Fifty thousand years of Arctic vegetation and megafaunal diet

Eske Willerslev, John Davison, Mari Moora, Martin Zobel, Eric Coissac, Mary E. Edwards, Eline D. Lorenzen, Mette Vestergård, Galina Gussarova, James Haile, et al.

Fifty thousand years of arctic vegetation and megafaunal diet

Eske Willerslev[1*], John Davison[2*], Mari Moora[2*], Martin Zobel[2*], Eric Coissac[3*], Mary E. Edwards[4*], Eline D. Lorenzen[1,5*], Mette Vestergård[1*], Galina Gussarova[6,7*], James Haile[1,8*], Joseph Craine[9], Gaddy Bergmann[10], Ludovic Gielly[3], Sanne Boessenkool[6†], Laura S. Epp[6†] Peter B. Pearman[11], Rachid Cheddadi[12], David Murray[13], Kari Anne Bråthen[14], Nigel Yoccoz[14], Heather Binney[4], Corinne Cruaud[15], Patrick Wincker[15], Tomasz Goslar[16,17], Inger Greve Alsos[18], Eva Bellemain[6†], Anne Krag Brysting[19], Reidar Elven[6], Jørn Henrik Sønstebø[6], Julian Murton[20], Andrei Sher[21§], Morten Rasmussen[1], Regin Rønn[22], Tobias Mourier[1], Alan Cooper[23], Jeremy Austin[23], Per Möller[24], Duane Froese[25], Grant Zazula[26], François Pompanon[3], Delphine Rioux[3], Vincent Niderkorn[27], Alexei Tikhonov[28], Grigoriy Savvinov[29], Richard G. Roberts[30], Ross D. E. MacPhee[31], M. Thomas P. Gilbert[1], Kurt H. Kjær[1], Ludovic Orlando[1], Christian Brochmann[6**] & Pierre Taberlet[3**]

1. Centre for GeoGenetics, Natural History Museum, University of Copenhagen, Oster Voldgade 5-7 DK-1350 Copenhagen K, Denmark.

2. Department of Botany, Institute of Ecology and Earth Sciences, University of Tartu, 40 Lai St., 51005 Tartu, Estonia.

3. Laboratoire d'Ecologie Alpine (LECA) CNRS UMR 5553, Univ. Joseph Fourier, BP 53 38041 Grenoble Cedex 9, France.

4. Geography and Environment, University of Southampton, Southampton SO17, 1BJ, UK.

5. Department of Integrative Biology, University of California Berkeley, 1005 Valley Life Sciences Building, Berkeley, CA 94720, USA.

6. National Centre for Biosystematics, Natural History Museum, University of Oslo, P.O. Box 1172, Blindern, NO-0318, Oslo, Norway

7. Department of Botany, St. Petersburg State University, Universitetskaya nab. 7/9 199034, St. Petersburg, Russia.

8. Ancient DNA Laboratory, Veterinary and Life Sciences School, Murdoch University, 90 South Street, Perth, Australia.

9. Division of Biology, Kansas State University, Manhattan KS 66506-4901 USA.

10. Ecology and Evolutionary Biology, University of Colorado, Boulder, USA.

11. Landscape Dynamics Unit, Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland.

12. Institut des Sciences de l'Evolution de Montpellier, UMR 5554 Université Montpellier 2, Bat.22, CC061, Place Eugène Bataillon 34095 Montpellier cedex5, France.

13. University of Alaska Museum of the North, Fairbanks AK 99775-6960, USA.

14. Department of Arctic and Marine Biology, UiT The Arctic University of Norway, NO-9037 Tromsø, Norway.

15. Genoscope, Institut de Genomique du Commissariat à l'Energie Atomique (CEA), Evry, France.

16. Adam Mickiewicz University, Faculty of Physics, Umultowska 85, 61-614 Poznań, Poland.

17. Poznań Radiocarbon Laboratory, Poznań Science and Technology Park, Rubież 46, 61-612 Poznań, Poland.

18. Tromsø University Museum, NO-9037 Tromsø, Norway.

19. Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, NO-0316 Oslo, Norway.

20. Permafrost Laboratory, Department of Geography, University of Sussex, Brighton, BN1 9QJ, United Kingdom.

21. Institute of Ecology and Evolution, Russian Academy of Sciences, 33 Leninsky Prospect, 119071 Moscow, Russia.

22. Department of Biology, Terrestrial Ecology, Universitetsparken 15, DK- 2100 Copenhagen Ø, Denmark.

23. Australian Centre for Ancient DNA, School of Earth & Environmental Sciences, University of Adelaide, South Australia, 5005 Australia.

24. Department of Geology/Quaternary Sciences, Lund University Sölvegatan 12, SE-223 62 Lund, Sweden.

25. Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, Alberta T6G 2E3, Canada.

26. Government of Yukon, Department of Tourism and Culture, Yukon Palaeontology Program, PO Box 2703 L2A, Whitehorse, Yukon Territory Y1A 2C6, Canada.

27. INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France.

28. Zoological Institute of Russian Academy of Sciences, Universitetskaya nab. 1, 199034 Saint-Petersburg, Russia.

29. Institute of Applied Ecology of the North of North-Eastern Federal University, Belinskogo str. 58, 677000 Yakutsk, Republic of Sakha (Yakutia), Russia.

30. Centre for Archaeological Science, School of Earth and Environmental Sciences, University of Wollongong, Wollongong, NSW 2522, Australia.

31. Division of Vertebrate Zoology/Mammalogy, American Museum of Natural History, New York, NY 10024, USA.

*These authors contributed equally and should be regarded joint first authors

**These authors should be regarded joint senior authors

§Deceased 2008

†Present addresses: Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, P.O Box 1066, Blindern, NO-0318, Oslo, Norway;.Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Research Unit Potsdam, Telegrafenberg A 43, 14473 Potsdam, Germany; SpyGen, Savoie Technolac, 17 allée du lac Saint André, BP 274, 73375 Le Bourget-du-Lac Cedex, France.

Although it is generally agreed that the arctic flora is among the youngest and least diverse on Earth, the processes that shaped it are poorly understood. Here we present 50 thousand years (kyr) of arctic vegetation history, derived from the first large-scale ancient DNA metabarcoding study of circumpolar plant diversity. For this interval we additionally explore nematode diversity as a proxy for modelling vegetation cover and soil quality, and diets of herbivorous megafaunal mammals, many of which became extinct around 10 kyr BP (before present). For much of the period investigated, arctic vegetation consisted of dry steppe tundra dominated by forbs (non-graminoid herbaceous vascular plants). During the Last Glacial Maximum (25–15 kyr BP), diversity declined markedly, although forbs remained dominant. Much changed after 10 kyr BP, with the appearance of moist tundra dominated by woody plants and graminoids. Our analyses indicate that both graminoids and forbs would have featured in megafaunal diets. As such our findings question the predominance of a late Quaternary graminoid-dominated arctic "mammoth steppe".

90 It can be argued that arctic vegetation during the proximal Quaternary (the last *c*. 50 kyr) is less
91 well understood than the ecology and population dynamics of the mammals that consumed it,
92 despite the overall uniformity and low floristic diversity of Arctic vegetation [1-2] Analyses of
93 vegetation changes during this interval have been based mainly on fossil pollen. Although highly
94 informative, records tend to be biased toward high pollen producers such as many graminoids
95 (grasses, sedges, and rushes) and *Artemisia*, which can obscure the abundance of other forms
96 such as many insect-pollinated forbs [1]. Arctic pollen records are rarely comprehensively
97 identified to species level, which underestimates actual diversity[3]. These problems are to some
98 extent ameliorated by plant macrofossil studies (e.g.[4]), which may provide detailed records of
99 local vegetation. However, macrofossil studies are far less common, have their own taxonomic
100 constraints, and usually cannot provide quantitative estimates of abundance.

101 In recent years, a complementary approach has emerged that utilizes plant and animal ancient
102 DNA preserved in permafrost sediments [5]. Such environmental DNA [11] does not derive primarily
103 from pollen, bones, or teeth, but from above- and below-ground plant biomass, faeces, discarded
104 cells and urine preserved in sediments [6-8]. Like macrofossils, environmental DNA appears to be
105 local in origin [9-12] and in principle the survival of a few fragmented DNA molecules is sufficient
106 for retrieval and taxonomic identification [13].

107 Environmental DNA can supply the fraction of the plant community not readily identifiable by
108 pollen analysis and, to some extent, macrofossils, particularly in vegetation dominated by non-
109 woody growth forms [6]. For most plant groups, DNA permits identification at lower taxonomic
110 levels than pollen[14]. Additionally, environmental DNA records have proven to reflect not only
111 the qualitative but also the quantitative diversity of aboveground plant [12] and animal taxa [8], as
112 determined from modern sub-surface soils.

113 Leaching of DNA through successive stratigraphic zones may be an issue in temperate
114 conditions [8,10] but not in permafrost [5] or in sediments that have only recently thawed [15]. Re-
115 deposition of sediments and organics can confound results, which is also the case for pollen and
116 macrofossils [6,16], but can be avoided and accounted for by careful site selection and by excluding
117 rare DNA sequence reads [16]. For Quaternary permafrost settings, at least, taphonomic bias due to
118 differences in DNA survival across plant groups does not appear to be of concern (see Methods

section 4.0 on taphonomy), as has been shown by a comparative permafrost ancient DNA study of plants and their associated fungi [17].

Reconstruction of Arctic vegetation from permafrost samples

We collected 242 sediment samples from 21 sites across the Arctic (Fig. 1, Extended Data Table 1). Ages were determined by accelerator mass spectrometry radiocarbon (^{14}C) dating, and are reported here in thousands of calibrated (calendar) years BP (Extended Data Figure 1, Supplementary Data 1). We sequenced the short P6 loop sequence of the *trn*L plastid region and a part of the ITS1 spacer region through metabarcoding (Methods section 3.0), generating a total of 14,601,839 *trn*L plant DNA sequence reads and 1,652,857 ITS reads. Reads were identified by comparison with (i) the arctic *trn*L taxonomic reference library [14], which we extended with ITS sequences for three families; (ii) a new north boreal *trn*L taxonomic reference library constructed by sequencing 1,332 modern plant samples representing 835 species; and (iii) GenBank, using the program ecoTag (Supplementary Data 2, Methods section 3.0). Basic statistics, *in silico* analyses, and additional experiments were carried out to check data reliability (Extended Data Figure 2, Extended Data Table 2). We grouped the identified molecular operational taxonomic units (MOTUs) into three distinct intervals (Fig. 2a): i) pre-LGM (50–25 kyr BP), a period of fluctuating climate; (ii) LGM (25–15 kyr BP), a period of constantly cold and dry conditions; and (iii) post-LGM (15-0 kyr BP), the current interglacial, characterised by relatively higher temperatures [17].

Shifts in plant composition and lower diversity during the LGM

To address compositional changes in vegetation across space and time we used a generalised linear model and permutational multivariate analysis of variance (Permanova) (Supplementary Data 3, Methods section 5.0). We find that (i) the composition of plant MOTU assemblages differed significantly across the three intervals (pseudo-F = 6.77, $p < 0.001$, Extended Data Figure 3a-e), with pre-LGM and post-LGM plant assemblages differing the most (Extended Data Figure 3f); (ii) the greater the spatial distance separating a pair of samples within each time period, the less similar their composition ($p < 0.001$); and (iii) LGM assemblages were most homogeneous across space and post-LGM assemblages were most heterogeneous (Fig. 2).

LGM pollen spectra show high floristic richness compared to other intervals (e.g. [1]). This is due to the limited occurrence of woody taxa with high pollen production, which in turn proportionately emphasizes lower pollen-producing taxa. In contrast, our DNA data reveal that plant diversity was lowest during LGM relative to other intervals (Fig. 2a). Plant assemblages became more similar to each other and the estimated number of MOTUs decreased from pre-LGM to LGM (Fig. 2a), with many taxa absent that had previously been well represented (Fig. 2b). In addition, while the LGM flora was largely a subset of the pre-LGM flora, the post-LGM flora was different (Fig. 2b), with pronounced geographic differentiation (Fig. 2c).

Steppe-tundra

Due to the low taxonomic resolution of previously published vegetation reconstructions, it remains undetermined whether arctic vegetation during the last part of the Quaternary was a form of tundra or more like steppe (e.g.[18,19]). Small-scale contemporary analogues range from low-productivity fellfields and cryoxeric steppe communities to more productive dry arctic steppe-to-tundra gradients. Our sediment DNA plant sequence data from ~50–12 kyr BP encompass taxa that typify both tundra and arctic steppe environments. These include taxa that are today typical of dry and/or disturbed sites (e.g. *Bromus pumpillianus*, *Artemisia frigida*, *Plantago canescens*, *Anemone patens*), saline soils (*Puccinellia*, *Armeria*), moist habitats (*Caltha*) and rocky or fellfield habitats (*Dryas*, *Draba*), plus a woody component dominated by *Salix* (Supplementary Data 4 and 5). A spatial and/or temporal mosaic of plant communities is indicated (Methods section 6.0), as is seen in floristically rich macrofossil records [4]. The most common MOTU in the pre-LGM and LGM samples is Anthemidae Group 1 (*Artemisia, Achillea, Chrysanthemum, Tanacetum*), which underscores the importance in regional pollen assemblages of Asteraceae in general and *Artemisia* in particular [1]. *Equisetum* and *Eriophorum* are important only in postglacial assemblages, reflecting moister soil conditions. Increases in aquatic taxa (Supplementary Data 4 and 5) also indicate a predominance of moister substrates in the later part of the post-LGM period. These findings indicate a shift from dry steppe-tundra to moist tundra in the early part of the post-LGM period—a change widely reported in other proxy studies.

Nematode assemblage composition is known to change significantly with vegetation cover [20] moisture [21] and organic resource inputs [22]. Therefore, to obtain a complementary proxy for vegetation cover and soil quality, we characterized the soil nematode fauna of contemporary mesic shrub tundra and subarctic steppe on well-drained loess soils in Yukon Territory, Canada (Fig. 1, Extended Data Table 3). The relative proportion of the nematode families Teratocephalidae and Cephalobidae varied among vegetation types ($p < 0.001$, nested ANOVA), and indicator species analysis [23] confirmed that Teratocephalidae (indicator value = 0.98, $p = 0.001$) and Cephalobidae (indicator value = 0.98, $p = 0.001$) are very good indicators of tundra and steppe vegetation, respectively (Fig. 3). These findings are in agreement with previous studies restricted to subarctic Sweden [24,25] and alpine and subalpine habitats [26-27]. We amplified short DNA sequences from these two taxa from 17 sediment samples analysed for plant DNA from Yukon and northeastern Siberia. We detected Cephalobidae DNA in almost all samples, while Teratocephalidae was detected at a higher frequency in samples younger than 10 kyr BP than in the pre-LGM and LGM samples (Extended Data Table 4). These results support our inferences from plant sequence data and indicate a transition from relatively dry tundra and steppe towards more moist tundra during the post-LGM interval.

Forb dominance and megafaunal diets

To assess structural and functional shifts in the plant assemblages, we investigated temporal changes in the relative abundance of different growth forms. Our DNA results show that pre-LGM vegetation was dominated by forbs, the relative share of which increased during the LGM, whereas graminoids constituted less than 20% of the total read count (Fig. 4a). These results persisted when we corrected for observed modern representational bias [12] (Methods sections 4.0 and 5.3).

Continued forb dominance during the LGM implies that similar proportions of forbs and graminoids were maintained through this period, despite the significant decline in floristic diversity (Fig. 2a,b). Our findings contrast with pollen-based reconstructions, which have emphasized dominance of graminoids in the unglaciated Arctic and adjacent regions, particularly during the LGM, and exemplified by the widely-used term “mammoth-steppe” [19]. Rather, our results show that vegetation was forb-dominated in both overall abundance of MOTUs and in

floristic richness (Fig. 4a,b, Extended Data Figure 3g,h), in agreement with macrofossil data that show a diversity of forbs of mixed ecological preference (e.g. [4]).

We explored whether forbs were prominent in habitats favoured by megafauna by analysing 25 dated (47-20 kyr BP) sediment samples from Main River, Siberia, using *trn*L plastid plant and 16S mtDNA mammal primers. We found that the mean proportion of forbs was higher in samples from which herbivorous megafaunal DNA had been retrieved (n = 18; e.g. woolly mammoth, woolly rhinoceros, horse, reindeer and elk) than in samples lacking such DNA (n = 7; Fig. 4c, Extended Data Table 5). Although suggestive of co-occurrence of megafauna in forb-dominated settings, these results should be regarded as tentative, and further studies are needed to verify if this is indeed a general trend.

We also investigated whether megafaunal diets revealed the level of forb dominance observed in permafrost sediment samples. Using standardised methods, we genetically characterised intestinal/stomach contents and coprolites recovered from 8 specimens of woolly mammoth, woolly rhinoceros, bison and horse from Siberia and Alaska, dated > 55–21 kyr BP (Extended Data Table 6, Methods sections 3.0 and 7.0). Although ingested plant remains are often difficult to identify morphologically, they can be accurately identified [28,29] and roughly quantified [30] using DNA. The majority of these samples are dominated by forbs, which comprise 0.63 ± 0.12 of the sequences, compared to 0.27 ± 0.16 expressing graminoid sequences (Fig. 4d, Supplementary Data 6). These results suggest that megafaunal species supplemented their diets with high-protein forbs rather than specializing more or less exclusively on grasses.

To confirm the reliability of our *trn*L approach for estimating herbivore diet, we analysed 50 rumen samples of sheep-feed diets with varying proportions of forbs (white clover, *Trifolium repens*) and graminoids (ryegrass, *Lolium perenne*) (Methods section 5.4). As seen in Figure 4e, the Pearson correlation coefficient between the actual fraction of forbs in these diets and the proportion of forbs estimated with the DNA-based approach was highly significant (r^2=0.75, $p < 10^{-15}$).

Perspectives

Our observations of high forb abundance in the terminal Pleistocene may merely reflect vegetation response to glacial climates, but there are other possibilities [1]. An abundant megafauna would have caused significant trampling [31], enhancing gap-based recruitment [32], which could favour forbs [33]. Coupled with nitrogen input from wide-ranging herbivores[34], forbs may out-compete grasses [35]. Furthermore, a diet rich in forbs may help explain how numerous large animals were sustained; forbs may be more nutrient-rich (e.g. [35]) and more easily digested [36] than grasses. However, a feedback loop that maintained nutritious and productive forage and supported large mammalian populations in glacial climate regimes may have been impossible to maintain after deglaciation, as C:N ratios increased with global warming [37], and the potential breakdown of the megafauna-forb interaction would have been exacerbated by declining mammalian populations. In contemporary tundra and steppe (the latter often called grasslands), graminoids are generally perceived to be the dominant growth form in large herbivore habitats (e.g. [38,39]). Our data, which unearth 50 kyr of arctic vegetation history, call this perception into question.

Methods summary

Plant fragments or soil matrix organics were ^{14}C-dated using accelerator mass spectrometry and measured [1] ages were converted into calendar years [40]. Permafrost sampling [5], DNA extraction [11], PCR amplification [41] and taxon identification (e.g. [12]) followed established procedures. Most vascular taxa are covered by [42], and nomenclature is provided accordingly; for the remaining taxa nomenclature follows [43]. Dissimilarity between plant assemblages was quantified using pairwise Bray-Curtis distance [44]. Variation in assemblage dissimilarity was decomposed using Permutational Multivariate Analysis of Variance (Permanova [45]) and visualised using non-metric multidimensional scaling [46,47]. We used a distance decay approach [49] and a generalized linear model to model variation in plant community assemblages over space and time. Growth form composition of communities was compiled from species trait databases [49]. Differences in the trait composition of assemblages in adjacent climatic periods were compared to a null model assuming random assortment from the previous interval. Nematode faunas of 35 contemporary sediment samples were morphologically determined. Presence of two indicator families (Teratocephalidae for tundra and Cephalobidae for steppe) was genetically determined in 17

ancient sediment samples. Megafaunal DNA and faeces and gut content were determined genetically following established methods. For a detailed discussion, see Methods.

References

1. Anderson, P. M., Edwards, M. E. & Brubaker, L. B. in *The Quaternary Period in the United States. Developments in Quaternary Science* (Gillespie, A. E., Porter, S. C. & Atwater, B. F.) 427–440 (Elsevier, New York, 2003).
2. Murray, D. F. in *Arctic and alpine biodiversity: Patterns, causes and ecosystem consequences* (Chapin, F. S. & Körner, C.) 21–32 (Springer, Heidelberg, 1995).
3. Lamb, H. F. & Edwards, M. E. in *Vegetation History. Handbook of Vegetation Science 7* (Huntley, B. & Webb, T., III) 519–555 (Kluwer Academic Publishers, Dordrecht, 1988).
4. Kienast, F., Schirrmeister, L., Siegert, C. & Tarasov, P. E. Palaeobotanical evidence for warm summers in the East Siberian Arctic during the last cold stage. *Quaternary Research* **63,** 283–300 (2005).
5. Willerslev, E. *et al.* Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* **300,** 791–795 (2003).
6. Jorgensen, T. *et al.* A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Mol. Ecol.* **21,** 1989–2003 (2012).
7. Lydolph, M. C. *et al.* Beringian paleoecology inferred from permafrost-preserved fungal DNA. *Appl Environ Microbiol* **71,** 1012–1017 (2005).
8. Andersen, K. *et al.* Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Mol. Ecol.* **21,** 1966–1979 (2012).
9. Parducci, L. *et al.* Glacial Survival of Boreal Trees in Northern Scandinavia. *Science* **335,** 1083–1086 (2012).
10. Haile, J. *et al.* Ancient DNA chronology within sediment deposits: Are paleobiological reconstructions possible and is DNA leaching a factor? *Molecular Biology and Evolution* **24,** 982–989 (2007).
11. Haile, J. *et al.* Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 22352–22357 (2009).
12. Yoccoz, N. G. *et al.* DNA from soil mirrors plant taxonomic and growth form diversity.

Mol. Ecol. **21,** 3647–3655 (2012).
13. Willerslev, E. & Cooper, A. Ancient DNA. *Proc. Biol. Sci.* **272,** 3–16 (2005).
14. Sønstebø, J. H. *et al.* Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular Ecology Resources* **10,** 1009–1018 (2010).
15. Hebsgaard, M. B. *et al.* The farm beneath the sand—An archaeological case study on ancient 'dirt'DNA. *Antiquity* **83,** 430–444 (2009).
16. Arnold, L. J. *et al.* Paper II - Dirt, dates and DNA: OSL and radiocarbon chronologies of perennially frozen sediments in Siberia, and their implications for sedimentary ancient DNA studies. *Boreas* **40,** 417–445 (2011).
17. Hopkins, D. M. in *Paleoecology of Beringia* (Hopkins, D. M., Matthews, J. V., Jr, Schweger, C. E. & Young, S. B.) 3–28 (Academic Press, New York, 1982).
18. Ritchie, J. C. & Cwynar, L. C. in *Paleoecology of Beringia* (Hopkins, D. M., Matthews, J. V., Jr, Schweger, C. E., Young, S. B. & Stanley, V.) 113–126 (Academic Press, New York, 1982).
19. Guthrie, R. D. *Frozen Fauna of the Mammoth Steppe*. (University of Chicago Press, 1990).
20. Yeates, G. W. Diversity of nematode faunae under three vegetation types on a pallic soil in Otago, New Zealand. *New Zealand Journal of Zoology* **23,** 401–407 (1996).
21. Sohlenius, B. Influence of climatic conditions on nematode coexistence - a laboratory experiment with a coniferous forest soil. *Oikos* **44,** 430–438 (1985).
22. Yeates, G. W. Nematodes as soil indicators: functional and biodiversity aspects. *Biology and Fertility of Soils* **37,** 199–210 (2003).
23. Dufrêne, M. & Legendre, P. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs* **67,** 345–366 (1997).
24. Ruess, L., Michelsen, A. & Jonasson, S. Simulated climate change in subarctic soils: responses in nematode species composition and dominance structure. *Nematology* **1,** 513–526 (1999).
25. Sorensen, L. I., Mikola, J., Kytoviita, M.-M. & Olofsson, J. Trampling and spatial heterogeneity explain decomposer abundances in a sub-Arctic grassland subjected to simulated reindeer grazing. *Ecosystems* **12,** 830–842 (2009).
26. Popovici, I. & Ciobanu, M. Diversity and distribution of nematode communities in

grasslands from Romania in relation to vegetation and soil characteristics. *Appl Soil Ecol* **14,** 27–36 (2000).

27. Hoschitz, M. & Kaufmann, R. Nematode community composition in five alpine habitats. *Nematology* **6,** 737–747 (2004).

28. Poinar, H. N. *et al.* Molecular coproscopy: Dung and diet of the extinct ground sloth Nothrotheriops shastensis. *Science* **281,** 402–406 (1998).

29. Hofreiter, M. *et al.* A molecular analysis of ground sloth diet through the last glaciation. *Mol. Ecol.* **9,** 1975–1984 (2000).

30. Soininen, E. M. E. *et al.* Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Front Zool* **6,** 16–16 (2009).

31. Zimov, S. A., Zimov, N. S., Tikhonov, A. N. & Chapin, F. S. I. Mammoth steppe: a high-productivity phenomenon. *Quaternary Science Reviews* **57,** 26–45 (2012).

32. Owen-Smith, N. Pleistocene Extinctions: The Pivotal Role of Megaherbivores. *Paleobiology* **13,** 351–362 (1987).

33. Austrheim, G. & Eriksson, O. Recruitment and life-history traits of sparse plant species in subalpine grasslands. *Canadian Journal of Botany-Revue Canadienne De Botanique* **81,** 171–182 (2003).

34. Wardle, D. A. & Bardgett, R. D. Human-induced changes in large herbivorous mammal density: the consequences for decomposers. *Frontiers in Ecology and the Environment* **2,** 145–153 (2004).

35. Gusewell, S. N : P ratios in terrestrial plants: variation and functional significance. *New Phytologist* **164,** 243–266 (2004).

36. Cornelissen, J. *et al.* Leaf digestibility and litter decomposability are related in a wide range of subarctic plant species and types. *Functional Ecology* **18,** 779–786 (2004).

37. McLauchlan, K. K., Williams, J. J., Craine, J. M. & Jeffers, E. S. Changes in global nitrogen cycling during the Holocene epoch. *Nature* **495,** 352–355 (2013).

38. van der Wal, R. Do herbivores cause habitat degradation or vegetation state transition? Evidence from the tundra. *Oikos* **114,** 177–186 (2006).

39. Brathen, K. A. *et al.* Induced shift in ecosystem productivity ? Extensive scale effects of abundant large herbivores. *Ecosystems* **10,** 773–789 (2007).

40. Reimer, P. J. *et al.* Intcal09 and Marine09 radiocarbon age calibration curves, 0-50,000 years cal BP. *Radiocarbon* **51,** 1111–1150 (2009).
41. Taberlet, P. *et al.* Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* **35,** e14 (2007).
42. Elven, R., Murray, D. F., Razzhivin, V. Y. & Yurtsev, B. A. *Annotated Checklist of the Panarctic Flora (PAF).* (Natural History Museum, University of Oslo, 2011). at <http://nhm2.uio.no/paf/>
43. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **37,** D5–15 (2009).
44. Bray, J. R. & Curtis, J. T. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs* **27,** 326–349 (1957).
45. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26,** 32–46 (2001).
46. Shepard, R. N. The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika* (1962).
47. Shepard, R. N. The analysis of proximities: multidimensional scaling with an unknown distance function. II. *Psychometrika* (1962).
48. Nekola, J. C. & White, P. S. The distance decay of similarity in biogeography and ecology. *Journal of Biogeography* **26,** 867–878 (1999).
49. Klotz, S., Kühn, I. & Durka, W. *BIOLFLOR*. (Bundesamt für Naturschutz, Bonn, 2002).
50. group, N. D. *IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series 2008-034 NOAA/NCDC Paleoclimatology Program, Boulder CO, USA. The Holocene* (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements

We thank A. Lister, R. D. Guthrie, M. Hofreiter, L. Parducci, for very valuable thoughts and fruitful discussions of our findings and K. Andersen for help identifying possible contamination. We thank T. B. Brand, P. S. Olsen, V. Mirré, L. J. Gillespie, J. M. Saarela, J. Doubt, M.

Lomonosova, D. Shaulo, J. E. Eriksen, Steffi Ickert-Bond ,T. Ager, D. Bielman ,M. Hajibabaei, A. Telka and S. Zimov for help and providing samples. We thank the Danish National Sequencing Centre. This work was supported by the European Union 6th framework project ECOCHANGE (GOCE-2006-036866, coordinated by P. Taberlet), the Danish National Research Foundation (Centre of Excellence to E. Willerslev), the European Regional Development Fund (Centre of Excellence FIBIR and SF0180098s08 to J. Davison, M. Moora, M. Zobel), the Research Council of Norway (191627/V40 to C. Brochmann), the Australian Research Council (DP0558446 to R.G. Roberts), a Marie Curie International Outgoing Fellowship (PIOF-GA-2009-253376 to E.D. Lorenzen) and a Carlsberg Foundation (to M. Vestergård).

Author contributions

The paper represents the joint efforts of several research groups, headed by various people within each group. Rather than publishing a number of independent papers, we have chosen to combine our data in this paper in the belief that this creates a more comprehensive story. The authorship reflects this joint effort. The ECOCHANGE Team designed and initiated the project. EW, MEE, JM, EDL, MV, GG, JH, JC, IGA, PM, DF, GZ, AT, JA, AS, GS, RGR, RDEM, MTPG, and KK collected the samples. GG, RE, AKB, JHS, CB, LG, EC and PT constructed the plant DNA taxonomic reference libraries and provided taxonomic assignments of the sediment data with input from IGA, EB, SB, LSE, MEE, and DM. EDL, MV, JH, LSE, SB, CC, PW, LG, GG and JHS conducted the genetics lab work. TG did the dating. FP, DR, and VN produced and analysed the data concerning the reliability of the *trn*L approach for estimating herbivore diet. JD, MM, MZ, EC, MV, MR, JC, SB, PBP, HB, RR, TM and PT did the analyses. EDL produced the figures. EW wrote most of the text with input from all authors, in particular JD, MM, MZ, EDL, MEE, MV, PBP, DM, KAB, NY, LO, CB, PT, and RDEM.

Author information

All the raw and filtered data concerning plants, nematodes, megafauna and sheep diet are available from the Dryad Digital Repository: http://doi.org/XXXXXX/XXXXXX. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.W. (ewillerslev@snm.ku.dk).

Figure legends

Figure 1. Sample localities. A total of 242 permafrost samples were collected from 21 sites, shown by green dots. Eight ancient megafauna gut and coprolite samples (A–H) are shown by grey hollow circles, seven modern nematode localities by grey hollow triangles.

Figure 2. Taxonomic diversity of arctic plant assemblages during the last 50 kyr. Taxon composition was estimated by high-throughput sequencing of DNA from 242 permafrost samples. A total of 154 molecular operational taxonomic units (MOTUs) were detected. **a**, Index of ambient temperature (continuous line; oxygen isotope concentration, GRIP[50]) and estimated MOTU number (horizontal bars; second-order jackknife), are shown for three palaeoclimatic periods: pre-LGM (> 25 kyr, n = 149), LGM (last glacial maximum; 25–15 kyr, n = 32) and post-LGM (< 15 kyr, n = 61). **b**, MOTU counts recorded uniquely in each palaeoclimatic period and shared among periods. **c**, Modelled decline in similarity (1-Bray-Curtis dissimilarity) between pairs of plant assemblages from the same palaeoclimatic period in relation to the spatial distance separating them.

Figure 3. Proportional abundance of two families (Teratocephalidae - dark; Cephalobidae - light) among the total soil nematode community at contemporary tundra and steppe sites in Yukon, Canada. Letters a–g correspond to sample localities (Fig. 1). Median (central dot), quartile (box), maximum and minimum (whiskers) and outlying values (points) are shown.

Figure 4. Plant growth form composition over time and across sample types, estimated by high-throughput sequencing of DNA from 242 permafrost samples. **a**, Proportions of DNA reads corresponding to taxa exhibiting different growth forms, binned over 5 kyr time intervals. The analysis included all sediment samples except 21 Svalbard samples and three further samples where no growth form information was available. **b**, Number of MOTUs exhibiting different growth forms as a proportion of total MOTU richness in all informative samples for each palaeoclimatic period. **c**, The proportional abundance of forbs in samples from Main River, Siberia (dated 47,100–19,850 yr BP) where megafauna were or were not detected. **d**, Proportions of DNA reads corresponding to different growth forms in megafauna diet, determined from analysis of eight gut and coprolite samples from late Quaternary megafauna species (woolly

mammoth, woolly rhinoceros, bison and horse). Letters A–H correspond to the individual samples (Fig. 1). The 95.4% calibrated age range of each sample is shown; '> 55' indicates that the sample was too old to provide a finite radiocarbon age. **e**, Reliability of the *trn*L approach for estimating forb and graminoid abundance in diet analyses. Sheep were fed with known amounts of forbs (*Trifolium repens*) and graminoids (*Lolium perenne*), and the rumen content analyzed using the same DNA-based approach as implemented above. Orange dots and lines represent the means and standard errors for diets containing different fractions of forbs. The grey line is a linear model fit. Numbers immediately below the columns in **a**, **b**, and **c** indicate sample sizes. Median (central bar), quartile (box), maximum and minimum (whiskers) values are shown in **a** and **c**.

Extended Data Figure 1. Permafrost sample locality details. **a**, Radiocarbon dating chronology for the main section at the Main River site, Russia, from which nearly all Main River samples are derived; **b**, View of the 2009 Duvanny Yar exposure, NE Siberia; **c**, *yedoma* sandy silt in upper *c*. 12 m of the exposure at Duvanny Yar exposure, NE Siberia. A large syngenetic ice wedge (top centre) within the yedoma is truncated by a thaw unconformity at a depth of *c*. 1.9 m below the ground surface, marking the maximum post-glacial thaw depth after deposition of the *yedoma* had ended. Persons for scale, with DNA sediment sample holes to the right of the person on right; **d**, Calibrated radiocarbon date distributions plotted against depth above river level at Duvanny Yar exposure, NE Siberia. Although there are some finite dates below ~20 m, the general curve shape suggests the radiocarbon dating limit occurs at about this level. The two Svalbard sites at **e**, Colesdalen and **f**, Endalen.

Extended Data Figure 2. MOTU characterization and data consistency. (**a**-**c**) Graphs showing the consistency of the DNA-based approach using permafrost samples across the different time periods: **a**, average marker size per sample; **b**, number of reads per sample; **c**, number of taxa per sample. **d**, WebLogos showing the match between the *gh* primers and their target sequences in the main plant families involved in the estimation of the proportions of forbs and graminoids [70].

Extended Data Figure 3. Temporal classification of samples, assemblage variation in time and data robustness. **a**–**d** (top panel), K-means clustering of permafrost plant assemblages: **a** cluster identity of samples derived from pre-LGM, LGM and post-LGM periods for values of k between

2 and 10. Each bar represents a separate sample; different colours reflect different cluster identities **b**, The Calinski-Harabasz criterion for different levels of k. Higher values indicate stronger support for a level of partitioning. **c**,**d** Heat maps showing the proportional occurrence of samples from pre-LGM, LGM and post-LGM periods in different clusters, for k=2 (**c**) and k=3 (**d**). Colours vary from red (low values) to white (high values). **e**–**g** (middle panel) Assemblage variation in time and space: **e,** Nonmetric multidimensional scaling (NMDS) ordination revealed significant variation (Permanova $p < 0.01$) in fossil/ancient plant assemblage composition during the three palaeoclimatic periods; **f**, The effect of spatial distance on similarity when assemblages from different palaeoclimatic periods were compared. The vertical axis represents similarity in floristic composition measured as 1-Bray-Curtis similarity, the horizontal axis depicts ln of distance between sampled communities in kilometres. The greater the spatial distance between pairs of assemblages, the more dissimilar they were. However, the rate of the decay differed depending on which two climatic periods were compared (full model $p < 0.001$). The weakest distance decay in similarity was observed in the case of comparisons between pre-LGM and post-LGM assemblages. Even if pre-LGM and post-LGM samples came from the same geographic area, their floristic compositions were dissimilar; **g**, Results of randomisation tests. Mean proportional composition of different growth form types in LGM and post-LGM samples. The bars around sample means indicate 95% quantiles derived from 999 bootstrap replicates (where bootstrap N was set to the number of samples in the post-LGM data set; see methods for details). **h** (lower panel) Counts of MOTUs exhibiting different growth forms binned over 5 kyr time intervals. The analysis included 218 of the 242 sediment samples, as described in Figure 4. Numbers immediately below the columns indicate sample sizes. Median (central bar), quartile (box), maximum and minimum (whiskers) counts are shown.

Extended Data Table 1. Site information of the 21 permafrost localities (shown in main text Fig. 1).

Extended Data Table 2. Statistics regarding length of the P6 loop amplified with the gh primers [42]for the most important plant families of the two growth forms (graminoids and forbs). These data were estimated from the arctic/boreal database built for this study.

Extended Data Table 3. Locality information of the seven contemporary tundra and steppe sites in Yukon, Canada, which were analysed for nematode faunal composition (shown in main text Fig. 3). Letters in parentheses refer to locality codes used in main text Figs. 1 and 3.

Extended Data Table 4. Proportion of 17 permafrost sediments with sequences of the two indicator nematode families Cephalobidae and Teratocephalidae.

Extended Data Table 5. Herbivorous mammal taxa derived from Main River permafrost samples for which plant data were available.

Extended Data Table 6. Sample information of the eight megafauna gut and coporolite samples (shown in main text Fig. 1).

505 **Methods1.0 Sites and sediment sampling**

506 Site details and related publications are provided in Extended Data Table 1. The sampled sites
507 are generally well characterised stratigraphically, but not all details are published. Complete site
508 and sample information is available from the ECOCHANGE database manager
509 (H.A.Binney@soton.ac.uk) and the Dryad database (also see section 8.0 for further details on
510 sites). The samples are of mixed provenance: the majority of samples representing the pre-LGM
511 (n = 149) and LGM (n = 32) come from exposures of frozen 'ice-complex' deposits, in which the
512 clastic component (silt and fine sand) derives mainly from aeolian deposition and surface runoff
513 in terrestrial permafrost settings characterized by ice-wedge polygons (e.g. [23]), whereas most of
514 the post-LGM samples (n = 61) come from modern soil and peat (37%), aeolian sediment (30%),
515 thermokarst-lake infill (13%) and fluvial terrace (11%) sequences, and a few samples of mixed
516 origin (9%). In most cases, frozen sediment samples were extracted by horizontal drilling using
517 established protocols to guard against sample-based contamination [5,6,53,54] and were kept frozen
518 until they were processed for DNA analyses. A list of samples and age estimates is given in
519 Supplementary Data 1.

---

520 In July 2009 we sampled soil from sites representing moist tundra and steppe vegetation from
521 seven different locations in Yukon Territory, North-western Canada (Extended Data Table 3).
522 Intact soil cores were excavated in 15 or 30 cm (depending on depth of the A-horizon or the
523 active layer over the permafrost) PVC tubes with a 5 cm diameter inserted into a hollow steel
524 auger forced vertically into the ground. PVC tubes were closed with close-fitting lids and
525 transported in an electric cooler to Whitehorse where they were temporarily stored at 5°C before
526 they were shipped to Centre for GeoGenetics, University of Copenhagen, for processing. From
527 each soil core the five top and bottom cm were processed. Additionally, the moss layer, when
528 present, was processed from tundra samples. Sample material from each layer was homogenized
529 before subsamples were taken for nematode extraction. Bulk density varies greatly between
530 moss, peat and soil, hence the weight of extracted fractions vary between the different sample
531 materials. Nematodes were extracted from 2.0-10.0 g of sample material for 48 h by a modified
532 Baermann tray method [53]. Nematodes were heat-fixed (80°C) in 4% formaldehyde, and a
533 minimum of 100 individuals per sample was identified to genus or family using a compound
534 microscope at 1000× magnification.

Ancient megafauna intestinal/stomach contents and coprolites were collected directly from permafrost and from permafrost-preserved animals (Extended Data Table 6). Interior parts were sampled for DNA analyses.

2.0 Chronology methods

For the majority of samples, plant fragments and soil matrix organics extracted from sediment samples using protocols described in e.g. [56] were radiocarbon-dated using accelerator mass spectrometry. ^{14}C ages were calibrated with the IntCal09 calibration curve [41], or (in the case of modern or near-modern soil samples) using the record of post-bomb atmospheric ^{14}C concentrations. Modern samples yielded ^{14}C concentrations over 100 pMC (percent modern carbon), which matched variations in the 20th century atmospheric carbon related to nuclear testing and other enrichment [57]. In the case where a series of ages from one profile was available, age-depth models were calculated, using the free-shape algorithm published by [58], allowing undated samples to be assigned ages. Age models were only applied to sequences for which stratigraphic evidence supported continuous accumulation of sedimentary units. For a few sequences with previously ascribed dates, calibrated ages were assigned based on the calibration routine available at www.neotomaDB.org. ECOCHANGE radiocarbon ages and supporting information are contained within the dating table of the ECOCHANGE meta-database (see above).

3.0 DNA extraction, amplification and sequencing

DNA extraction of permafrost samples and coprolites and intestinal/stomach contents followed the protocols of [11,59,60]. For construction of the new northern boreal plant reference library, DNA was extracted from leaves taken from taxonomically verified museum specimens originating from across the circumboreal region and sequenced for the plastid *trn*L intron, following the protocols of [14].

3.1 Amplification of plant DNA from sediments

For the ancient plant DNA from sediments, PCR amplification was done using nine base-pair tagged generic plant primers [41] for the P6 loop of the *trn*L plastid region (GH primers). We did

not use the two standard barcoding markers *rbc*L and *mat*K in this study, despite the extensive reference database available, as they are not appropriate for working with degraded DNA, as demonstrated in [12]. First, these markers are too long (*c*. 500 bp for rbcL and 800 bp for matK) for reliably amplifying degraded DNA, and second, it is not possible to shorten them by designing versatile primers on protein-coding genes. Hence some short amplification products can be obtained, but with a strong bias among plant groups according to the variations of the primer target sequences.

Each *trn*L tag was distinguished from any other tag by at least three base differences. The list of tags was generated using the oligoTag program [61]. In order to increase the taxonomic resolution of the analysis for three plant families, three additional primer pairs were used:

ITS1-F: GATATCCGTTGCCGAGAGTC [62]

ITS1Poa-R: CCGAAGGCGTCAAGGAACAC [62]

ITS1Ast-R: CGGCACGGCATGTGCCAAGG [62]

ITS1Cyp-R: GGATGACGCCAAGGAACAC, this study.

They target the first internal transcribed spacer (ITS1) of nuclear ribosomal DNA in Poaceae (ITS1-F and ITS1Poa-R), Asteraceae (ITS1-F and ITS1Ast-R) and Cyperaceae (ITS1-F and ITS1Cyp-R). These primers were tagged in the same way as the P6 loop primers to allow the assignment of sequence reads to the relevant sample. PCR conditions followed the protocol of [12]. Each permafrost and modern soil sample was amplified five times with the *gh* primer pair and once with each of the ITS1 primer pairs. Amplicons were sequenced using the Illumina GA IIx platform as 2 x 108 base pairs (bp) pair end reads.

3.2 Amplification of plant DNA from coprolites and intestinal/stomach contents

DNA amplifications were carried out with the *trn*L *gh* primers [17] with MID incorporated tags. For each sample, PCR was carried out twice with the same-tagged primers and with the use of HiFi (Invitrogen) polymerase and 5ul of extract with 50 cycles of PCR. PCR products were pooled equimolarly and subsequently sequenced on the Roche FLX DNA sequencing platform

(Copenhagen) following previously established protocols [12]. 16SMamm1 and 16SMamm2 primers [63] were used to PCR DNA from the environmental faeces extracts, and the amplicons cloned in order to identify the species of origin.

3.3 Amplification of megafauna DNA from sediments

For megafauna DNA in permafrost, PCRs were performed in the ancient DNA laboratory of the Natural History Museum at the University of Oslo, using the 16Smam1 and 16Smam2 primers [64] and a human-specific blocking primer (16Smam_blkhum3 [59]). Fusion primers containing the Lib-L forward and reverse primers (Roche 454) were used, and 16Smam1 included a Multiplex Identifier (MID) sequence to allow multiplexing of PCR products for sequencing. PCR mixture and profile were as described in [59]. All samples, including extraction blanks, were amplified a maximum of six times in an attempt to obtain two positive PCR replicates, where positive PCRs are those that produced a visible band of the correct size on an agarose gel. When successful, the two PCR replicates were combined, and purified and normalised together using Sequalprep™ Normalisation plates (Invitrogen).

The purified PCR products were sequenced on three machines following the manufacturer's guide for amplicon sequencing. All the plant *trn*L introns and ITS products were sequenced on the Illumina GA IIx platform, the Norwegian Sequencing Centre was used for sequencing of megafauna DNA (Roche 454 GS FLX Titanium).

3.4 Amplification of nematode DNA from sediments

For nematodes, PCR amplification was attempted on a subset of samples using two primer sets. The Cep (fw primer CepF: 5'-CCGATAACGAGCGAGACTC-3', rv primer CepR 5'-CGGCTAAACACCGAAAATCC-3') and Ter (fw primer TerF: 5'-GCTCTCAAGGTGTATATCGC-3', rv primer TerR: 5'-AAACCAGCAGTATTAGCC-3') primers target a 90 bp region of the 18S rDNA of the Cephalobidae and a 118 bp region of the 18S rDNA of the Teratocephalidae, respectively. All primers were flanked by the Lib-L fw and rv primers (Roche 454), and the 5' primers were further flanked by an 8-bp DNA tag [65]. PCRs were performed with 2 μl template DNA in a mixture described by [6] under the following conditions: initial denaturation at 94°C for 5 min, followed by 65 cycles of denaturation at 94°C

for 30 s, annealing at 52°C or 50°C for Cep and Ter primers, respectively, for 30 s, and extension at 68°C for 30 s. Cycling was completed at 72°C for 7 min. PCR products of the correct size (checked on a 2% agarose gel) were purified using the QIAquick Gel Extraction kit (Qiagen) according to the manufacturer's protocol. PCR reactions were repeated at least three times; five times for samples that failed to produce amplicons using either of the primer pairs.

4.0 Taphonomy and contamination issues

[66] recently emphasized the need to understand the taphonomy of a palaeo-proxy system. Here we further assess taphonomic bias and possible contamination of the samples. Fossil assemblages do not represent life assemblages exactly due to post-mortem processes, including differential decomposition, depositional changes, and addition of removal of material [63]. Our landscape-scale taphonomic model for plant DNA derives it from *in situ* burial of above- and below-ground plant parts, downslope transport of material in above-ground and below-ground flow as particles or with DNA as part of soil-water colloidal complexes, and possible deposition from a vector such as animals or wind. Tests in Svalbard (ECOCHANGE, unpublished data) indicate that local (3-50m^2) sources provide almost all plant DNA in modern soils.

For *yedoma*, the surface vegetation was rooted in an accreting substrate that had insufficient time for full profile development prior to burial and freezing (e.g., inceptisols, see [67]. The active layer (estimated at ~50 cm for the LGM of Alaska by [67] acts as a time-averaging moving window, with penetration of unfrozen material to a level by roots potentially occurring until the freezing front reaches that level. We estimate that most yedoma samples record DNA over ~1000 yr of accumulation, but with a bias toward the first few hundred years, this based on observations on how deeply roots penetrate modern soils and average accumulation rates of sediment. We also tested for differences in accumulation rate between time periods that might lead to bias in diversity estimates [68] (Supplementary Data 1). There was no significant difference between pre-LGM and LGM rates (1.12 and 1.25 mm yr^{-1}, respectively). The post-LGM had significantly greater average rates (3.82 mm yr^{-1}), but this estimate is based on only a few sites and samples and more diverse forms of sedimentation; furthermore, beta diversity increases, rather than diminishes, as would be expected if there were bias, in the post-LGM. We conclude that time-averaging effects in our samples have not biased the diversity estimates.

Peat and lacustrine sediment samples will have finer (decadal) temporal resolution as demonstrated by numerous other proxy studies, and the loess-derived sediments sampled beneath the rapidly deposited Dawson tephra at Quartz Creek (see section 8.3) may be time-averaged over only decades or centuries. The few samples drawn from thermokarst lake deposits could potentially include a wider age range of material derived from lake-bank collapse, but the Holocene ^{14}C chronologies suggest that these sequences can be reliably compared with the late-Pleistocene records.

The study by [12] showed that graminoid DNA occurs in soil in about the same proportions as graminoids occupy the above-ground biomass. We might expect woody plants to release environmental DNA at a lower rate in relation to their above-ground biomass as much of their production goes into woody stems and roots, which have a relatively slow rate of decomposition; this is in the case, with woody taxa, when at low proportions in the biomass, being under-represented in DNA by a ratio of approximately 5:1. Pollen and macrofossil data from numerous sites including our own attest to the rarity of woody taxa in the pre-LGM and LGM periods. In these two periods, woody taxa are likely under-represented in our DNA record, but even allowing for this they still form a minor component of all assemblages.

[12] show forbs to be represented in DNA compared with above-ground biomass at a ratio of about 2:1. [12] suggest that this difference may reflect different litter turnover rates; graminoids are richer in lignins than are forbs [69]. Alternatively, forbs may invest resources into below-ground parts if they are perennials while others (not many in the Arctic) are annuals and largely decompose every year, yielding a range of root-shoot ratios [70]. It is unlikely that differential preservation of ancient forb tissue has occurred because this would predict a lessening of forb dominance through time; rather there is continuous forb dominance through the pre-LGM and LGM and an abrupt diminution of forb DNA in the post-LGM. Further, there is no bias in the length of sequences recovered through time (see below), which could otherwise conceivably generate a bias as some of the longer *trn*L sequences occur in the Cyperaceae.

Established protocols for permafrost sampling were followed to control for sample-based contamination [5,6,11,53,54]. All ancient pre-PCR work (i.e. sub-sampling, extraction, and PCR set up) was conducted in full body suits in state-of-the-art dedicated ancient DNA laboratories in

Copenhagen and Oslo that are physically separated from any other biological laboratories, with positive air pressure and nightly UV-exposure of surfaces, and equipped with positive flow hoods. Occasionally, common contaminants were detected: *Homo sapiens, Mus musculus, Sus, Bos*, *Canis, Felis cattus, Solanum lycopersicum, Zea mays* and *Cedrus*. The current control setup does not allow contamination of individually tagged PCR products to be detected. To mitigate this problem we removed haplotypes which have previously been detected as contaminants in PCR reagents, are exotic to the study sites, or represented likely artificial diversity caused by sequencing-error-by-products of contaminant or exotic haplotypes. The taxonomic assignment of these sequences includes for example Rutaceae, Solanaceae, Solanoideae, Loasaceae and Musaceae. Additionally, the following plant MOTUs occurred in sequencing blanks: Salicaceae (Group 1), containing *Populus* and *Salix*; *Equisetum* (Group 2), containing *E. arvense, E. sylvaticum,* and *E. fluviatile*; and *Taraxacum.* These MOTUs are likely to genuinely occur in the study samples but were excluded as a conservative measure. We also note that *Eritrichium,* (Group 1) Triticeae (Group 1), containing *Elymus* spp., *Leymus* spp., Apiaceae (Group 1), *Betula* (Group 1), *Dryas* (Group 1) though not found in the bank controls of this study, have been recorded as possible sources of contamination in other studies.

Importantly, to avoid possible contamination from re-deposition of organics or DNA in the exposures sampled, we did not include any low-abundance sequences in the analyses (see below), as such sequences may be due to re-deposition of material [16].

For further evidence of reliability of results and their interpretations please see section 5.3 MOTUs characterization and data consistency.

All the raw and filtered data concerning plants, nematodes and megafauna are available from the Dryad Digital Repository: http://doi.org/XXXXXX/XXXXXX.

5.0 Plant DNA reference libraries, sequence groupings and MOTU characterization

5.1 DNA reference libraries

We identified plant sequences retrieved from the ancient samples taxonomically using (i) the arctic plant *trn*L reference library developed by [14], comprising 842 species representing all

widespread or ecologically important taxa of the circum-arctic flora, (ii) a new extension of this library constructed by sequencing the nuclear ribosomal ITS1 region to improve species resolution in three families (Cyperaceae, Poaceae and Asteraceae), (iii) a new north boreal plant *trn*L reference library constructed by sequencing DNA extracted from 1332 herbarium specimens representing 835 of the most common north circumboreal species, of which most also occur in present-day arctic vegetation, and (iv) the EMBL database for sequences not matching taxa contained in these three reference libraries. The specimens used to construct the new north boreal library were sampled after taxonomic verification from the following collections: Herbarium of the Natural History Museum, University of Oslo, Norway (O); Popov Herbarium, Siberian Central Botanical Garden, Novosibirsk, Russia (NSK); National Herbarium of Canada, Canadian Museum of Nature, Ottawa, Canada (CAN); and University of Alaska Museum of the North (ALA). Quality checking and cleaning of this new library was performed by comparing all sequences with published sequences using NCBI/BLAST and by phylogenetic analyses of each family, including sequences from closely related taxa to verify taxonomic identity. All reference databases are available from the Dryad Digital Repository: http://doi.org/XXXXXX/XXXXXX.

5.2 Sequence groupings and identifications of sedimentary plant DNA

For plant DNA data obtained from the sediment samples, each pair of reads was assembled to reconstruct full-length marker sequence using the Solexapairend program from the OBITools package (http://metabarcoding.org/obitools). Sequences were associated with their corresponding sample according to the primer tags, and identical sequences were clustered to form molecular operational taxonomic units (MOTUs). MOTUs occurring less than five times in the whole data set or containing ambiguous base symbols were discarded. Only PCR repeats with more than 1000 sequences for the *gh* primers and 500 sequences for the ITS1 primers were considered for the following process. For *gh* PCR amplification, a MOTU was considered as belonging to a sample if it occurred in the majority of the usable repeats for this sample. Taxonomic assignment of MOTUs was done with the ecoTag program [12] using our plant reference libraries as reference databases: Only MOTUs having at least 95% similarity with a sequence in one of the reference libraries or in the EMBL database were kept in the final dataset. Identifications realized with our reference libraries were given priority over EMBL. The final set of MOTUs associated with a sample was based on all MOTUs retrieved from all repeats of this sample. Initial identifications

to the species level were in some cases adjusted to a higher taxonomic level based on the completeness of our reference libraries. Results are listed in Supplementary Data 2, 4 and 5.

5.3 MOTUs characterization and data consistency

Basic statistics were used to check data consistency among time periods. Results are presented in Extended Data Figure 2a-c, and clearly show that older samples did not present any bias compared with more recent samples. A bias could have been introduced (i) if the size of the *trn*L P6 loop would have been smaller in taxa identified in older samples, (ii) if the number of identified taxa were smaller in older samples, or (iii) if the number of sequence reads were lower in older samples. This was not the case and we conclude that the reconstructed plant assemblages from different time periods did not suffer from such biases.

We also checked if the primers used could explain the differences observed between forbs and graminoids. The WebLogos [51] presented in Extended Data Figure 2d show that the target sequences of the *trn*L *gh* primers [42] are very well preserved in the main families leading to the estimation of the relative proportions of forbs and graminoids. According to the very good match of the *gh* primers in the different families, it is highly unlikely that these minor differences can produce any significant bias in the observed proportions of forbs and graminoids.

Finally, we carried out length statistics of the P6 loop of the *trn*L intron for several plant families (Extended Data Table 2), knowing that shorter sequences are likely to be preferentially amplified than longer sequences. According to the mean length in the different families, Cyperaceae (graminoid) might be under-represented in our results, and Plumbaginaceae (forb) and Polygonaceae (forb) over-represented. In any case, the bias was identical for all samples (permafrost and diet), and for all periods as no size difference among the amplified sequences were observed among period (Extended Data Figure 2a-c). For all the other families, the size difference is minor, and is unlikely to generate any significant bias.

5.4 Reliability of the *trn*L approach for estimating the diet of herbivores

To test the reliability of the *trn*L approach for estimating the diet of herbivores, we conducted an experiment on sheep. During the period of May-July 2011, pure plots of white clover (*Trifolium*

repens, cv Merwi) and ryegrass (*Lolium perenne*, cv Aberavon) were used to test five mixtures of green fodder (i.e. five diets differing by their clover:ryegrass ratios of 0:100, 25:75, 50:50, 75:25 and 100:0).

The five diets were allocated to five 1-year-old Texel sheep fed *ad libitum*. For each sheep and each diet, one rumen sample was collected on 2 successive days. The collection started 13 days after the beginning of the diet in order to prevent from an effect of the previous diet. Each of the 50 samples consisted of about 5 g of rumen content.

Total DNA was extracted from about 25 mg of rumen content with the DNeasy Blood and Tissue Kit (QIAgen GmbH, Hilden, Germany) following the manufacturer's instructions. The DNA extracts were amplified with the *trn*L *gh* primers (g: GGGCAATCCTGAGCCAA; h: CCATTGAGTCTCTGCACCTATC [24]) targeting a short portion of the *trn*L intron of the chloroplast DNA. For each sample two independent PCR replicates were carried out. Paired-end sequencing (100 nucleotides on each extremity of the DNA fragments) was carried out at the French National Sequencing Centre (CEA Genoscope, Evry, France) on a Illumina HiSeq 2000 (Illumina Inc.).

A total of 216,586 and 163,328 sequence reads corresponded to *Trifolium repens* (forb) and to *Lolium perenne* (graminoid), respectively. The Pearson correlation coefficient between the actual fraction of forb in diet and the proportion of forb estimated using the DNA-based approach is highly significant (r^2=0.75, $p < 10^{-15}$) (Fig. 4e).

All the data concerning the sheep diet experiments are available from the Dryad Digital Repository: http://doi.org/XXXXXX/XXXXXX.

6.0 Analysis of MOTU assemblage data

Each sediment sample provided a molecular characterization of a local plant assemblage. To analyse gross changes in plant assemblages through space and time we used 242 dated samples from 21 sites (56 entities, i.e., individual sections), which provided a total of 7,738,725 chloroplast *trn*L (UAA) intron reads. For these analyses we used only the MOTUs identified with the *gh* primers (see Section 3.0), because the reads of these MOTUs are proportional to

vegetation (see [12]). In total, 154 taxa (MOTUs) were identified, of which 47 were assigned to species level (Supplementary Data 4). Supplementary Data 5 lists the MOTUs and constituent taxa for the ITS identifications.

6.1 Temporal classification of samples and data robustness

Each sample was allocated to one of three broad age categories: (i) 50–25 thousand years ago (kyr; pre-LGM), a period of fluctuating climate [71]; (ii) 25–15 kyr, the Last Glacial Maximum (LGM), a period of constant cold and dry conditions [17]; (iii) 15 kyr–present, the current interglaciation (post-LGM), which, subsequent to deglacial warming, is characterised by climate stability and relatively high temperatures [71]. Our specification of LGM timing represents a period between the transition of Marine Isotope Stage (MIS) 3 to MIS 2 and the transition to the Bølling (Gi-1e). This time window incorporates the period of lowest global sea level, which is traditionally used to define the LGM (22–18 kyr), along with flanking periods during which the development of glaciation or deglaciation occurred. The use of a fairly wide window was also intended to allow for some regional variation in the timing of the maximum. We assessed the robustness of our analyses to alternative definitions of LGM timing using Permanova (implemented using R package vegan [72]) to test the fit of models including LGM specifications with different duration and timing, falling in the range 30–11 kyr. In general, there were not large differences between many of the alternative definitions, and all detected the large shifts in plant assemblages occurring around that time (Supplementary Data Table 3). To assess whether our temporal definition of post-LGM masked changes prior to and including the onset of the Holocene at ~11 kyr, we extracted the post-LGM subset of data, i.e. 15–0 kyr, and used Permanova to test whether splitting the data into two time periods (15–11 and 11–0 kyr) improved the fit. The results indicated that given the data we have, the split of post-LGM into two consecutive time bins did not significantly improve the null model ($P = 0.08$).

We compared our approach of defining *a priori* groups based on radio carbon dating with an unsupervised approach whereby variation between samples was used to define groups. To partition samples into clusters we used k-means clustering with the Hartigan-Wong algorithm, values of k between 2 and 10 and 100 random starting configurations for each value of k. The Calinski-Harabasz criterion was used to identify the best supported values of k [73].The results of

814 unsupervised clustering largely coincided with our supervised analysis (Extended Data Figure 3
815 a-d). The two- and three-cluster solutions, which were best supported, revealed the clearest
816 distinction between post-LGM communities on one hand and pre-LGM and LGM samples on the
817 other. This is in accordance with our diversity analysis, which showed that that the species list of
818 the LGM was essentially a subset of the pre-LGM species list, although considerably fewer
819 species were recorded from LGM samples. The higher values of k indicated more subtle
820 differences between LGM and pre-LGM samples.

821 As a further investigation of data robustness, we repeated the analyses, but imposed an upper
822 limit of 40 kyr to the pre-LGM period and excluded older samples, thus equalizing the duration
823 of the pre- and post-LGM periods (both 15,000 years). The results of these analyses were
824 qualitatively identical to those based on the whole data set. However, while MOTU richness
825 remained highest in the pre-LGM in the equalized analysis, it was less clearly so (equalized
826 analysis: total richness: pre-LGM = 103, LGM = 48, post-LGM = 74; jackknife second order
827 estimator: pre-LGM = 169, LGM = 85, post-LGM = 159).

828 **6.2 Functional characterization of molecular taxa**

829 We characterized MOTUs in terms of their coarse growth form; 147 of the 154 taxa identified
830 could be placed into four primary groups: forbs, graminoids (grasses + sedges + rushes), dwarf
831 shrubs or other woody plants (i.e., shrubs and trees). Information on growth form was derived
832 from BiolFlor, a database covering more than 60 plant species traits for 3659 plant species from
833 the German flora [49,74]. Where data were lacking, we excluded the taxon from analysis.

834 **6.3 Assemblage variation in time and space**

835 **6.3.1 Ordination**

836 Variation in assemblage characteristics among time periods was visualised using two-
837 dimensional non-metric multi-dimensional scaling (NMDS). The composition of samples was
838 estimated by the proportion of reads corresponding to particular MOTUs.

839 Dissimilarity between pairs of plant assemblages was defined using Bray-Curtis

dissimilarity (BC)[44]. For some analyses similarity was calculated as 1-BC. Bray-Curtis dissimilarity is frequently used in plant community ecology and is recommended by several basic sources due to its properties [75 pg 51] and elsewhere [72,76,77]. In particular, Bray-Curtis shows a good ability to mirror environmental distances [75,78 pg 50-54]. The Bray-Curtis index also works well with proportional abundance data [78,79 pg 287]. Euclidean distance is also widely used with proportional abundance data. While so-called proportion indices like BC depend on the number of shared species and thus measure distance as proportions of the maximum distance possible, Euclidean distance concentrates only on differences in relative proportional abundances [80]. Thus, the choice of distance measure depends on the emphasis of a particular study, e.g. how much attention is paid to different aspects of community assemblage structure. We considered the co-occurrence of taxa in samples to be an important feature of palaeocommunity assembly, and this is why Bray-Curtis was our primary choice. However, since Euclidean distance could add another aspect of community assembly, we performed a parallel analysis (Permanova, NMDS and distance decay) using Euclidean distance. We found that our quantitative results and the qualitative patterns were robust to the choice of distance measure.

First, the ordination was conducted for the whole data set. Second, since the spatial distribution of the total data set was not balanced between time periods, we identified four replicated locations (two in North America, one in western Siberia, one in eastern Siberia) where samples were collected from sites within 100 km of each other in all palaeoclimatic periods. We based a further ordination on an equal number of samples per location per period (15 samples per period, 45 samples in total). Because the results of analyses based on the two data sets coincided, only the results of the first analysis are presented, except in Extended Data Figure 3e where it was impossible to portray all 242 samples and the results of the second analysis are presented. Stress values for the ordinations were in the range 0.05–0.17. Permanova was used to compare the similarity of floristic composition in different periods.

6.3.2 Richness estimation

Nonparametric richness estimators are usually recommended due to their precision and low susceptibility to sampling bias [81]. In particular, the second order jackknife has been shown to be one of the most effective estimators [82,83], especially for highly sparse palaeontological data [84]. We used the second order jackknife to estimate species richness in climatic periods.

6.3.3 Distance-decay measures

We modelled variation in plant communities using a distance-decay in similarity approach [48], using as a dependent variable all pairwise similarities between samples in terms of floristic composition. We used a generalised linear model to describe variation in the dependent variable. The dependent variable was bounded by 0 and contained a large proportion of exact 0s (i.e. achieved when pairs of samples contained no shared taxa). The data were also theoretically bounded by 1, but in practice no samples were identical and the data exhibited a strong positive skew. To adequately model variation in such a dependent variable, we used a compound Poisson error distribution (using R package tweedie [85]), with an index parameter for the power variance function of 1.45 (estimated using maximum likelihood) and a log link function. The geographic distance separating points was included as an independent variable. This distance was calculated as the natural logarithm of the orthodromic distance between points, i.e. calculated as the shortest earth-surface distance between two sets of latitude and longitude coordinates (the earth was assumed to be spherical with a radius of 6371 km). The second independent variable consisted of a categorical variable representing the combination of the time periods being compared. Thus, this variable had six levels, consisting of all pairwise combinations between these periods (pre-LGM *vs* pre-LGM, pre-LGM *vs* LGM, pre-LGM *vs* post-LGM etc.).

An interaction term between the independent variables was included in the model. Since each sample was represented multiple times in the data set, observations were not independent, biasing model estimates of variance and statistical significance. To estimate the true significance of model terms, we recalculated each model a further 999 times using data sets where the community data underlying the dependent variable were randomised (values were permuted within samples using the permatfull function from the R package vegan). The change in deviance associated with dropping a term in the empirical model was then compared to the corresponding statistics derived from randomised models; significance was estimated based on the number of randomised statistics higher than the empirical value.

6.3.4 Randomisation tests used to assess functional changes between time periods

We used a randomisation procedure (BC_{diff}; described in [86]) to assess whether the growth form composition of plant communities of the LGM and post-LGM (target) periods represented a random sample from the directly preceding (source) period. To do this we calculated the BC between the observed mean growth form composition of the target period and each of 999 means derived from a bootstrapped selection (sampled with replacement; the sample size corresponding to that of the target period) of samples from the source period (BC[observed *vs* random]=BC_{or}). In parallel, BC was calculated 999 times between two random means, calculated as described above (BC[random *v*s random]=BC_{rr}). The latter calculation provided a population of BC measures that might be expected to arise by chance. The vector of 999 BC_{rr} values was subtracted from the vector of 999 BC_{or} in a random pairwise manner to produce a final vector of 999 values (BC_{diff}=BC_{or}−BC_{rr}). BC_{diff} has an expected value of 0 if community composition is random.

This approach indicated that LGM growth form structure did not differ from a random draw from the pre-LGM community (95% quantiles of BCdiff; LGM: -0.14–0.11). However, post-LGM growth form composition was not a random subset of that from the LGM (95% quantiles of BCdiff: 0.05–0.30). The abundance of forbs decreased while the abundance of all other growth form types increased in the post-LGM compared with the LGM period (Extended Data Figure 3g).

6.3.5 Overview of vegetation change through time

We classified a subset of samples (those of finite age) into 5000-year age classes (from 50,000–45,000 to 5000–0 kyr) across the region encompassing central and northeast Siberia and Alaska-Yukon. These regions were unglaciated and inhabited by the megafauna in the Pleistocene, and they are the regions from which the dietary samples originated. The samples from Svalbard used in the previous analyses were omitted here as Svalbard was almost entirely glaciated in the LGM and did not host megafauna. We plotted the abundance of the key groups (described above) as estimated by the abundance of DNA sequence reads through time to provide an overview of their shifting importance. We also calculated the number of MOTUs detected for each group through time (Extended Data Figure 3h).

7.0 Filtering and taxonomic inference of nematode and megafauna data

7.1 Nematode data

Nematode sequences were sorted according to the DNA tag used. Within individual PCR products, sequences represented by less than five reads were discarded. The remaining sequences were assigned to taxa using the statistical assignment package SAP [87].

We used Dufrêne–Legendre indicator species analysis [23] to identify nematode taxa that acted as good indicators of modern tundra or steppe habitat (as implemented by the indval function from the R package labdsv [88]). The function calculates an indicator value for each taxon that is the product of its relative frequency and relative average abundance in sample groups (the groups in this case being steppe and tundra). The value varies from 0 to 1 and would be maximal if all examples of a taxon were distributed among all samples from only one of tundra or steppe. By morphologically determining the nematode faunas of 35 sediment samples from contemporary tundra and steppe sites in Yukon, Canada, we discovered two indicator families: Teratocephalidae for tundra and Cephalobidae for steppe. We tested whether the proportion of the two families differed between tundra and steppe with a nested ANOVA (site nested within vegetation type) (SAS Enterprise Guide, version 4). Data on proportions were square root transformed to obtain homogeneity of variance (Bartlett test). The ANOVA was executed on non-normally distributed data, but the ANOVA is quite robust to non-normality [89]. We genetically determined the presence of the two indicator families in 17 of the 242 ancient sediment samples; results are listed in Extended Data Table 4.

7.2 Ancient megafauna sediment data

Sequences were filtered and sorted using the programs included in the OBITools package (http://metabarcoding.org/obitools). For filtering, only reads containing both primers and the tag were kept in the data, permitting two errors in the primers and no errors in the tags. Filtering and taxonomic identification was performed as described in [59] with the following two adjustments: (i) an additional denoising step using the program Obiclean was included [90], and (ii) the electronic PCR was performed on the EMBL standard sequences release 111. Within each sample, only sequences represented by > 10 reads and an identification to at least genus level

with an identity > 0.95 were kept in the final dataset. Identified taxa in each of the samples for which plant data are available are given in Extended Data Table 5.

7.3 Ancient megafauna diet data

The plant DNA amplified from coprolites and intestinal/stomach contents was sorted using the OBITools package (http://metabarcoding.org/obitools). Sequences shorter than 10 basepairs, or containing ambiguous nucleotides, or with occurrence ≤ 5 were excluded. Strictly identical sequences were merged and taxonomic assignation was achieved using the ecoTag program and reference libraries described in sections 5.1 and 5.2.Only unique sequences with an identity of 100% to at least one of the reference sequences were kept for further analysis. Where 100% identities were obtained from multiple reference libraries, priority was given to taxon assignment using the Arctic and boreal libraries.

We obtained a total of 15,951 sequence reads that could be assigned to the eight coprolite/gut samples using the MID tags, of which 1,663 reads were unique. Out of these reads, 13,735 passed filtering and a final 9,084 reads could be assigned with 100% identity to a plant species in one of the reference databases. Sequence data and compositional data for the fossil diet samples are given in Supplementary Data 6.

8.0 Permafrost site information

8.1 Published sites, Eurasia

Bol'shaya Balakhnaya, Buor Kaya and Khatanga, NW Siberia

Three locations in NW Siberia with perennially frozen deposits are described in [16]. Buor Kaya is located on the east side of the bay formed by the Lena Delta, is a 3-m exposure of sandy silt with organic inclusions, interpreted as lacustrine sediment, Holocene in age. Khatanga material was sampled from Holocene river terrace deposits (< 5m) along a small tributary stream. Material ranged from clay, to weakly laminated sands and silts, to peat. Bol’shaya Balakhnaya is also a Holocene fluvial terrace locality featuring weakly laminated sands and minor interspersed lenses of peat and clay.

Baskura Peninsula, Cape Sabler, Federov Island, Ovrazhny Peninsula, and Upper Taymyr River, Taimyr Peninsula, NW Siberia

These localities are described by [6], who report perennially frozen sediments taken along the shore of Lake Taimyr. Deposits are silt-dominated but range from organic to inorganic, and massive to laminated; all sediments are of late-Pleistocene age (~40 to 12 kyr). Further stratigraphic information from the "type locality" of this type of sediment – the Cape Sabler site – is provided in [91]. Sediment depth/age curves in [6] show that depositional rates were in the order of 1–2 mm/yr. This implies a high temporal resolution of the trapped macroflora elements and other biogenic matter, as the ground surface rose due to the vertical accretion of silt and fine sand that was transported and deposited by aeolian and surface runoff processes.

Main River, E Siberia

The Main River (Ice Bluff) exposure extends for about 1 km at an elevation of 30 m on the left bank of the Main River. It has been previously reported by [91] and [93]. The northern exposure, from where our samples are derived, is dominated by ice-rich deposits interpreted as a facies of *yedoma* by [92]. At the time of sampling, the lower portion of the exposure was covered by slump material; the oldest exposed deposits are ~40 kyr. We dated further samples to improve the previously established chronology of the site (Extended Data Figure 1a). Samples form a consistent progression suggesting continuous sedimentation without major hiatus between ~40 and 20 kyr.

8.2 Unpublished sites, Eurasia

Taimyr Lake, Taimyr Peninsula, NW Siberia

A 3 m high cliff section at the western side om the Cape Sabler Peninsula. Vaguely laminated silt with some sand intrabeds. Four radiocarbon dates suggest a mid-Holocene age between 4.7-7.1 kyr for the sediment sequence, except for the uppermost sample that is modern in age.

Anadyr, E Siberia

Holocene deposits, beside the Anadyr River, 2 km West of Anadyr, Chukotka. Materials excavated from a pit lying 3.0–5.1 m above sea level.

Duvanny Yar, NE Siberia

The site is the type section for the late Pleistocene in NE Siberia and has been much studied (e.g. [94-96]). The extensive set of exposures runs for ~4 km along the east bank of the Kolyma River and features high cliffs of *yedoma* (ice complex), dominated by silt and large syngenetic ice wedges, depressions representing the drained basins of thermokarst lakes (alasy), and large areas of slumped and partially vegetated material. The exposure we studied and sampled in 2009 (Extended Data Figure 1b) is from the centre of remnant 7E of the *yedoma* surface identified by [94]. We levelled in and logged 23 sections and sampled for DNA, radiocarbon and palaeoecological analysis from just above the Kolyma River level to ~40 m above it.

The sampled stratigraphic unit comprised *yedoma* sandy silty at least 34 m thick, underlying a thaw unconformity at a depth of ~1.9 m below the ground surface (Extended Data Figure 1c). The *yedoma* unit was characterized by grey sandy silt to silty fine sand with low and varying amounts of organic matter, the most prominent of which were abundant fine *in situ* roots pervasive throughout the unit. The sediment is interpreted primarily as loess and contains a number of weakly developed palaeosols (J.B. Murton unpublished data). The upper 1.9 m of the sedimentary sequence comprised the post-glacial transition zone and overlying modern active layer.

A ^{14}C age-depth model is presented in Extended Data Figure 1d. The upper part of the model, above an elevation of 20 m above river level, is considered to be robust, based on ^{14}C ages that decrease overall in stratigraphic order towards the top of the unit. ^{14}C ages from below 15–20 m above river level are close to the limit of radiocarbon dating, and the age-depth model of this lower part of the *yedoma* should be treated as less definitive, although supported by OSL age at 14.5 m. The basal units of the exposure are not represented in this study.

Svalbard: Colesdalen and Endalen

Samples were taken from the upper organic horizon of tundra soils in two valleys directly into sterile tubes and sealed. Sites are Colesdalen and Endalen (Extended Data Figure 1e,f). Both valleys have vegetation dominated by mid-Arctic tundra.

8.3 Published sites, North America

Zagoskin Lake, Alaska

Zagoskin is a maar lake in western Alaska with sediments dating from ~37 kyr BP. Details are reported in [97,98]. The sediments are relatively inorganic and dominated by silt, interpreted as loess [98]. Loss on ignition values are generally < 10%, except in the top 1.4 m of the 15-m section. Biostratigraphic changes related to deglaciation (~15 kyr BP) are recorded at 5 m depth. Sediments dating to the LGM are present.

Quartz Creek, Yukon Territory, Canada

This locality is described by [99]. Sections are exposed in mining cuts and comprise silt-rich facies and palaeosols. The silt is loess-derived and sometimes finely bedded, reflecting re-working. Samples are associated with the Late Pleistocene Dawson tephra [99,100] and immediately underlie the bed, ranging from 31 kyr BP to 30 kyr BP, consistent with the Dawson tephra chronology.

Goldbottom, Yukon Territory, Canada

This locality also comprises several exposures in mining cuts and is described in [101]. Frozen silt-dominated sediments, interpreted as loess or retransported loess, and organic deposits are present, and the Dawson tephra provides a late MIS 3-ealry MIS 2 stratigraphic marker. Previous dating of the tephra at *ca.* 30 kyr BP [101,102] are consistent with ages associated with the samples in this study. Samples at the site range include pre-LGM samples (*ca.* 45 to 27 kyr BP) and early LGM ages (ca. 24 to 23 kyr BP).

Stevens Village, Alaska

The locality is described in [11]. The exposure is ~15 m high and lies on the Yukon River in central Alaska. Frozen silt, interpreted as loess, overlies basal fluvial gravel and contains interbedded

organic layers (regosols) dating to the early Holocene. A single sample collected from an early Holocene soil dates to 11.2 kyr BP.

8.4 Unpublished sites, North America

Purgatory, Alaska

The Purgatory site is located a few kilometres upstream from the Stevens Village site and consists of aeolian sands with plant detritus. Two samples from near the base of the exposure date to the post-LGM interval.

Ross Mine, Canada

The Ross Mine site is located in the southern Klondike goldfields of central Yukon. One sample from a floodplain silt unit within fluvial deposits dates to the LGM, while the remaining samples, collected from within a Holocene peatland date to the post-LGM interval.

References

51. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Research* **14,** 1188–1190 (2004).

52. Schirrmeister, L. *et al.* Sedimentary characteristics and origin of the Late Pleistocene Ice Complex on north-east Siberian Arctic coastal lowlands and islands - A review. *Quaternary International* **241,** 3–25 (2011).

53. Willerslev, E. *et al.* Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* **317,** 111–114 (2007).

54. Epp, L. S. *et al.* New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Mol. Ecol.* **21,** 1821–1833 (2012).

55. Vestergård, M. Nematode assemblages in the rhizosphere of spring barley (Hordeum vulgare L.) depended on fertilisation and plant growth phase. *Pedobiologia* **48,** 257–265 (2004).

56. Brock, F., Higham, T., Ditchfield, P. & Ramsey, C. B. Current pretreatment methods for AMS radiocarbon dating at the Oxford Radiocarbon Accelerator Unit (ORAU). *Radiocarbon* **52,** 103–112 (2010).

57. Hua, Q. & Barbetti, M. Review of tropospheric bomb 14C data for carbon cycle modeling and age calibration purposes. *Radiocarbon* **46,** 1273–1298 (2007).

58. Goslar, T., Van der Knaap, W. O., van Leeuwen, J. & Kamenik, C. Free-shape 14C age–depth modelling of an intensively dated modern peat profile. *Journal of Quaternary Science* **24,** 481–499 (2009).

59. Boessenkool, S. *et al.* Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. *Mol. Ecol.* **21,** 1806–1815 (2012).

60. Haile, J. in *Methods in Molecular Biology – Ancient DNA* (Shaprio, B. & Hofreiter, M.) 57–63 (Humana Press Series, 2012).

61. Coissac, E. E. OligoTag: a program for designing sets of tags for next-generation sequencing of multiplexed samples. *Methods Mol Biol* **888,** 13–31 (2012).

62. Baamrane, M. A. A. *et al.* Assessment of the food habits of the Moroccan dorcas gazelle

in M'Sabih Talaa, west central Morocco, using the trnL approach. *PLoS ONE* **7,** e35643 (2012).

63. Birks, H. J. B. & Birks, H. H. *Quaternary Palaeoecology*. (London Edward Arnold, 2004).

64. Taylor, P. G. Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. *Molecular Biology and Evolution* **13,** 283–285 (1996).

65. Binladen, J. *et al.* The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE* **2,** –e197 (2007).

66. Jackson, S. T. Representation of flora and vegetation in Quaternary fossil assemblages: known and unknown knowns and unknowns. *Quaternary Science Reviews* **49,** 1–15 (2012).

67. Höfle, C. & Ping, C.-L. Properties and soil development of late-Pleistocene paleosols from Seward Peninsula, northwest Alaska. *Geoderma* **71,** 219–243 (1996).

68. Tomašových, A. & Kidwell, S. M. Predicting the effects of increasing temporal scale on species composition, diversity, and rank-abundance distributions. *Paleobiology* **36,** 672–695 (2010).

69. Cornelissen, J. H. C. *et al.* Global negative vegetation feedback to climate warming responses of leaf litter decomposition rates in cold biomes. *Ecol Lett* **10,** 619–627 (2007).

70. Aerts, R. & Chapin, F. S. I. The mineral nutrition of wild plants revisited: a re-evaluation of processes and patterns. *Advances in Ecological Research* **30,** 1–67 (1999).

71. Dansgaard, W. *et al.* Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364,** 218–220 (1993).

72. Oksanen, J. Multivariate analysis of ecological communities in R: vegan tutorial. *R package version 1.17-7* (2011).

73. Caliński, T. & Harabasz, J. A dendrite method for cluster analysis. *Communications in Statistics* **3,** 1–27 (1974).

74. Kühn, I., Durka, W. & Klotz, S. BiolFlor - a new plant-trait database as a tool for plant invasion ecology. *Diversity and Distributions* **10,** 363–365 (2004).

75. McCune, B., Grace, J. B. & Urban, D. L. Analysis of ecological communities. (2002).

76. Clarke, K. R., Somerfield, P. J. & Chapman, M. G. On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray–Curtis coefficient for denuded assemblages. *Journal of Experimental Marine Biology and Ecology* **330,** 55–80 (2006).
77. Austin, M. P. Inconsistencies between theory and methodology: a recurrent problem in ordination studies. *Journal of Vegetation Science* **24,** 251–268 (2012).
78. Faith, D. P., Minchin, P. R. & Belbin, L. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* **69,** 57–68 (1987).
79. Legendre, P. & Legendre, L. *Numerical Ecology*. (Elsevier, Amsterdam, 1998).
80. Anderson, M. J., Connell, S. D. & Gillanders, B. M. Relationships between taxonomic resolution and spatial scales of multivariate variation. *J Anim Ecology* **74,** 636–646 (2005).
81. Gotelli, N. J. & Colwell, R. K. in *Biological diversity: Frontiers in measurement and assessment* (Magurran, A. E. & McGill, B. J.) 39–54 (Oxford University Press, Oxford, 2011).
82. Walther, B. A. & Moore, J. L. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography* **28,** 815–829 (2005).
83. Hortal, J., Borges, P. A. V. & Gaspar, C. Evaluating the performance of species richness estimators: sensitivity to sample grain size. *J Anim Ecol* **75,** 274–287 (2006).
84. Vavrek, M. J. fossil: Palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica* **14,** 1T:16p (2011).
85. Dunn, P. K. tweedie: tweedie exponential family models. *R package version 2.1.7* (2011).
86. Davison, J., Öpik, M. & Daniell, T. J. Arbuscular mycorrhizal fungal communities in plant roots are not random assemblages. *FEMS Microbiology Ecology* **78,** 103–115 (2011).
87. Munch, K., Boomsma, W., Huelsenbeck, J. P., Willerslev, E. & Nielsen, R. Statistical assignment of DNA sequences using Bayesian phylogenetics. *Systematic Biology* **57,** 750–757 (2008).
88. Roberts, D. W. labdsv: ordination and multivariate analysis for ecology. *R package*

version 1.5-0 (2007).

89. Underwood, A. J. *Experiments in ecology: their logical design and interpretation using analysis of variance*. (Cambridge University Press, Cambridge, 1997).

90. Shehzad, W. *et al.* Prey preference of snow leopard (Panthera uncia) in South Gobi, Mongolia. *PLoS ONE* **7,** e32104 (2012).

91. Möller, P., Bolshiyanov, D. Y. & Bergsten, H. Weichselian geology and palaeoenvironmental history of the central Taymyr Peninsula, Siberia, indicating no glaciation during the last global glacial maximum. *Boreas* **28,** 92–114 (1999).

92. Kotov, A. N., Lozhkin, A. V. & Ryabchun, V. K. Permafrost-facial conditions of the Upper Pleistocene deposit creation of the Main River valley (Chukotka). *Forming of Relief, Correlated Deposit and Gravels of the Northern-east of USSR, SVKNII DVO AS USSR, Magadan* 117–131 (1989).

93. Kuzmina, S. A., Sher, A. V., Edwards, M. E. & Haile, J. The late Pleistocene environment of the Eastern West Beringia based on the principal section at the Main River, Chukotka. *Quaternary Science Reviews* **30,** 2091–2106 (2011).

94. Sher, A. V. *et al.* Late Cenozoic of the Kolyma Lowland. *XIV Pacific Science Congress, Khabarovsk* 1–116 (1979).

95. Wetterich, S., Schirrmeister, L. & Kholodov, A. L. The joint Russian-German expedition Beringia/Kolyma 2008 during the International Polar Year (IPY) 2007/2008. *Reports on Polar and Marine Research* **636,** 43 (2011).

96. Zanina, O. G., Gubin, S. V. & Kuzmina, S. A. Late-Pleistocene (MIS 3-2) palaeoenvironments as recorded by sediments, palaeosols, and ground-squirrel nests at Duvanny Yar, Kolyma lowland, northeast Siberia. *Quaternary Science Reviews* **30,** 2107–2123 (2011).

97. Ager, T. A. Late Quaternary vegetation and climate history of the central Bering land bridge from St. Michael Island, western Alaska. *Quaternary Research* **60,** 19–32 (2003).

98. Muhs, D. R., Ager, T. A., Been, J., Bradbury, J. P. & Dean, W. E. A late Quaternary record of eolian silt deposition in a maar lake, St. Michael Island, western Alaska. *Quaternary Research* **60,** 110–122 (2003).

99. Sanborn, P. T., Smith, C. A., Froese, D. G. & Zazula, G. D. Full-glacial paleosols in perennially frozen loess sequences, Klondike goldfields, Yukon Territory, Canada.

Quaternary Research **66,** 147–157 (2006).

100. Zazula, G. D., Froese, D. G., Elias, S. A. & Kuzmina, S. Arctic ground squirrels of the mammoth-steppe: paleoecology of Late Pleistocene middens (~24000–29450 14 C yr BP), Yukon Territory, Canada. *Quaternary Science Reviews* **26,** 979–1003 (2007).

101. Froese, D. G., Zazula, G. D. & Reyes, A. V. Seasonality of the late Pleistocene Dawson tephra and exceptional preservation of a buried riparian surface in central Yukon Territory, Canada. *Quaternary Science Reviews* **25,** 1542–1551 (2006).

102. Demuro, M. *et al.* Optically stimulated luminescence dating of single and multiple grains of quartz from perennially frozen loess in western Yukon Territory, Canada: comparison with radiocarbon chronologies for the late Pleistocene Dawson tephra. *Quaternary Geochronology* **3,** 346–364 (2008).

103. Wetterich, S., Schirrmeister, L. & Kholodov, A. L. The joint Russian-German expedition Beringia/Kolyma 2008 during the International Polar Year (IPY) 2007/2008. *Reports on Polar and Marine Research* **636,** 43 (2011).

104. Beilman, D. W. Holocene and recent carbon accumulation in Svalbard mires. Svalbard Geology Workshop –Tromso Norway 27-29 April (2011).

105. Kosintsev, P. A. et al. Environmental reconstruction inferred from the intestinal contents of the Yamal baby mammoth Lyuba (Mammuthus primigenius Blumenbach, 1799). *Quaternary International* **255**, 231–238 (2012a).

106. Boeskorov, G. G. et al. Woolly rhino discovery in the lower Kolyma River. *Quaternary Science Reviews* **30**, 2262–2272 (2011).

107. Harington, C. R. & Eggleston-Stott, M. Partial carcass of a small Pleistocene horse from Last Chance Creek near Dawson City, Yukon. *Current Research in the Pleistocene* **13**, 105–107 (1996).

108. Sulerzhitsky, L. D. & Romanenko, F. A. Age and dispersal of 'mammoth' fauna in Asian Polar region (according to radiocarbon data). *Kriosfera Zemli (Earth Cryosphere)* **1**, 12–19 (1997).

109. Lazarev, P. A. Skeleton of the woolly rhinoceros from Churapcha. In: Labutin, Y. (ed.): Mammals of the Yakutian Anthropogene: 55–97; Yakutsk (Russian Academy of Sciences, Yakutian Scientific Centre) [in Russian] (1998).

110. Kosintsev, P. A., Lapteva, E. G., Korona, O. M. & Zanina, O. G. Living environments and diet of the Mongochen mammoth, Gydan Peninsula, Russia. Quaternary International 276, 253–268 (2012b).