Y., Fortier, D., Jorgenson, M.T., Stephani, E., 2011.
496 Cryostratigraphy of late Pleistocene syngenetic permafrost (yedoma) in northern
497 Alaska, Itkillik River exposure. *Quaternary Research* 75, 584-596.

498 Lammers, Y. Heintzman, P.D. 1, Alsos, I.G. ND. Environmental palaeogenomic
499 reconstruction of an Ice Age algal population. Preprint available at
500 https://doi.org/10.1101/2020.04.10.035535.

501 Mackay, H., Henderson, A. C. G., van Hardenbroek, M., Cavers, G., Crone, A.,
502 Davies, K. L., Fonville, T. R., Head, K., Langdon, P. G., Matton, R., McCormick, F.,
503 Murray, E., Whitehouse, N. J., Brown, A. G. 2017. Prehistoric Human-environment
504 Interactions and Their Impact on Aquatic Ecosystems. American Geophysical Union,
505 Fall Meeting 2017, abstract #GC13A-0764.

506 Pawłowska, J., Lacka, M., Kucharska, M., Pawlowski, J., Zajaczkowski1, M. 2020.
507 Multiproxy evidence of the Neoglacial expansion of Atlantic Water to eastern
508 Svalbard. *Climate of the Past* 16, 487–501.

509 Parducci, L., 2019. Quaternary DNA: A multidisciplinary research field. *Quaternary* 2,
510 Article No.37.

511 Parducci, L., Alsos, I.G., Unneberg, P., Pedersen, M.W., Han, L., Lammers, Y.,
512 Salonen, J.S., Valiranta, M.M., Slotte, T., Wohlfarth, B., 2019. Shotgun
513 environmental DNA, pollen, and macrofossil analysis of Lateglacial lake sediments
514 from Southern Sweden. *Frontiers in Ecology and Evolution* 7, Article No. 189.

515 Parducci, L., Bennett, K.D., Ficetola, G.F., Alsos, I.G., Suyama, Y., Wood, J.R.,
516 Pedersen, M.W., 2017. Ancient plant DNA in lake sediments. *New Phytologist* 214,
517 924-942.

518 Parducci, L., Edwards, M.E., Bennett, K.D., Alm, T., Elverland, E., Tollefsrud, M.M.,
519 Jorgensen, T., Houmark-Nielsen, M., Larsen, N.K., Kjaer, K.H., Fontana, S.L., Alsos,
520 I.G., Willerslev, E., 2012. Response to Comment on "Glacial Survival of Boreal Trees
521 in Northern Scandinavia". *Science* 338, 742.

522 Parducci, L., Jorgensen, T., Tollefsrud, M.M., Elverland, E., Alm, T., Fontana, S.L.,
523 Bennett, K.D., Haile, J., Matetovici, I., Suyama, Y., Edwards, M.E., Andersen, K.,
524 Rasmussen, M., Boessenkool, S., Coissac, E., Brochmann, C., Taberlet, P.,
525 Houmark-Nielsen, M., Larsen, N.K., Orlando, L., Gilbert, M.T.P., Kjaer, K.H., Alsos,
526 I.G., Willerslev, E., 2012. Glacial Survival of Boreal Trees in Northern Scandinavia.
527 *Science* 335, 1083-1086.

528 Parducci, L., Valiranta, M., Salonen, J.S., Ronkainen, T., Matetovici, I., Fontana,
529 S.L., Eskola, T., Sarala, P., Suyama, Y., 2015. Proxy comparison in ancient peat
530 sediments: pollen, macrofossil and plant DNA. *Philosophical Transactions of the*
531 *Royal Society B* 370, Article No. 20130382.

532 Pedersen, M.W., Ginolhac, A., Orlando, L., Olsen, J., Andersen, K., Holm, J.,
533 Funder, S., Willerslev, E., Kjaeer, K.H., 2013. A comparative study of ancient
534 environmental DNA to pollen and macrofossils from lake sediments reveals
535 taxonomic overlap and additional plant taxa. *Quaternary Science Reviews* 75, 161-
536 168.

537 Pedersen, M.W., Overballe-Petersen, S., Ermini, L., Sarkissian, C.D., Haile, J.,
538 Hellstrom, M., Spens, J., Thomsen, P.F., Bohmann, K., Cappellini, E., Schnell, I.B.,
539 Wales, N.A., Caroe, C., Campos, P.F., Schmidt, A.M.Z., Gilbert, M.T.P., Hansen,
540 A.J., Orlando, L., Willerslev, E., 2015. Ancient and modern environmental DNA.
541 *Philosophical Transactions of the Royal Society B* 370, Article No. 20130383.

542 Pedersen, M.W., Ruter, A., Schweger, C., Friebe, H., Staff, R.A., Kjeldsen, K.K.,
543 Mendoza, M.L.Z., Beaudoin, A.B., Zutter, C., Larsen, N.K., Potter, B.A., Nielsenlo,
544 R., Rainville, R.A., Orlando, L., Meltzer, D.J., Kjaer, K.H., Willerslev, E., 2016.
545 Postglacial viability and colonization in North America's ice-free corridor. *Nature* 537,
546 45-49.

547 Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: The Barcode of Life Data System
548 (www.barcodinglife.org). *Molecular Ecology Notes* 7, 355-364.

549 Shapiro, B., Drummond, A.J., Rambaut, A., Wilson, M.C., Matheus, P.E., Sher, A.V.,
550 Pybus, O.G., Gilbert, M.T.P., Barnes, I., Binladen, J., Willerslev, E., Hansen, A.J.,
551 Baryshnikov, G.F., Burns, J.A., Davydov, S., Driver, J.C., Froese, D.G., Harington,
552 C.R., Keddie, G., Kosintsev, P., Kunz, M.L., Martin, L.D., Stephenson, R.O., Storer,
553 J., Tedford, R., Zimov, S., Cooper, A., 2004. Rise and fall of the Beringian steppe
554 bison. *Science* 306, 1561-1565.

555 Sjogren, P., Edwards, M.E., Gielly, L., Langdon, C.T., Croudace, I.W., Merkel,
556 M.K.F., Fonville, T., Alsos, I.G., 2017. Lake sedimentary DNA accurately records
557 20th Century introductions of exotic conifers in Scotland. *New Phytologist* 213, 929-
558 941.

559 Smith, O., Momber, G., Bates, R., Garwood, P., Fitch, S., Pallen, M., Gaffney, V.,
560 Allaby, R.G., 2015. Sedimentary DNA from a submerged site reveals wheat in the
561 British Isles 8000 years ago. *Science* 347, 998-1001.

562 Sonstebo, J.H., Gielly, L., Brysting, A.K., Elven, R., Edwards, M., Haile, J.,
563 Willerslev, E., Coissac, E., Rioux, D., Sannier, J., Taberlet, P., Brochmann, C., 2010.
564 Using next-generation sequencing for molecular reconstruction of past Arctic
565 vegetation and climate. *Molecular Ecology Resources* 10, 1009-1018.

566 Svendsen, J.I., Haflidason, H., Henriksen, M., Hovland, M.N., Mangerud, J.,
567 Nazarov, D., Lohne, O.S., Regnell, C., Gyllencreutz, R., Faerseth, L.M.B. &
568 Schaefer, J. 2019: Glacial and environmental changes during the last 60,000 years
569 in the Polar Ural Mountains, Arctic Russia, inferred from a high resolution lake record
570 and observations from adjacent areas. *Boreas* 48, 407-431.

571 Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat,
572 T., Corthier, G., Brochmann, C., Willerslev, E., 2007. Power and limitations of the
573 chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* 35,
574 Article No. e14.

575 Taberlet, P., Bonin, A., Zinger, L. Coissac, E. 2018. *Environmental DNA for*
576 *Biodiversity Research and Monitoring*. Oxford. Oxford University Press. 253 pp.

577 Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M.E.,
578 Lorenzen, E.D., Vestergard, M., Gussarova, G., Haile, J., Craine, J., Gielly, L.,
579 Boessenkool, S., Epp, L.S., Pearman, P.B., Cheddadi, R., Murray, D., Brathen, K.A.,
580 Yoccoz, N., Binney, H., Cruaud, C., Wincker, P., Goslar, T., Alsos, I.G., Bellemain,
581 E., Brysting, A.K., Elven, R., Sonstebo, J.H., Murton, J., Sher, A., Rasmussen, M.,

582 Ronn, R., Mourier, T., Cooper, A., Austin, J., Moller, P., Froese, D., Zazula, G.,
583 Pompanon, F., Rioux, D., Niderkorn, V., Tikhonov, A., Savvinov, G., Roberts, R.G.,
584 MacPhee, R.D.E., Gilbert, M.T.P., Kjaer, K.H., Orlando, L., Brochmann, C., Taberlet,
585 P., 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* 506,
586 47-51.

587 Willerslev, E., Hansen, A.J., Binladen, J., Brand, T.B., Gilbert, M.T.P., Shapiro, B.,
588 Bunce, M., Wiuf, C., Gilichinsky, D.A., Cooper, A., 2003. Diverse plant and animal
589 genetic records from Holocene and Pleistocene sediments. *Science* 300, 791-795.

590 Wilmshurst, J.M., Moar, N.T., Wood, J.R., Bellingham, P.J., Findlater, A.M.,
591 Robinson, J.J., Stone, C., 2014. Use of Pollen and Ancient DNA as Conservation
592 Baselines for Offshore Islands in New Zealand. *Conservation Biology* 28, 202-212.

593 Wood, J.R., Diaz, F.P., Latorre, C., Wilmshurst, J.M., Burge, O.R., Gutierrez, R.A.,
594 2018. Plant pathogen responses to Late Pleistocene and Holocene climate change
595 in the central Atacama Desert, Chile. *Scientific Reports* 8, Article No. 17208.

596 Yoccoz, N.G., Brathen, K.A., Gielly, L., Haile, J., Edwards, M.E., Goslar, T., von
597 Stedingk, H., Brysting, A.K., Coissac, E., Pompanon, F., Sonstebo, J.H., Miquel, C.,
598 Valentini, A., de Bello, F., Chave, J., Thuiller, W., Wincker, P., Cruaud, C., Gavory,
599 F., Rasmussen, M., Gilbert, M.T.P., Orlando, L., Brochmann, C., Willerslev, E.,
600 Taberlet, P., 2012. DNA from soil mirrors plant taxonomic and growth form diversity.
601 *Molecular Ecology* 21, 3647-3655.

602 Zimmermann, H.H., Raschke, E., Epp, L.S., Stoof-Leichsenring, K.R., Schirrmeister,
603 L., Schwamborn, G., Herzschuh, U., 2017. The History of Tree and Shrub Taxa on

Bol'shoy Lyakhovsky Island (New Siberian Archipelago) since the Last Interglacial Uncovered by Sedimentary Ancient DNA and Pollen Data. *Genes* 8, Article No. 273.

Zobel, M., Davison, J., Edwards, M.E., Brochmann, C., Coissac, E., Taberlet, P., Willerslev, E., Moora, M., 2018. Ancient environmental DNA reveals shifts in dominant mutualisms during the late Quaternary. *Nature Communications* 9, Article No. 139.

Figure Captions

Figure 1: Approximate number of published articles on ancient DNA through time, based on the Web of Knowledge search engine.

Figure 2: PCR metabarcoding can introduce bias during laboratory processing that is more significant than that affecting other biological proxies. Upper: a “true” proportion of sequences in an extraction prior to PCR. Lower: after PCR, variable capture of sequences by primers means a rare sequence may be missed and never amplified, while after several cycles of amplification, a dominant sequence may swamp other sequences. In this example, the rare sequence is captured in only one of three replicates (see Ficetola *et al.,* 2015).

Figure 3: Numerical abundance of sequence reads (histograms) and proportion of replicates (diamonds) for the functional group “dwarf shrubs” and the genus *Dryas* through time at Bolshoi Shchuchye Lake, Russia (after Clarke *et al.*, 2019)

Figure 4: Upper: Svalbard tundra soil samples vs vegetation—all but one recovered MOTU matched observed taxa in vegetation, but only 50% of vegetation taxa were identified in the DNA.   Lower: Taxa detected in both vegetation and soil DNA at

626 sampling points. Nearly all such occurrences were found within 0.5 m of the sample

627 point (see Edwards *et al.*, 2018).

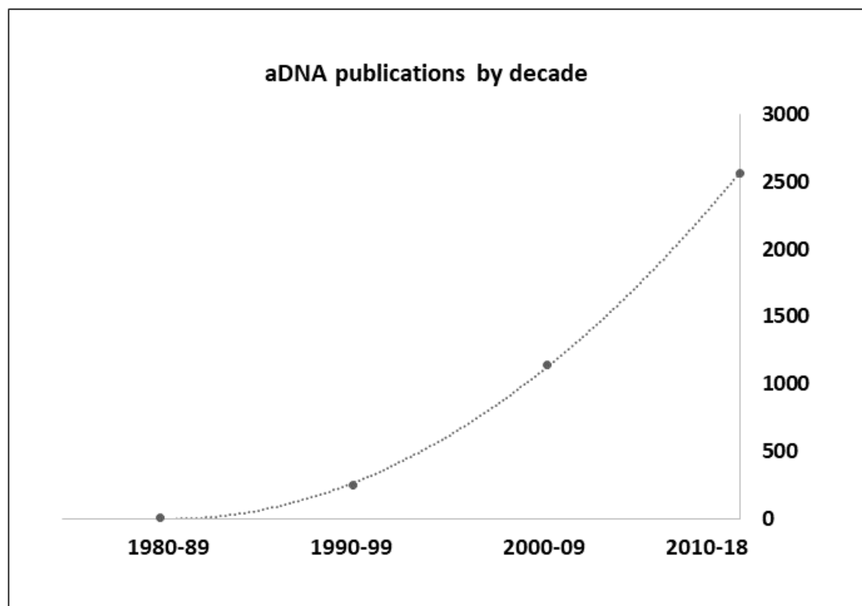
628

629

630

631

632 **Fig 1**



633

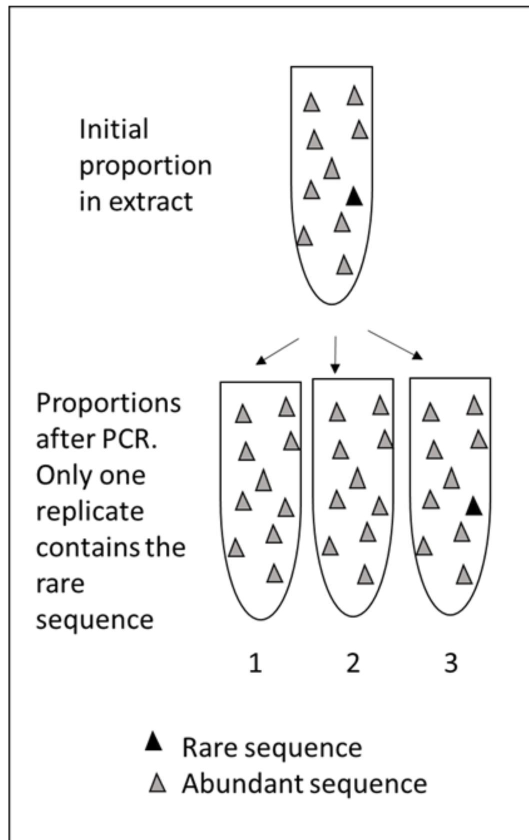
634

635

636

637

638 **Fig 2**



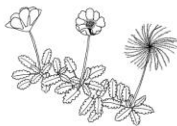
639

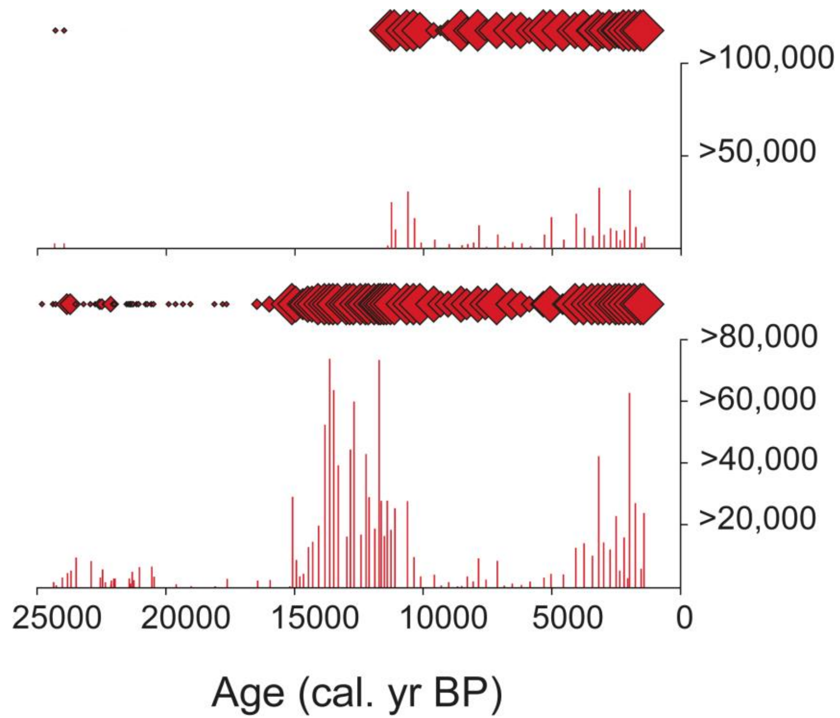
640

Fig 3

Dwarf shrubs

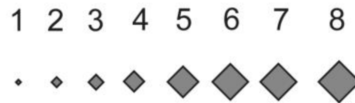
Dryas



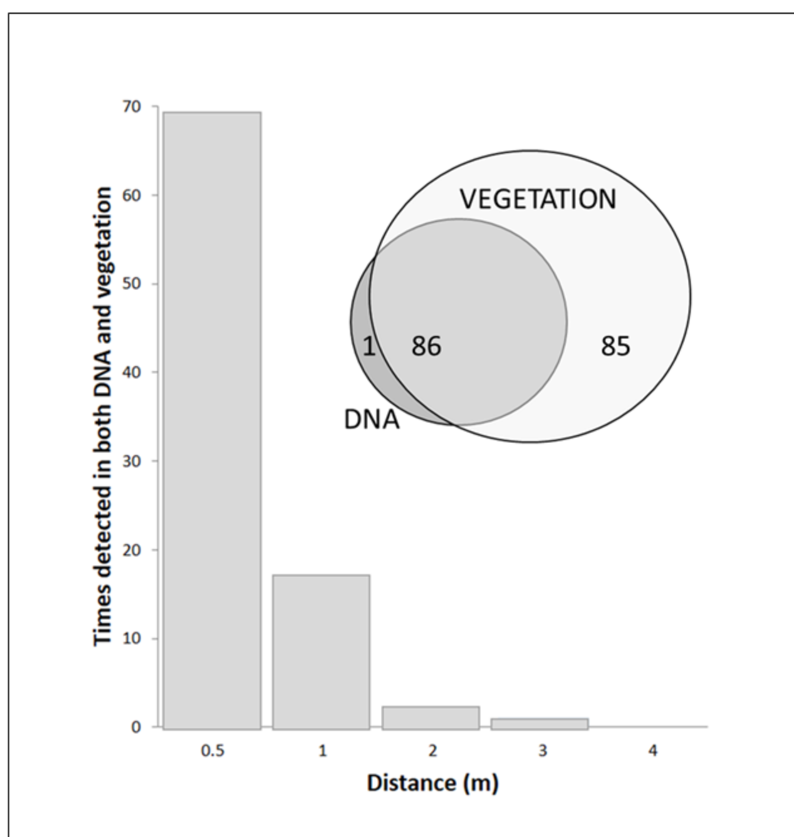


Number of sequence reads

Number of PCR repeats with taxon/functional group present (out of eight)



646 **Fig 4**



647

648