Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago

Oliver Smith[1], Garry Momber[2], Richard Bates[3], Paul Garwood[4], Simon Fitch[5], Mark Pallen[6¶], Vincent Gaffney[7¶], Robin G. Allaby[1¶*].

1. School of Life Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom.
2. Maritime Archaeology Trust, Room W1/95, National Oceanography Centre, Empress Dock, Southampton SO14 3ZH, United Kingdom.
3. Department of Earth Sciences University of St Andrews, St Andrews, Fife,KY16 9AL, Scotland.
4. Department of Classics, Ancient History and Archaeology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom.
5. School of History and Cultures, University of Birmingham, IBM VISTA ERI building, Pritchatts road, Birmingham, B15 2TT, United Kingdom.
6. Warwick Medical School, University of Warwick, Coventry CV4 7AL, United Kingdom.
7. Division of Archaeological, Geographical and Environmental Sciences, University of Bradford , Bradford . West Yorkshire. BD7 1DP , United Kingdom.

¶ Equal senior authors
* Corresponding author r.g.allaby@warwick.ac.uk

Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago

Abstract

The Mesolithic-to-Neolithic transition which coincided with rising sea levels, marked the time when a hunter-gatherer economy gave way to agriculture. Bouldnor Cliff is a submarine archaeological site with a well-preserved Mesolithic palaeosol dated to 8000 years BP. We analyze a core obtained from sealed sediments, combining evidence from microgeomorphology and microfossils with sedimentary ancient DNA (sedaDNA) analyses to reconstruct floral and faunal changes during occupation of this site, before it was submerged. In agreement with palynological analyses, the sedaDNA sequences suggest a mixed habitat of oak forest and herbaceous plants. However, in later sediments, they also provide evidence of wheat 2000 years earlier than expected. These results suggest that sophisticated social networks linked the Neolithic front in southern Europe to the Mesolithic peoples of northern Europe.

The Mesolithic-to-Neolithic transition is associated with the replacement of a hunter-gatherer economy by arable farming of crops such as einkorn, emmer and barley. Although it is generally accepted that the Neolithic had arrived by 6000 years BP on the British mainland, controversy surrounds the timing and mode of Neolithisation in the British Isles[1]. It also remains unclear whether the arrival of Neolithic technologies on the mainland was rapid, facilitated by the arrival of migrating farmers, who displaced or acculturated existing hunter-gatherers[2], or whether hunter-gatherers gradually transitioned to a Neolithic economy with increasing dependency on cereals over millennia[3].

The Neolithic arrived on the British mainland during a warming period in which sea levels rose and inundated land between Britain and Europe, forming the English Channel and much of the North Sea[3-7]. We hypothesized that the earliest stages of Neolithisation in the British Isles occurred in these lowland regions.

Some Mesolithic palaeosols, representing the old land surface, have been preserved under marine sediments, including a palaeosol from Bouldnor Cliff, off the Isle of Wight in the Western Solent. The site has been dated to 8030-7980 cal BP [8] (Fig. 1, Table S1), placing it in the late Mesolithic of the British Isles, a period which is represented by few assemblages and still little understood. This palaeosol formed along the edge of an ancient valley that was dissected by a river and fed by tributaries from the surrounding hills. The area then became wetter, forming a peat bog, before eventual marine inundation over a period of about 30-100 years, followed by deposition of marine sediments[9]. Archaeological artefacts from this palaeosol include worked and burnt flint, corded fibre, worked wood and burnt hazelnut shells[10]. Many of these artefacts represent early instances of such technologies and suggest that the Mesolithic peoples of Bouldnor Cliff were connected to more advanced groups from

Europe relative to those on mainland Britain. We thus analyzed sedimentary ancient DNA (sedaDNA) from sediment cores from the Bouldnor Cliff site. The sedaDNA approach has been applied to a range of terrestrial and marine sediments and has been shown to be highly informative in environmental contexts, providing more information than macrofossil assemblages[12-14]

Before taking samples for DNA analysis, microgeomorphological and microfossil analyses were undertaken on four sediment cores (MS-04, MS-05, MS-07 and MS-08[9], Fig. S2). The sediment layers were of low porosity, with the palaeosol and peat layers sealed beneath dense, silty-clay marine alluvial sediments. We found a sharp boundary between the palaeosol and overlying peat, with no evidence of mixing of particles. Diatom and foraminifera analysis revealed a range of species in the superficial marine alluvial sediments. However, these did not penetrate into the underlying peat layer, indicating a lack of vertical movement in the sediment column. Given the absence of evidence of soil erosion (as might be revealed by illuviation or podsolization), we concluded that archaeological artefacts had been deposited *in situ* on a pristine land surface rather than entered the samples through alluvial deposition from another site. Radiocarbon dates obtained from 21 samples of wood and plant macrofossils from the sediment cores[9] (Table S1, Fig. S1, Fig. S2) allow an inference of marine inundation beginning 8020-7980 years cal BP, which represents the latest date for human activity at the site, with inundation complete by 7990-7900 years cal BP.

We took four palaeosol sediment samples (S308 0-2cm, S308 2-4cm, S308 4-6cm and S308 6-8cm) from a location at the site associated with Mesolithic food debris (burnt hazelnut shells)[9]. The samples were taken at successive 2cm intervals from the top of the stratum, each roughly representing the period of a decade. Samples

were taken on site[15], examined for macrofossils and subjected to ancient DNA extraction in a dedicated laboratory[16]. Samples were found to be devoid of macrofossils, apart from a few *Alnus glutinosa* (common alder) twigs.

We made Illumina libraries from the sediment cores and generated 71,856,199 256-bp single-end reads on the MiSeq platform (Table S2). To overcome the problem of bias resulting from partial representation of species in the publicly available sequence databases, we designed a phylogenetic intersection analysis for phylogenetic assignation of sedaDNA[15]. that we estimate has an accuracy of 81%[15]. Sedimentary DNA sequences were sorted into major taxonomic groups in a preliminary phylogenetic clustering stage[17] and phylogenetic intersection analysis on major tetrapod groups (except for primates) and flowering plants (Magnoliophytes) (Table S3), were performed under high-stringency conditions (≥ 99% of read length aligning to database entries, with ≥ 3 taxa on which to base the phylogenetic intersection analysis). Reads that met the stringency criteria were used to construct a palaeoenvironmental profile (Fig. 2, Tables S4, S5). It was not necessary to reject any data, *post hoc*[15], allowing us to avoid imposing any *a priori* assumptions about the past environment.

The sedaDNA profile revealed a wooded landscape that included oak, poplar, apple and beech family members, with grasses and few herbs present. Oak and poplar were also detected in the pollen profile[9,18] , while oak, apple and alder have been reported in archaeological worked wood remains at the site[9,10].

Palynological analysis shows an abundance of true grass species (Poaceae) at the site, which is reflected in more detail in the sedaDNA profile (Figures S3-S6). There is a marked difference in the profile of grasses and fauna between the top half (0-4cm) and bottom half (4-8cm) of the palaeosol. The lower strata contain a varied

and abundant representation of major clades of grasses. Surprisingly, we also found sequences assigned to Triticeae, which is the grass tribe within the Pooideae that encompasses genera with many domesticated species of cereal crop in the lower strata, albeit at relatively low levels (4% of the sedaDNA signal for flowering plants). However, reads assigned to Triticeae grow to dominate the plant profile in the upper strata (81% of the signal for flowering plants). The Poaceae type pollen from this zone 1b does not indicate the presence of larger cereal type pollen[9], suggesting that the source of the *Triticum* signal is unlikely to be from wheat that was grown on site.

We specifically examined the possibility that the *Triticum* signal could be due to a false positive or may have been caused by the members of the Triticeae that are known to be native to the British Isles (*Leymus, Elymus*, *Agropyron* and *Hordelymus*). All instances in which these species were detected in the analysis showed them to be less similar to the sedaDNA than the *Triticum*/*Aegilops* group. Many sedaDNADNA sequences showed 100% nucleotide identity with sequences from *Triticum*, particularly *Triticum monococcum* (einkorn), with decreasing similarity next to its sister genus *Aegilops* (Table S5). The British Triticeae all occur outside this taxonomic group and were excluded as a possible source of the signal. Since both the *Triticum* and *Aegilops* are Near-Eastern genera with no known wild members in northern Europe, we conclude that these are genuine wheat sequences. We considered the possibility that the sequences could be due to modern contamination, but discounted this because wheat has not been studied in the ancient DNA facility used. Furthermore, the sediment samples were taken during the winter months and sealed underwater and the stratified sedaDNA signal would not be expected from contamination. We also tested for the presence of wheat DNA in both our reagents and the core samples by preparing blank libraries from the same reagent batch, which

we sequenced on the MiSeq platform. This resulted in no evidence of wheat in this data set, or indeed any higher plant. We thus conclude that the *Triticum* sequences derive from the core itself, and therefore are associated with Mesolithic human activity at the site.

SedaDNA analysis of the upper strata further revealed a faunal profile compatible with human activity, with an abundant presence of *Canis* and Bovidae. *Canis* may be interpreted as either dogs or wolves. Two of the bovid reads sat at the intersection of *Bos* with the sister genus *Bison*, which we interpreted as most likely *Bos*, and which was supported by a subsequent find of an auroch bone at the site (Fig. S8). We also detected the presence of deer, members of the grouse family, and rodents compatible with the contents of a Mesolithic diet shared by humans and dogs.

The occurrence of wheat 8,000 years ago on the British continental shelf appears early given its later establishment on the UK mainland. Neolithic assemblages first appear in northwest Europe in the 8th millennium BP, from 7500 BP in the central Rhineland[19], 7300 BP in the Rhine/Maas delta and adjacent areas[20,21] and 7400 BP in western France[22]. These developments were driven by the spread of the LBK culture from trans-Danubia into central Europe 7600 BP[19], and the Cardial culture from the Mediterranean into Western France around 7600-7400 BP[23]. These dates suggest only a 400 year gap between Britain and the earliest known presence of farming geographically in nearby. However, the spread of the Cardial culture is still incompletely understood, with dates of 8000 BP in southern France a recurring theme[24,25]. Given the littoral route of the Cardial spread it is possible that earlier sites may also be submerged in Southern Europe. Such dates are contemporaneous with the Mesolithic site at Bouldnor cliff and given the high mobility of Mesolithic communities it is plausible that communication occurred over such distances. It has

been suggested that agricultural products moved ahead of the front of Neolithisation into Mesolithic zones[26].

In the absence of significant environmental barriers to the dissemination of agriculture across Europe, there is no *a priori* reason why the spread of farming products from the Balkans to the Atlantic zone was not swift: as the only constraints were the scale, intensity and spatial and temporal articulation of social and demographic networks. It is possible that the isolation of Britain from mainland Europe by sea represented an environmental barrier. Although sea levels clearly rose during the early Holocene, the identification of coastlines within the North Sea during the early Holocene are complex[27] and estimates of coastline have been speculative[28]. We explored the plausibility of a direct connection between Britain and Europe at the time of the palaeosol by plotting a generalized map on the basis on C^{14} and OSL dated marine cores of early Holocene sediments taken from the east coast of the UK[29] (Fig 3). This map represents an estimate of coastal extent around southern Britain from dated evidence. These data support two possible points of direct contact with northern France and the Netherlands respectively, supporting the possibility of contact between the Mesolithic peoples of the British area with both LBK and Cardial cultures.

Our sedaDNA analysis has revealed the presence of wheat, a domesticated plant associated with the Neolithic, at a site on the British continental shelf 2000 years earlier than would be expected from the known archaeology of the British mainland. We obtained no archaeological evidence suggesting cultivation at this site. The Poaceae pollen profile does not show an expansion indicating an open environment suitable for farming until higher strata above the peat zone overlaying the palaeosol[15] (pollen zone 2). Therefore, in the absence of direct evidence, we suspect that this wheat represents foodstuffs imported from the continent rather than cultivation of this

cereal crop at this locale. The presence of wheat, along with pioneering technological artefacts at the site, provides evidence for a social network between well-developed Mesolithic peoples of north-western Europe and the advancing Neolithic front. In this light, recent debates concerning chronologies of transition, geographical origins, respective contributions of migrants and natives, and colonisation and acculturation processes of the earliest farmers of north-west Europe during the late 8th and 7th millennia BP[2,30,31], may in fact have focused on the second half of the time-frame of early farmer/hunter-forager interaction and cultural change in this region.

References

1. C.J. Stevens, D.Q. Fuller, *Antiquity* **86**, 707-722 (2012).
2. P. Rowley-Conwy, *Curr. Anthrop.* **52**(S4), S431–S451 (2011)
3. J. Thomas, *Understanding the Neolithic. A revised second edition of rethinking the Neolithic.* (Routledge, London, 1999).
4. V. Gaffney, S. Fitch, D.Smith, *Europe's Lost World, the rediscovery of Doggerland*. (CBA Research Report 160, York 2009).
5. . G. Momber (2014) In *Prehistoric Archaeology of the Continental Shelf: a Global Review*. Eds. A. Evans, J. Flatman, N.C. Flemming (Springer, New York, 2014) pp. 194-212
6. J.H.M Peeters, G. Momber, *Netherlands J. Geosci.*. **93,** 55-70 (2014).
7. V. Gaffney, K. Thomson, S. Fitch, *Mapping Doggerland*. (Archaeopress, Oxford, 2007).
8. D. Tomalin, G. Momber, N. Nayling. In *Mesolithic occupation at Bouldnor Cliff and the submerged prehistoric landscapes of the Solent.* Eds. G. Momber, D. Tomalin, R. Scaife, J. Satchell, J. Gillespie, Council for British Archaeology Research Report 164. (CBA, York, 2011) pp, 3-14.
9. G. Momber, *et al.*. In *Mesolithic occupation at Bouldnor Cliff and the submerged prehistoric landscapes of the Solent.* Eds. G. Momber, D. Tomalin, R. Scaife, J. Satchell, J. Gillespie, Council for British Archaeology Research Report 164. (CBA, York, 2011) pp 19-65.
10. G. Momber, *et al.*,. In *Mesolithic occupation at Bouldnor Cliff and the submerged prehistoric landscapes of the Solent.* Eds. G. Momber, D. Tomalin, R. Scaife, J. Satchell, J. Gillespie, Council for British Archaeology Research Report 164. (CBA, York, 2011) pp 66-93.

11. D. Tomalin. In *Mesolithic occupation at Bouldnor Cliff and the submerged prehistoric landscapes of the Solent.* Eds. G. Momber, D. Tomalin, R. Scaife, J. Satchell, J. Gillespie, Council for British Archaeology Research Report 164. (CBA, York, 2011) pp 155-163.

12. E. Willerslev *et al.*. *Science* **317**, 111-114 (2007).

13. F. Lejzerowicz *et al*., *Biology Letters* **9**, 20130283.

14. J. Haile *et al.*, *Mol. Biol. Evol.* **24**, 982-989.

15. See Supplementary Information online for methods.

16. M.T. Gilbert, H.J. Bandelt, M. Hofreiter, I. Barnes *Trends Ecol. Evol.* **20**, 541–544 (2005).

17. D.H. Huson, A.F. Auch, J. Qi, S.C. Schuster, *Genome Res.,* **17**, 377-386 (2007).

18. G. Momber *et al.*, In *Mesolithic occupation at Bouldnor Cliff and the submerged prehistoric landscapes of the Solent.* Eds. G. Momber, D. Tomalin, R. Scaife, J. Satchell, J. Gillespie, Council for British Archaeology Research Report 164. (CBA, York, 2011) pp 105-119.

19. J. Robb, *Curr. Anthrop.* **54** (6), 657-83 (2013).

20. L.P. Louwe Kooijmans, *Proc. Brit. Acad.***144**, 287-309 (2007)

21. P. Crombé, B. Vanmontfort, *Proc. Brit. Acad.***144**, 263-85 (2007).

22. G. Marchand, *Proc. Brit. Acad.* **144**, 225-42 (2007)

23. A. Tresset, J.-D. Vigne, *Proc. Brit. Acad.* **144**, 189-210 (2007).

24. J. Guilane, C. Manen, **144,** 21-51. *Proc. Brit. Acad* (2007)

25. C. Joly, L. Visset, *Rev. Palaeobot. Palynol.* **154**, 124-79 (2009)

26. C. Jeunesse, *Revue d'Alsace* **129**, 97-112 (2003).

27. T. Bell, A. O' Sullivan, R. Quinn, *Archaeol. Ireland* **20**, 2, 12-17 (2006).

28. B.J. Coles, *Proc. Prehist. Soc.* **64**, 45–81 (1998).

29. D.R. Tappin, *et al., The Humber Regional Environmental Characterisation.* British Geological Survey Open Report OR/10/54. (Marine Aggregate Levy Sustainability Fund, Lowestoft 2011) pp 194-230.

30. M. Collard, K. Edinborough, S. Shennan, M.G. Thomas, *J. Archaeol. Sci.* **37**, 866–70 (2010).

31. J. Thomas, In J. Pollard, J, (ed), *Prehistoric Britain*, (Blackwell, Oxford, 2008) pp58-89.

32. G. Momber, In *Mesolithic occupation at Bouldnor Cliff and the submerged prehistoric landscapes of the Solent.* Eds G. Momber, D. Tomalin, R. Scaife, J. Satchell, J. Gillespie, Council for British Archaeology Research Report 164. (CBA, York, 2011) pp 123-135.

33. G. Momber, In J. Benjamin et al. (eds.) *Submerged Prehistory*. (Oxbow Books. Oxford, 2011). pp 85-98.

34. X. Ke, M. Collins, In T. Healy, Y. Wang and J. A. Healy (eds.) *Muddy Coasts of the World: Processes, Deposits and Function*. (Elsevier Science, Amsterdam, 2002). pp 411-440.

35. S.A. Palmer, *et al.*, *Mol. Biol. Evol.* **29**, 2031-2038 (2012).

36. A.W. Briggs, *et al.*, *Nucl. Acids Res.* **38**, e87 (2010).

37. C. Amante, B.W. Eakins, ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA, doi:10.7289/V5C8276M (2009).

38. T. Bell, A. O' Sullivan, R. Quinn. *Archaeol. Ireland* **20**, 2, 12-17 (2006).

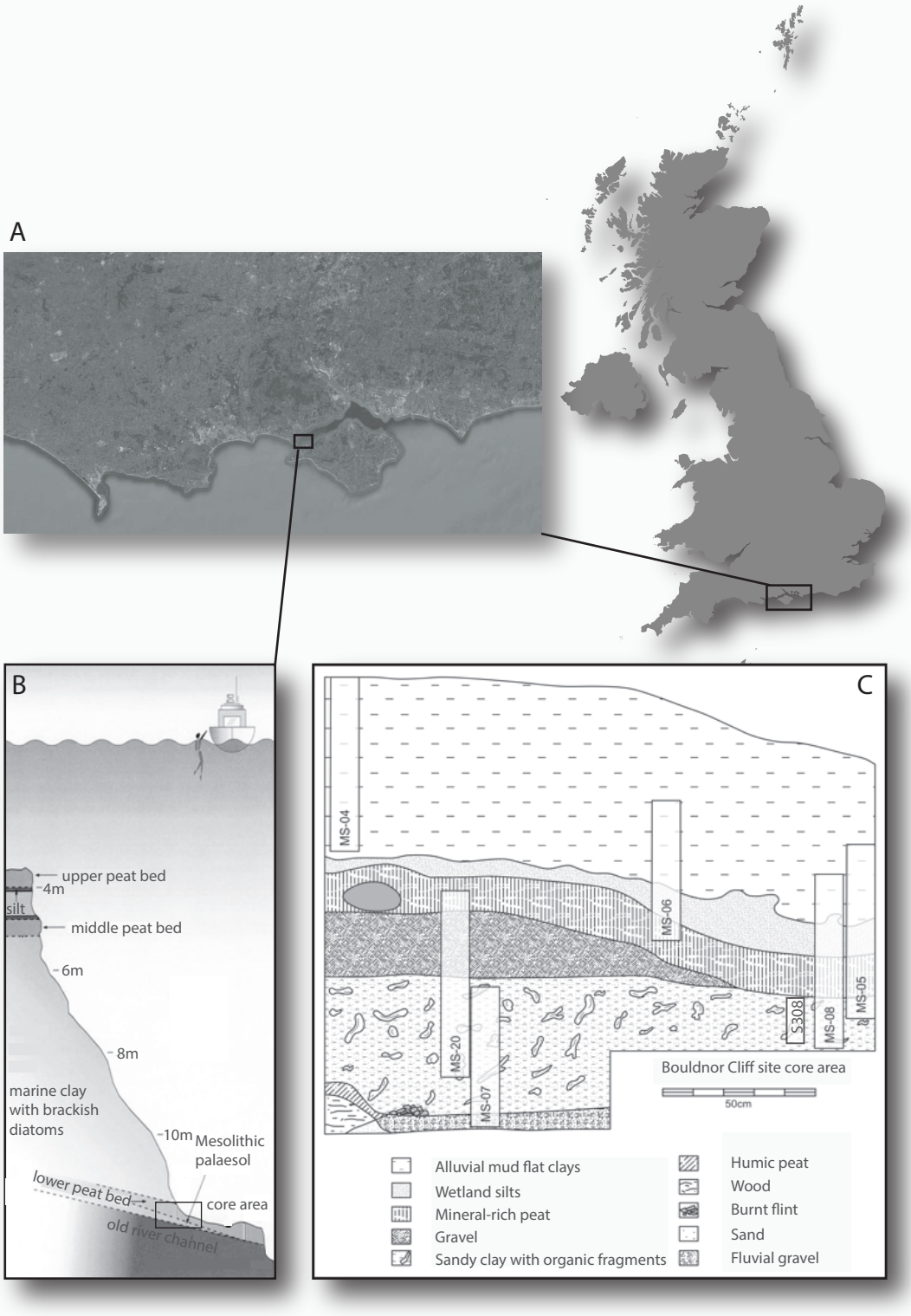
Acknowledgements

The authors wish to thank Alan Clapham for identification of *Alnus glutinosa* fragments and Jacqueline Z.M. Chan and Gemma L. Kay for assistance with operating the MiSeq platform, and the SplashCOS and Deucalion networks for continued support. O.S. was funded by Warwick Medical School. Sequence data was deposited at EMBL-EBI, accession number PRJEB6766 (ERP006391).

Figure 1. Sampling location of sedaDNA from Bouldnor Cliff. A, Location of the Bouldnor Cliff site in the Solent on the coast of the Isle of Wight. B, Large-scale stratigraphic profile of the site indicating the depth and location of the Mesolithic palaeosol, and location of the area from which cores were taken. C, Core area in detail, stratigraphic profile of the site indicating core sites (MS-04-8, and MS-20), and approximate location of sediment sample taken for sedaDNA analysis (S308).

Figure 2. Floral and faunal composition of the Mesolithic palaeosol. Depths are measured from the top of the stratum. Compositions are based on read assignations detailed in Table S3.

Figure 3. Generalized map of potential coastal extent around southern Britain 9840-7830 cal years BP. Vibrocores through submersed old land surfaces off the coast of Britain at depths of 31.68m (VC39), 24.01m (VC51) and 23.90m (VC29) were OSL dated[33]. Map extrapolates the contour of the VC29 vibrocore site.


A
B
upper peat bed
4m
silt
middle peat bed
6m
8m
marine clay
with brackish
diatoms
10m
Mesolithic
palaesol
lower peat bed
core area
old river channel
C
MS-04
MS-06
MS-20
MS-07
S308
MS-08
MS-05
Bouldnor Cliff site core area
50cm
Alluvial mud flat clays
Wetland silts
Mineral-rich peat
Gravel
Sandy clay with organic fragments
Humic peat
Wood
Burnt flint
Sand
Fluvial gravel

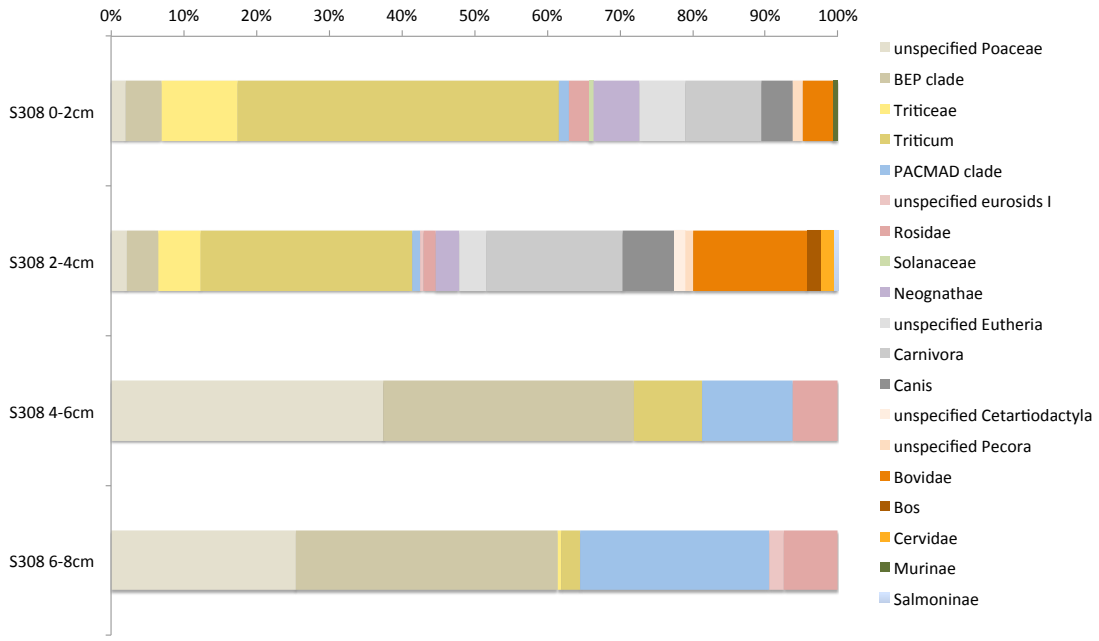
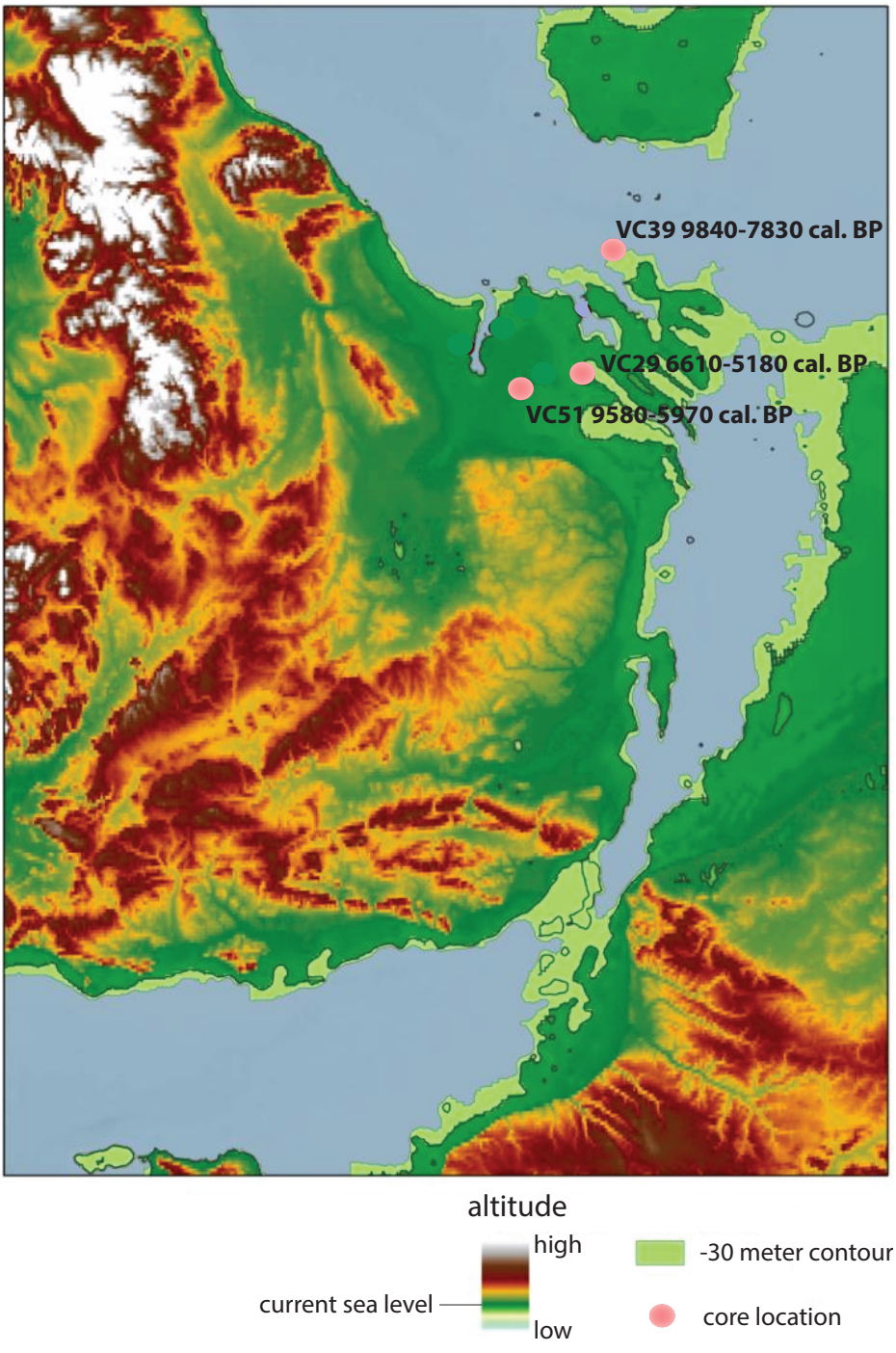

0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
S308 0-2cm
S308 2-4cm
S308 4-6cm
S308 6-8cm
unspecified Poaceae
BEP clade
Triticeae
Triticum
PACMAD clade
unspecified eurosids I
Rosidae
Solanaceae
Neognathae
unspecified Eutheria
Carnivora
Canis
unspecified Cetartiodactyla
unspecified Pecora
Bovidae
Bos
Cervidae
Murinae
Salmoninae

Figure 2


VC39 9840-7830 cal. BP
VC29 6610-5180 cal. BP
VC51 9580-5970 cal. BP
altitude
high
current sea level
low
-30 meter contour
core location

!

Supplementary Information

Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago

Oliver Smith[1], Garry Momber[2], Richard Bates[3], Paul Garwood[4], Simon Fitch[5], Mark Pallen[6¶], Vincent Gaffney[7¶], Robin G. Allaby[1¶*].

1. School of Life Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom.
2. Maritime Archaeology Trust, Room W1/95, National Oceanography Centre, Empress Dock, Southampton SO14 3ZH, United Kingdom.
3. Department of Earth Sciences University of St Andrews, St Andrews, Fife,KY16 9AL, Scotland.
4. Department of Classics, Ancient History and Archaeology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom.
5. School of History and Cultures, University of Birmingham, IBM VISTA ERI building, Pritchatts road, Birmingham, B15 2TT, United Kingdom.
6. Warwick Medical School, University of Warwick, Coventry CV4 7AL, United Kingdom.
7. Division of Archaeological, Geographical and Environmental Sciences, University of Bradford , Bradford . West Yorkshire. BD7 1DP , United Kingdom.

¶ Equal senior authors
* Corresponding author

Contents

Methods

1.Sediment samples.

1.1 Formation of Bouldner Cliff

Bouldner Cliff is a Mesolithic site that preserves a diverse range of materials from worked wood to stone tools which rests in a submerged landscape in the Solent, UK, 11.2m below Ordnance Datum and dates to 6030–5980 cal BC [OxA-15718][9]., Table S3, Figure S1. The terrestrial material was inundated 8,000 years ago during the Flandrian transgression when it was protected and preserved beneath peat and a thick deposit of brackish water sediment[32]. As the sea level rose, estuarine mud flat built up over the peat. It reached a thickness of over 7m during a period of c.1,400 years, after which, the pace of sea level rise slowed. The top of the sediment cover is dated to 4525–4330cal BC [Beta-140102].

The mud flats were formed in a sheltered inland estuary within a valley set back 4km from the sea. It was drained by the river Yar to the south while the Lymington River flowed into the system from the north. This resulted in a natural sink that continued to attract sediment until the east and west flanks of the valley were overtopped to form a tidal strait some time after 1900–1690 cal BC (Beta-270797)[33,34]. The new strait ran perpendicular to the established fluvial system and cut through the estuarine sediments that had built up over the previous 4,000 years. The tidal currents lowered the valley floor but left steep sided sections to the north and south. The archaeological landscape is currently exposed at the foot of the sediment sequence remaining at Bouldnor Cliff on the south side of the channel. The sample was collected from an exposure of relict terrestrial material that is being eroded by approximately 30cm a year.

1.2 Retrieval of sediment samples

Six monolith cores (MS-04, MS-05, MS-06, MS-07, MS-08 and MS-20) were taken from Bouldnor Cliff, as illustrated in Figure 1, from which micromorphology, palynology, diatom and foraminifera, macrofossil studies were carried out (see section 2 below) to establish the stability and nature of the stratigraphy at the site. The direct sediment sample for DNA analysis (S3-08) was taken in the proximity MS-08 and MS-05, which were adjacent to each other, Figure 1. In order to take samples that would minimize disturbance of the sediment matrix, purpose built tins of marine grade stainless steel were designed measuring 500mm long, 250mm wide and 150mm deep. These tins were knocked into the exposed sediment surface and then released from the cliff with the aid of handsaws. Once retracted from the cliff and still underwater, the tins were sealed with industrial cling film and transported to the surface. The tins were then transported to a clean laboratory where they were opened and internal sections were retrieved with a scalpel.

2.Taphonomic studies and leaching issues

A number of studies were carried out to establish the nature of the stratigraphy[9]. The results are summarized in Figure S2.

2.1 Micromorphology analysis

The aim of this analysis was to establish whether material moved between strata through leaching, and whether the palaeosol layer represented a pristine soil surface or an eroded depositional environment[9]. By understanding this an interpretation could be made about whether artifacts were deposited *in situ* or are likely to represent reworked material. Monoliths (MSO8 and MS05) were sampled

with Kubiena tins and thin sections prepared. The micromorphology was examined at x5.8 and x400 using transmitting polarizing light microscopes. Three 100mm sections were examined spanning boundaries between 1 overlaying organic mud and peat (units 2.1 and 3.1 on Figure S2), 2 Peat deposit and palaeosol (units 3.2 and 7.1 on Figure S2), 3 Palaeosol and underlying sediments (unit 7.1 and 7.2 in Figure S2).

Boundary 1 The peat deposit was mineral rich, the mineral component made up of 95% quartz, with accessory grains of glauconite, tourmaline and zircon. Bone fragments occurred were embedded but showed no particular distribution. Secondary minerals occurred of pyrite, gypsum and jarosite. Porosity was low (<10%). Pedofeatures of infilling were rare, indicating some periods of drying and infilling of the peat.

Boundary 2 The boundary units between the peat and underlying palaeosol are graded suggesting no hiatus or erosion of the underlying soil prior to peat formation. The palaeosol is a structureless organic mud, with an unsorted mineral component comprised of 98% quartz. Evidence of pyrite formation was found, and porosity was found to be very low (<5%).

Boundary 3 The layer underlying the palaeosol is made up of a structureless minerogenic sandy clay. Sandy grains are poorly sorted, subrounded to angular, and dominated by quartz (95%), with accessory grains of glauconite, tourmaline and zircon. Organic content was lower than the overlaying palaeosol, and randomly distributed. Porosity was low (<10%). No evidence of diagnostic pedogenic processes (illuviation or podsolization) was found, but the morphology of the void space suggests some biopertubation and soil formation through weathering.

The micromorphology analysis shows that there was little or no erosion of the land surface before peat formation occurred, therefore the material of the palaeosol

represents an *in situ* context, rather than exogenous material that may have been washed in. Once laid down there has been little evidence of disturbance, generally no sorting of granular material, combined with very low porosities. Together, these data suggest that there has been little or no vertical movement of material, so no evidence for a leaching process was found. A subsequent thick overlaying silt was laid down by the marine inundation of the site (unit 1 on Figure S2), sealing the layers beneath.

2.2 Palynology

Pollen analysis of monoliths (MS05, MS08, MS04 and MS07) was carried out[9] as well as a core from across the valley at Tanners Hard[18] (Figs. S3-S7). This second off-site coring location was selected at Tanners Hard because it reflects the northern extent of the Bouldnor Cliff palaeo-valley that has also been associated with archaeological artefacts[18], and so provides a more complete picture. Pollen zones were subdivided into two principal zones and four subzones for MS05, MS04 and MS08, and there were three principal zones in MS07 (Figs. S4-S7). The first zone spans the lower humic sands (S08:1a, S08:1b and S08:1c on Figure S2), and includes the Mesolithic Palaesol (zone 1b). These zones are dominated by *Pinus* (mean values of 21%, 18%, 4%, 3% for zones 1a, 1b, 1c and 1d respectively), *Quercus* (7%, 13%, 14% and 18%), *Alnus glutinosa* (32%, 32%, 46% and 40%) and *Corylus avellana* (32%, 28%, 24% and 22%). There is a presence of *Ulmus* (4%, 3%, 1%, 3%), and sporadic appearances of *Betula, Sorbus, Prunus/Malus, Fraxinus* and *Tilia*. Herbs are present, but in low pollen frequency and dominated by Poaceae (1%, 2%, 3% and 8%). Although a very low frequency of large Poaceae pollen occurs, no cereal like grains were identified, and no large grains occur in the upper section of zone 1b in which the Mesolithic palaeosol occurs. Marsh fen types occur including *Typha angustifolia, Filipendula ulmaria* and Cyperaceae. Herb families present include

Rosaceae, Brassicaceae, Chenopodiaceae, Fabaceae, Apiaceae and Primulaceae. Across the valley at Tanners Hard a number of additional pollen types were identified from this zone, including *Populus* and Solanaceae. The base of zone 1 (1a) represents a woodland in which pine dominates, characteristic of the end of the Boreal pine maxima and progressive expansion of *Quercus* and *Corylus*. Zone 1b contains the palaeosol and is typical of the late Boreal (Flandrian Ic) prior to the separation of Britain from Europe. This zone is typified by an expansion of *Quercus* and the appearance *Ulmus* and *Tilia*, while *Pinus* declines. During zone 1c, *Pinus* continues to decline as does *Ulmus*, and *Quercus* continues to expand, and the Poaceae increase in frequency. This zone is one of peat accretion indicating the development of a damp fen woodland. Zone 1d occurs towards the top of the peat layer and is characterized by an expansion of Poaceae, and marsh taxa such as *Typha*. Herbs generally become more important and diversify. The vegetation reflects a rising water table, fresh water at first, but becoming brackish later. Zone 2 is characterized by a change from peat formation to the overlaying alluvial sediments with increasing water levels. While the tree taxa still remain important, there are expansions in the marsh-associated herbs and a large expansion of Poaceae pollen. This phase represents the Boreal Atlantic expansion of sea level. The three pollen zones identified in MS07 broadly equate to the pollen zones 1a, 1b and 1c identified in the other three cores in species composition.

2.3 Diatom and foraminifera analysis

Sediments from across the monolith cores were examined for the presence of diatoms and foraminifera[9]. Several species of both diatoms and foraminifera were found in the upper silty marine layers (Unit 1 on Figure S2), but not in underlying sediments. This again supports the notion that the sediment is sealed under the marine

silts. The species composition diatoms suggests a brackish water in the lower levels represented by species such as *Nitzschia navicularis* which gives way to species associated with more open deep marine waters such as *Paralia sulcata*. The foraminifera profile reflected the findings with diatoms, with mudflat associated species such as *Haynesina germanica* towards the base of the clays giving way to deeper sea species in the upper strata such as *Gavellinopsis inflate*. Together the diatoms and foraminifera support the marine inundation scenario, and demonstrate that no mixing of the sediments has occurred since deposition.

3.DNA extraction, amplification and sequencing.

DNA extractions were carried out in a dedicated ancient DNA laboratory at Warwick following standard protocols for work with ancient DNA[16]. The ancient DNA facility at Warwick is a positive pressured, restricted access ante-chambered laboratory situated in a different building to PCR and library amplification facilities.

3.1 DNA extraction

For each sediment block, 4.3g - 6.3g was subsampled and transferred to a sterile 15ml falcon tube, then immersed in 5ml CTAB buffer (2% w/v CTAB, 1% w/v PVP, 0.1M Tris pH 8.0, 20mM EDTA, 1.4M NaCl). Samples were incubated at 37°C for 7 days. Buffer was separated from sediment material by centrifugation. DNA was extracted once from the supernatant using 24:1 chloroform : isolamyl alcohol, and subject to column-based cleanup using an Isolate kit (Bioline). An additional cleanup / drying phase using 300µl acetone was performed to ensure complete buffer removal before elution in 100µl elution solution. DNA was quantified using the Qubit (Invitrogen) platform. Aliquots of ~50ng were taken from each DNA extraction for

fragmentation analysis by being run on a 2% Ultrapure (Invitrogen) agarose gel, pre-stained with GelRed.

3.2 Illumina library construction and sequence generation

Libraries were constructed using a TruSeq Nano kit (Illumina), with extended (3x protocol) incubation for end repair and adapter ligation to maximize efficiency of the library construction. The DNA yield from samples S308-1 and S308-2 was significantly less than the others and so 5ng DNA was used as library input for these samples, compared to 50ng for the rest. Bead-based size selection using Ampure XP beads (Agencourt) was used on samples S308-4, 38-42-1 and 148-152-1 due to apparent traces of higher molecular weight DNA outside of the otherwise normally-distributed degraded DNA. Size selection where used was optimized to capture fragments of 0-500 bp. Bead-based cleanup prior to PCR amplification was used for all samples. PCR conditions were executed according to Illumina's protocol, for the maximum 10 rounds. Aliquots of final libraries were visualized on a 2% Ultrapure agarose gel and further aliquots were quantified using the Qubit platform. All samples exhibited normally distributed fragmentation patterns, from which molarity was calculated based on Qubit readings using the Promega Biomath calculator. Libraries were subsequently diluted to 4nM concentration. Completed libraries were loaded on to a MiSeq sequencer using the v2 reagent kit (Illumina) and sequenced as FastQ only, with automated adapter removal.

A control library was also prepared following the same protocol as described above in which the same batch of illumina reagents were used as for the sedaDNA analysis, but with the absence of the DNA extract.

4.Bioinformatics

4.1 Initial checks

Adapter sequences were removed automatically by the MiSeq software. FastQ sequences were initially analyzed using FastQC, showing expected levels of nucleotide presence and fragment length distribution , Figure S10. To compensate for the known propensity of elevated C > U nucleotide substitutions at 5' overhangs, the 10 most 5' bases were removed from each read to reduce BLASTn false negatives resulting from extensive overhang damage, and allowing a more stringent level of criteria to be applied to the PIA analysis.

4.2 Preliminary phylogenetic assortment

All sequences were subjected to a preliminary metagenomic BLASTn analysis, using the full NCBI nt database and a multithreading 64-bit standalone build on a Redhat Linux server. Sequence outputs were limited to 500 hits and one visualized alignment. BLASTn output files were parsed for metagenomic analysis using MEGAN (Metagenome Analyzer)[17]. The majority of reads were assigned by MEGAN to microbial (Table S2). These were, however, discarded, and several taxonomic groups were taken forward for analysis (Table S3). The choice of groups to take forward was restricted to groups of mammals, birds and plants were placed into one group (Magnoliophytes). As was expected from variable database representation of reads in the NCBI database, a number of likely false positives were produced in the preliminary analysis, such as the duck-bill platypus and Tasmanian devil (indicated in bold in Table S3). As a control some of these groups were included in the PIA as a check to see whether they were automatically removed during the analysis.

4.3 Phylogenetic Intersection Analysis (PIA)

Reliable phylogenetic assignation of reads of metagenomic DNA is a non-trivial problem because of variable database representation across organism kingdoms for different genomic regions. If the genomic region of a DNA sequence from the sediment core is not represented by its organism species in the database, then this can result in the DNA sequence being attributed to the 'next closest' species available. Therefore there can be a tendency to assign sequences to taxa that are over represented in the database, which may make no sense ecologically and will often include model organisms. Such sequences may be identified as contamination and require *post hoc* removal that demands an interpretation of the data based on factors other than the information value of the genetic data alone. This is especially problematic in instances where the true species are both unexpected in the archaeological context and model organisms. Many protocols rely on good database representation, for instance through using barcoding targets for which a large number of species have been surveyed, and assignation is dependent on a high threshold of similarity between sedaDNA and database entries.

We elected to use a shotgun sequencing approach rather than a barcoding approach to generate data for phylogenetic analysis. Our reasoning was that bar coding targets represent in the order of 0.001% of the genetic material laid down by organisms, an observation borne out by our previous metagenomic studies[35], and this study (Table S2). Under the conditions of extensive DNA degradation that we expected from the marine environment such targets could be lost by chance relative to the genome as a whole. By using an unbiased approach that obtained DNA from any part of the genome of an organism we aimed to produce a more sensitive assay of the sediments because of a higher probability of representation of all organisms present. While we expected the shotgun approach to elevate our ability to capture ancient

DNA of organisms from the Mesolithic palaeosol, we also expected highly variable database representation of the DNA retrieved. We designed a phylogenetic intersection analysis (PIA) for phylogenetic assignation of sedaDNA to be robust to variable database representation and which did not require a *post hoc* removal of sequences that do not make ecological sense.

The PIA was carried out using a group of scripts written for the project (available from http://www2.warwick.ac.uk/fac/sci/lifesci/research/archaeobotany/downloads/pia), which utilize the names and node files of the NCBI Tree of Life (names.dmp and nodes.dmp). The analysis takes as input a header list of reads from MEGAN that denotes a particular phylogenetic classification and the BLAST output file from BLAST (input file for MEGAN). The program PIA.pl reads through the MEGAN output file and retrieves each read header and then extracts the taxa from corresponding entry from a large BLAST file. The taxa are converted to NCBI identifier codes using the names.dmp file. The taxonomic structure of the taxa is then retrieved using the nodes.dmp file for the first two BLAST hits. The first identical taxonomic code identified between the two taxa is taken as the taxonomic rank of intersection. A taxa diversity (*td*) score was calculated as

$$td = \frac{t-1}{c} \qquad (1)$$

where t is the total number of different taxa that were found in a blast entry to be similar to a query sequence, and c is a predefined cap of taxa. For convenience we set c to 10 for our analyses. The output file then recorded the top hit, level of similarity and read length, the identity of the next and last hits, the total number of hits, the taxon diversity, the taxonomic rank of intersection and type of rank of the intersection.

The next step of the analysis was to prune data that did not meet stringent criteria of length of query covered by first hit and taxonomic diversity. We set our criteria for length coverage at 99%, and taxonomic diversity at 0.2 (ie at least three different species in blast hits). For downstream analysis, reads were split into three bins; those requiring no further analysis, those requiring a higher cap cutoff because they had more than ten taxa retrieved from the blast search, and those requiring further BLASTn analysis (where all 500 initial results are limited to a single organism). In the latter case up to 2000 BLAST hits were considered and examined to see if the requisite taxon diversity was achieved. In the final pruning stage data was removed in which the similarity did not reduce between the first and second hits and the query. In cases where a number of taxa occurred with the same level of similarity before a drop off was observed, the taxonomic intersection was raised to the common point for the group.

4.4. Analysis of robustness of PIA

The aim of the PIA was to produce an approach that is robust to variable database representation of the true species identity of sedaDNA reads, and avoid a specific taxonomic assignation to the closest species represented in the database which may not be the true species. The assumption of the analysis is illustrated in figure S11, where x_n denotes the similarity between the *n*th taxon and the query sequence. If x_1 represents the true species identity of the query sequence, then the PIA assumes that $x_1 > x_n$. If all species were represented in the database, then a classification intersection of $N_{1,2}$ would be assigned. If the true species was not represented in the database, but the other species were, then an intersection of $N_{1,2.3}$ would be assigned and so on. The approach is generally expected to be robust to violations in descending similarity through variable rates of evolution in different

lineages since these would be expected to lead to more distant taxa being intersected resulting in higher taxonomic ranks being assigned. However, under conditions of variable database representation local false positives may occur if the true species is absent, but two species forming a sister clade are present. For instance, if x_3 were the true species (100% similarity), but absent from the database, an incorrect assignation of $N_{1,2}$ could be made between x_1 and x_2 (e.g. 99% and 98% similarity respectively). A control analysis was carried out to estimate the accuracy of the PIA approach in the absence of database representation of the true species. The test was trained on our real dataset in order to establish the accuracy of the test in a way that reflected the areas of the database used in the real analysis. We selected the top BLAST hits to sedaDNA reads. In each case all entries associated with the taxon of the top hit were removed from the database so making the BLAST identified sequence unrepresented in the database. The BLAST identified sequence was then subject to PIA, and it was noted whether the taxonomic assignation was correct or not. Using this approach we estimate that the PIA has an accuracy of 81% in the absence of the true species in the database, which is a generally acceptable level in bioinformatic approaches. The true accuracy of the approach is therefore higher than this value assuming a proportion of the species sampled in the sedaDNA are in fact present in the database.

Under the stringency conditions applied (99% of query sequence covered by first hit, taxon diversity of 0.2), all of the clear false positives originally identified in the preliminary metagenomic analysis were automatically removed. None of the intersections identified in the analysis appeared to indicate species or taxonomic ranks that could not be explained as part of the expected profile. One of the control taxa (*Orcinus orca*) produced 2 reads that passed through the PIA, but were attributed to

the Cetartiodactyla, which is completely compatible with our findings of bovids and cervids.

The PIA identified six genera level intersections, Table S4. The most surprising was that of *Triticum*. Many of these reads were 100% identical to *Triticum* with decreasing similarity to the closest known sister taxa in *Aegilops*, so do not constitute a possible false positive as outlined above, Table S5. The next most frequent genus identified was *Canis*. Again these often represented 100% similarity, with decreasing similarity through different *Canis* species. *Bos* was inferred by intersections between *Bos taurus* and *Bison bison*. Two of the intersections indicating *Lolium* and *Quercus* respectively had levels of similarity of 97% and 90% respectively before dropping off across more distance taxa. The remaining genus level hits in *Lolium*, *Quercus* and *Populus* were characterized by 100% similarity between sedaDNA and database entry, followed by a falling similarities across more distant taxa.

4.5 Control library results.

The control library gave rise to 534068 reads from the MiSeq indicating the very low content of DNA in this sample (data not shown, but available from the authors on request). Of these 420161 were unassignable and most likely adaptor-dimers, the majority of the remainder were bacterial in origin. No higher plants were identifiable from this sample precluding the possibility that the *Triticum* signal we observed in the sedaDNA analysis could have originated from the library reagents.

4.6 Shotgun versus barcoding analysis for sensitivity

After the analysis we re-evaluated whether the evidence supported our notion that a shotgun approach had likely proved more sensitive than a barcoding approach would have to detect organisms in the palaeosol. The cost to a shotgun approach is

that although more genetic material can be accessed, at current database levels of content it is likely that most organisms will not be represented at most loci. Barcoding addresses this issue by having an extensive library for cross-reference, analogous to a pollen library for palynology. While the latter typically has tens of thousands of entries, barcode libraries typically contain between one and two thousand entries. The number of flowering plants alone exceeds 220,000 species, therefore the rate of false assignation even among barcoding approaches is likely to be high. Therefore both shotgun and barcoding approaches are prone to error from variable database representation, although undoubtedly the problem is more acute for shotgun data. We managed to assign with an estimated 81% accuracy shotgun data using the PIA. We noted how many reads were assignable as mitochondrial DNA (Table S2), and made a simplifying assumption that all mtDNA would be assignable. Assuming an even coverage of the mitochondrial genome, we estimate that loci making up the proportion of the mtDNA genome covered by a barcode target (circa 1%) would equate to about 10% of the total number of reads assigned in the PIA (Table S3). We therefore estimate that this sedaDNA analysis was able to access about tenfold more genetic space than a barcoding approach alone, and by inference was therefore at least tenfold more sensitive. Analysis of the reads accepted in the PIA analysis show that about 30% were from mitochondrial DNA. Assuming equal representation across the mitochondrial genome this would equate to 0.3% of the data coinciding with barcode targets which suggests our initial estimate of a tenfold increase in exploitable genetic space and therefore sensitivity is likely to be an underestimate.

4.7 Postmortem analysis

To assess postmortem deamination patterns, we recovered the original sequence data (i.e. with 10 5' bases reinstated) for all matches passing the filter (see

section 4.3). We then re-aligned those sequences to their most likely match as per the BLASTn top hit. However, the level of coverage of each top hit was far too low (mean 1.0) to obtain meaningful results from the mapDamage 2.0[36]. Each reference sequence was extended by 15 bases to allow for insertions subsequent to the core dates. Reverse complement alignments were used where BLASTn top hits were antisense alignments. The analysis showed that the modal fragment length of the sedaDNA accepted in the analysis was in the order of 110bp, Figure S12. The results indicate that DNA preservation under the constant conditions of a marine sediment may be surprisingly good, although the analysis was based on the subset of DNA which gave robust phylogenetic information which probably caused a selection bias for longer and less damaged fragments. A more in depth understanding of the base modification profile for this type of ancient DNA will require approaches that give high coverage of regions.

5. Early Holocene coastal assessment

5.1 Vibrocores samples and dates

A series of vibrocores were taken from 8 locations off the SE coast of Britain and subject to an extensive palaeoenvironmental analysis in a previous study[29]. The vibrocores sample Early Holocene land surfaces from various depths below the current sea level. Three of the core locations in particular yielded clear optically stimulated luminescence (OSL) dates, Fig 3 and Fig. S9, which gives an estimate of how long ago minerals were exposed to sunlight. These dates can be taken as the date of marine inundation of the old land surface at these locations.

5.2 Generalized map construction.

The vibrocore VC39 gave an OSL date most compatible with the date of marine inundation of the palaeosol at Bouldnor Cliff. The Global Relief model ETOPO1[37] is a 1 arc-minute global relief model of Earth's surface that integrates land topography and ocean bathymetry. It was built from numerous global and regional data sets. We utilized ETOPO1 to generate a regional coastline map with the sea level rendered to 30 meters below its current level around the southern coast of England. VC39 is therefore depicted on the coastal edge. Predicting past coastlines is complex[38] and the resulting palaeogeographical reconstruction from these data should be taken as an approximate guide of the past coastline. Some validation is provided for this reconstruction by the progressive OSL dates in shallower waters, and the location the Bronze Age inundated site of sea-henge of the Norfolk coast.

Table S1 Radiocarbon dates of organic material from Bouldnor Cliff (reference 9).

Dating Laboratory Identifier	Sample identifier	Material type†	13C (‰)	Radiocarbon age (BP)	Calibrated date¶ (cal BC)	Posterior density¥ (cal BC)
SUERC-11286	SO70 BCII 06	*Alnus glutinosa* r/w	-28.0	7030 ± 35	6000-5840	5990-5890
SUERC-11284	SO71 BCII 06	bark, unidentified	-28.0	7060 ± 35	6020-5880	6000-5920
SUERC-7579	MSO6 12 BCII	Monocotyledon leaf	-26.8	6925 ± 35	5900-5720	-
OxA-15698/721	MS05 16 BCII	*Betula* sp	-26.1	6938 ± 26	5890-5730	-
SUERC-11285	SO72 BCII 06	*Prunus* sp r/w	-28.6	7065 ± 35	6000-5880	6010-5940
OxA-15696	MS08 14 BCII	*Alnus glutinosa*, twig	-24.4	7013 ± 36	6000-5800	6000-5910
SUERC-7560	MS08 08 BCII	*Alnus glutinosa* r/w	-29.3	7105 ± 35	6050-5910	6010-5960
SUERC-7580	SO31 10 BCII	*Alnus glutinosa* r/w	-22.7	7115 ± 35	6060-5910	6020-5970
SUERC-7562	MS20 03 BCII	*Alnus glutinosa* r/w	-28.5	7130 ± 35	6070-5920	6030-5980
OxA-15718	MS08 40 BCII	*Corylus avellana*, twig	-27.2	7175 ± 45	6100-5980	6030-5980
OxA-15720	MS07 10-12 BCII	*Alnus glutinosa* r/w	-24.5	7125 ± 45	6070-5910	6040-5990
SUERC-8157	MS08 05 BCII	*Alnus glutinosa* r/w	-27.7	7110 ± 40	6060-5900	6040-5990
OxA-15697	MS07 22 BCII	*Betula* sp	-26.9	7110 ± 34	6060-5910	6060-5990
SUERC-7561	MS07 01 BCII	*Alnus glutinosa* r/w	-29.3	7175 ± 40	6090-5980	6060-5990

† r/w root or wood
¶ 95% confidence
¥ 95% probability

Table S1

Core ID	Mass (g)	DNA extracted (ng)	DNA (ng per g sediment)	Reads	Microbial	Non-microbial	No hits¶	Not assigned§	Mitochondrial
S308-1	4.36	40	9.17	16227092	1686437	41949	123857	14374437	510
S308-2	6.31	23	3.65	12634237	662182	33187	319858	11618683	697
S308-3	6.13	840	137.03	23028200	724480	38921	240184	22023382	1510
S308-4	6.14	692	112.70	19966670	658191	37775	192401	19077415	1990

¶ No similar sequences in database identified through BLAST
§ No taxonomic group assigned in the preliminary phylogenetic analysis

Table S2 Summary breakdown of sedaDNA reads.

Cluster assignation	Preliminary clusters counts 2308 0-2cm	S208 2-4cm	S308 4-6cm	S308 6-8cm	Post PIA counts 2308 0-2cm	S208 2-4cm	S308 4-6cm	S308 6-8cm
Anas platyrhynchos	11	0	26	31	1	0	0	0
Anolis carolinensis §	**793**	**150**	**113**	**104**	**0**	**0**	**0**	**0**
Bos taurus	44	121	17	20	9	31	0	0
Canis	246	447	28	20	23	43	0	0
Capreolus capreolus	0	7	0	0	0	2	0	0
Ceratitis capitata	2718	411	564	481	0	0	0	0
Cervus nippon	0	5	0	0	0	1	0	0
Culex pipiens	348	166	129	90	0	0	0	0
Drosophila melanogaster	35	66	78	74	0	0	0	0
Echinops telfairi §	**106**	**24**	**35**	**34**	**0**	**0**	**0**	**0**
Equus caballus	15	10	16	17	0	0	0	0
Felis catus	216	15	16	19	0	0	0	0
Gallus gallus	408	349	40	32	8	6	0	0
Meleagris gallopavo	13	0	0	17	0	0	0	0
Monodelphis domestica §	**135**	**8**	**26**	**27**	**0**	**0**	**0**	**0**
Odobenus rosmarus divergens §	**0**	**0**	**23**	**14**	**0**	**0**	**0**	**0**
Orcinus orca §	**0**	**6**	**0**	**7**	**0**	**3**	**0**	**0**
Ovis aries	13	16	17	25	0	5	0	0
Rodentia	164	180	381	336	3	0	0	0
Salmo salar	165	18	17	15	0	1	0	0
Sarcophilus harrisii §	**2845**	**2943**	**10509**	**10859**	**0**	**0**	**0**	**0**
Sorex araneus	0	0	54	53	0	0	0	0
Sus scrofa	27	23	28	25	2	2	0	0
Ursus maritimus	55	83	13	14	1	5	0	0
Mustela	186	68	31	26	2	4	0	0
Magnoliophyta	2979	1865	3107	3574	96	84	34	142
Totals	11522	6981	15268	15914	145	187	34	142

§Control taxa are shown in bold

Table S3 Read counts from taxonomic clusters before and after PIA

Taxonomic rank					Core			
higher rank	class/order	family	subfamily	subfamily/tribe	genus	S308 0-2cm	S308 2-4cm	S308 4-6cm	S308 6-8cm
		Poaceae				3	4	12	48
			BEP clade			2	6	11	63
				Pooideae		4	2		5
					Lolium	1	1		
				Triticeae		15	11		1
					Triticum	63	54	3	5
			PACMAD clade			1			9
				Arundineae					3
				Centothecoideae		1	2		10
				Panicoideae					
				Andropogoneae				4	26
				Paniceae					1
Gunneridae						1			
Eurosids I							1		4
	Rosidae					2		1	3
		Fagaceae							3
					Quercus				5
		Cucurbitaceae					2		
		Rosaceae							
				Maloideae					1
		Brassicaceae				1	1		1
		Salicaceae			Populus	1		1	1
	Asterids	Solanaceae				1			
Neognathae						4	2		
		Phasinidae				2	4		
			Phasianinae			3			
Eutheria						8	7		
	Carnivora					2	7		
	Caniformia					8	22		
		Ursidae					1		
		Canidae				5	4		
			Caninae				1		
					Canis	6	13		
Cetartiodactyla						1	3		
	Pecora					3	2		
		Bovidae				6	29		
					Bos		4		
		Cervidae							
			Odocoileinae				2		
			Cervinae				1		
Rodentia			Murinae			1			
Teleostomi			Salmoninae				1		

Table S4 Flora and fauna profile of Bouldnor Cliff given by frequencies of read assignations of sedaDNA.

Figure S1Chronological model of Bouldnor Cliff based on radiocarbon dates[9]. For details on samples see Table S1.

Figure S2. Summaries of monoliths from Bouldnor Cliff 2 showing depths and relationships with dated stratigraphic units.

Figure S3 Map of Bouldnor Cliff and Tanners Hard core locations (shown in red).

Figure S4 Pollen diagram[9] for Bouldnor Cliff based on cores MS04/05/08

Figure S5 Pollen diagram[9] for Bouldnor Cliff based on core MS07

Figure S6 Pollen diagram[18] for Tanners Hard lower peat section

Figure S7 Pollen diagram[18] for Tanners Hard higher peat section

Figure S8 Auroch left astragalus bone excavated from Bouldnor Cliff in 2014.

Figure S9 Posterior densities[29] of OSL dates obtained from vibrocores 29, 39 and 51.

Figure S10 Base quality analysis of Illumina reads from cores A. S2308 0-2cm B. S2308 2-4cm C. S2308 4-6cm D. S2308 6-8cm

Figure S11 Phylogenetic structure of BLAST distributions. Theoretical phylogenetic distribution of species identified to be a similar to a query sequence. The percentage similarity (over 100% of the query length) is denoted by x_n for the *n*th taxon. N denotes nodes.

Figure S12 Frequency distribution of Illumina read lengths of combined cores.

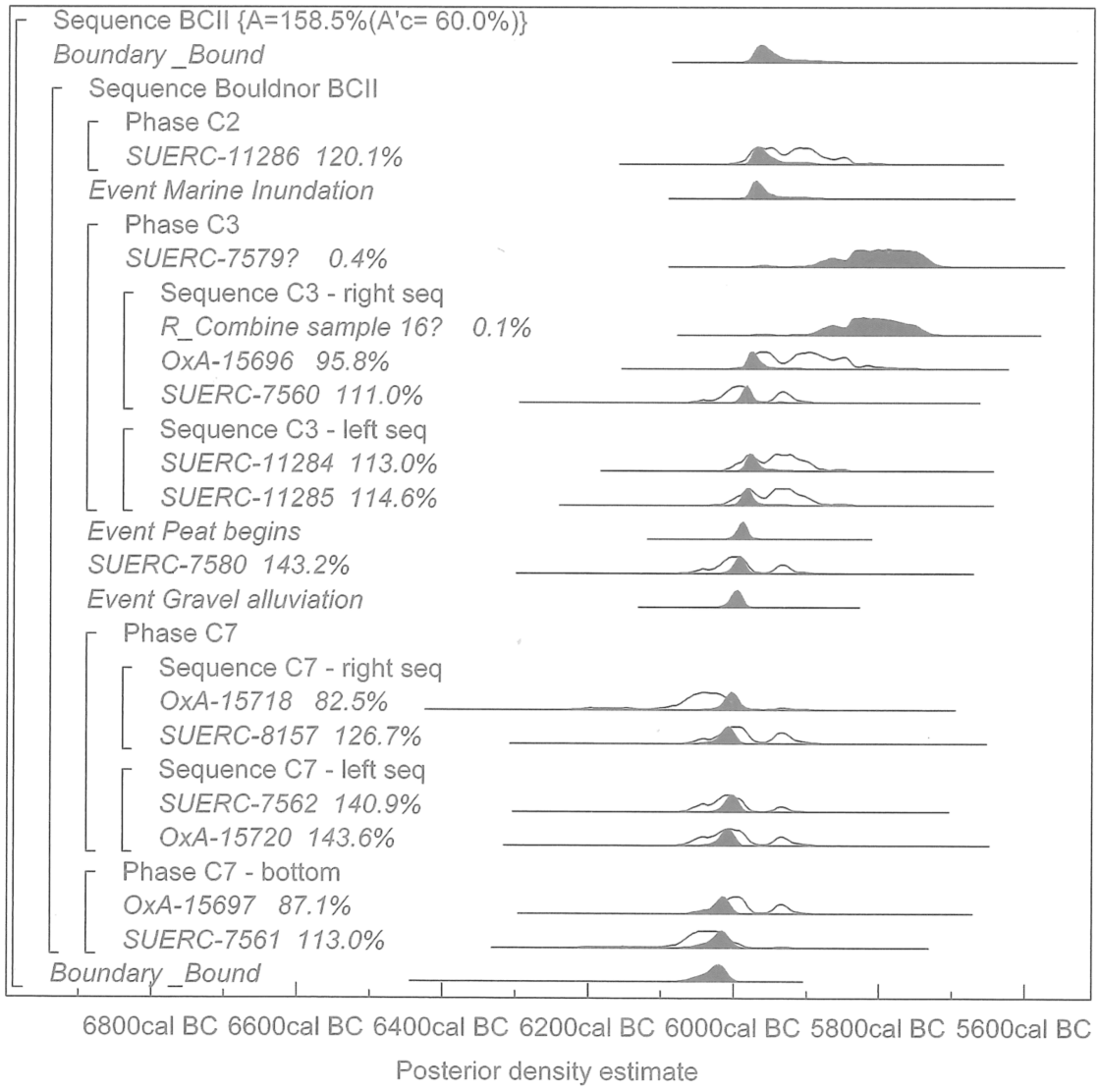

Sequence BCII {A=158.5%(A'c= 60.0%)}
Boundary _Bound
Sequence Bouldnor BCII
Phase C2
SUERC-11286 120.1%
Event Marine Inundation
Phase C3
SUERC-7579? 0.4%
Sequence C3 - right seq
R_Combine sample 16? 0.1%
OxA-15696 95.8%
SUERC-7560 111.0%
Sequence C3 - left seq
SUERC-11284 113.0%
SUERC-11285 114.6%
Event Peat begins
SUERC-7580 143.2%
Event Gravel alluviation
Phase C7
Sequence C7 - right seq
OxA-15718 82.5%
SUERC-8157 126.7%
Sequence C7 - left seq
SUERC-7562 140.9%
OxA-15720 143.6%
Phase C7 - bottom
OxA-15697 87.1%
SUERC-7561 113.0%
Boundary _Bound
6800cal BC 6600cal BC 6400cal BC 6200cal BC 6000cal BC 5800cal BC 5600cal BC
Posterior density estimate

Figure S1

Monolith Sample MS04 & MS07

Absolute depths (+/- 0.1m)	Dates	Unit	Stratigraphy	Morphology	*Pollen [Zone]*	Diatoms/Forams
10.04	*Posterior density estimate* (95% probability)	**Unit 1**	*Grey silt*	Brackish water	*Changes from alluvial or marine sediments accompanied by changes in pollen spectra.*	
10.06				Tidal mudflat to lower saltmarsh	*Poaceae important and reworking of pre-Quaternary pollen.*	
10.08						
10.10						
10.12						
10.14						
10.16						
10.18			**MS04**			
10.20						Forams: typical mud-flat fauna
10.22						
10.24						
10.26					*Final Boreal-Atlantic sea-level expansion*	
10.28						
10.30						
10.32						
10.34						
10.36						
10.38						
10.40						
10.42						Planktonic and shelf specimens
10.44						
10.46						
10.48						
10.50						Forams becoming scarce
10.52						
10.54						
10.56						
10.58					*Increase in Poaceae and reed swamp/fen taxa*	
10.60		**Unit 2**	*Transition silty clay*			
10.62	*5990–5890 cal BC*				*Increasing wetness caused by rising local ground water*	
10.64	*[SUERC-11286]*					
10.66						
10.68			*Peat*			
10.70	*6000–5920 cal BC*					
10.72	*[SUERC-11284]*	**Unit 3**				
10.74						
10.76						
10.78						
10.80						
10.82	*6010–5940 cal BC*					
10.84	*[SUERC-11285]*	**Unit 6**	*Gravel*			
10.86						
10.88	*6020–5970 cal BC*					
10.90	*[SUERC-7580]*					
10.92						
10.94						
10.96						
10.98						
11.00						
11.02						
11.04						
11.06			*Med/coarse sand*			
11.08	*6030–5980 cal BC*					
11.10	*[SUERC-7562]*	**Unit 7**				
11.12						
11.14				**[SO7:3]**		
11.16				(Comparable to SO8:1c, Zone 1c)		
11.18						
11.20	*6040–5990 cal BC*					
11.22	*[OxA-15720]*		**MS07**			
11.24		**Unit 7**	*Humic sand*			
11.26				**[SO7:2]**		
11.28						
11.30						
11.32						
11.34						
11.36				**[SO7:1]**		
11.38						
11.40						
11.42						
11.44			*Wood*	Mesolithic artefacts found		
11.46			*Peaty sand*	Horizon of darker more humic sands/silts		
11.48						
11.50						
11.52			*Humic sand*			
11.54	*[SUERC-7562]*					
11.56	*6060–5990 cal BC*					
11.58	*6060–5990 cal BC*			Fluvial Gravels		
11.60	*[OxA-15697]*		*Gravel*			
11.62						
11.64		**Unit 12**				
11.66						
11.68			VOID			

Monolith samples MS05 & MS08

Dating	Unit	Stratigraphy	Morphological units
	Unit 1	**MS05**	
			brackish water
		Soft silty clay	Prob tidal mudflat to lower saltmarsh
		Graded boundry	
	Unit 2	Graded boundry	
		Stiff silty clay	Low energy
	Unit 2.1	*Organic mud*	brackish or freshwater
		Clear boundary	(Probably erroded boundary)
	Unit 3.1	Clear boundary *Peat*	Oxidised & desiccated
[SO8:1c] *Predominantly peats Alnus dominates with Quercus and Corylus Pinus declines*	Unit 3.2	**MSO8**	Prob significant period of lowered ground water levels
6000–5910 cal BC [OxA-15696]		Graded boundary	
[SO8:1b]	**Unit 7.1**	*Med/course sand*	Organic mud
6010–5960 cal BC [SUERC-7560]		Sharp boundary	Possible old landsurface prior to peat development.
	Unit 7.2	*Med/course sand* Clear boundry	deposit formed in fluvial or brackish water environment
Pinus declines Quercus and Ulmus increase	Unit 7.3	*Silty clay* Graded boundary	Deposit laid down in standing water
	Unit 7.4	*Med/course sand*	Light olive brown sandy/clay
[SO8:1a] *[OxA-15718]*	Unit 7.5		unit 7.4 & 7.5: Either; channel lag deposits, tidal creeks with episodic running and standing water conditions or point-bar braided freshwater channel
6030–5980 cal BC 6040–5990 cal BC [SUERC-8157]			

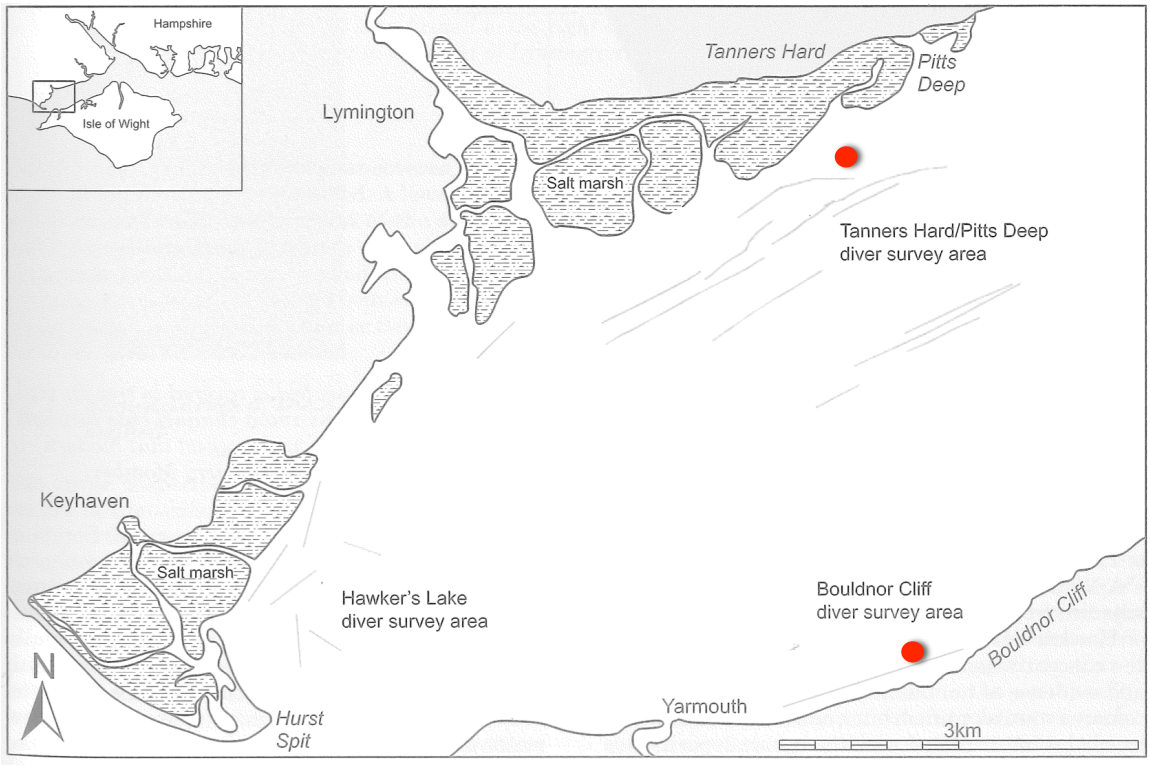

Hampshire
Isle of Wight
Lymington
Tanners Hard
Pitts Deep
Salt marsh
Tanners Hard/Pitts Deep diver survey area
Keyhaven
Salt marsh
Hawker's Lake diver survey area
Bouldnor Cliff diver survey area
Bouldnor Cliff
N
Hurst Spit
Yarmouth
3km

Figure S3

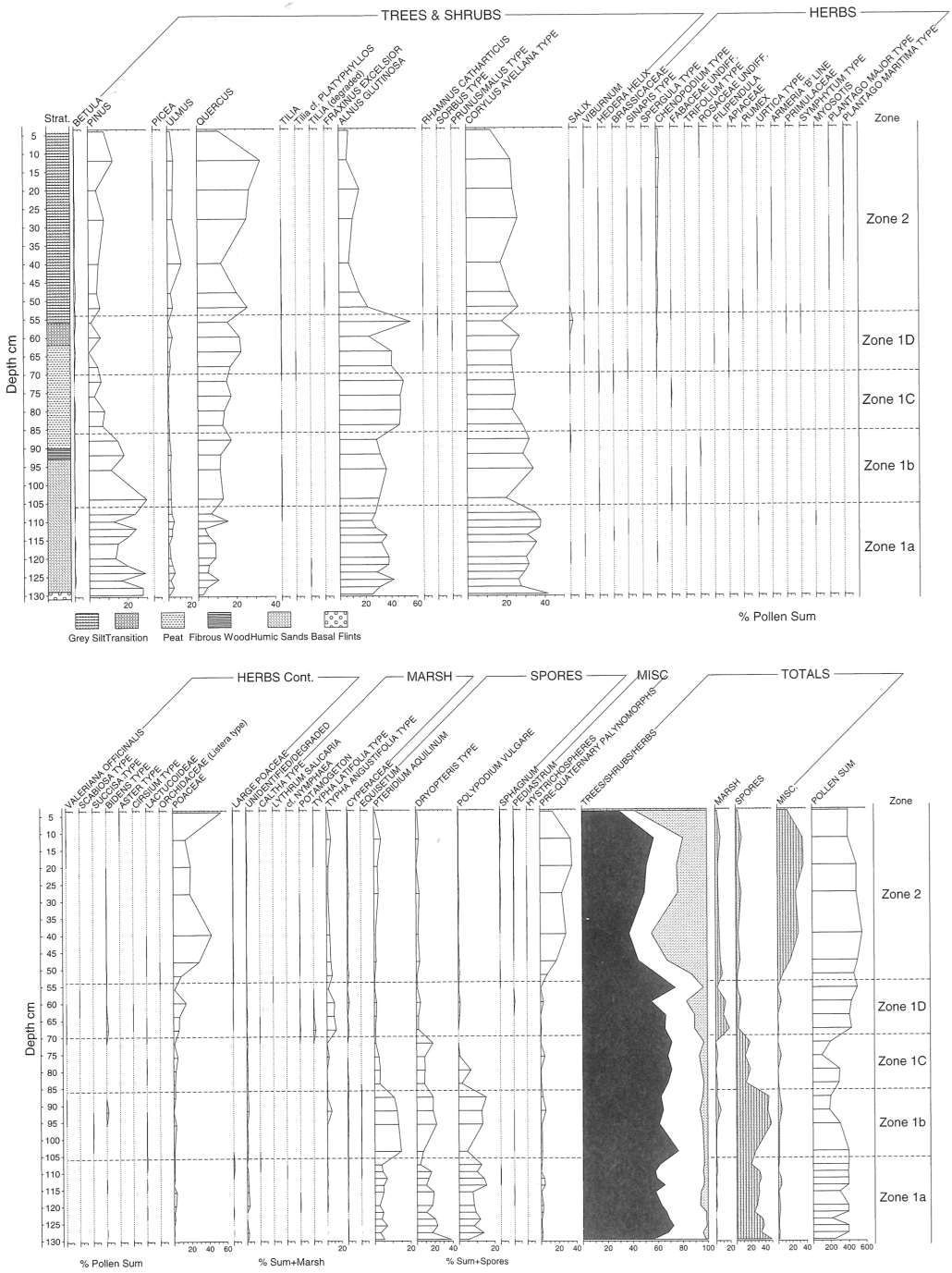

TREES & SHRUBS
HERBS
HERBS Cont.
MARSH
SPORES
MISC
TOTALS
Strat.
Depth cm
Zone
Zone 2
Zone 1D
Zone 1C
Zone 1b
Zone 1a
% Pollen Sum
% Sum+Marsh
% Sum+Spores
Grey Silt
Transition
Peat
Fibrous Wood
Humic Sands
Basal Flints
BETULA
PINUS
PICEA
ULMUS
QUERCUS
TILIA
Tilia cf. PLATYPHYLLOS
TILIA (degraded)
FRAXINUS EXCELSIOR
ALNUS GLUTINOSA
RHAMNUS CATHARTICUS
SORBUS TYPE
PRUNUS/MALUS TYPE
CORYLUS AVELLANA TYPE
SALIX
VIBURNUM
HEDERA HELIX
BRASSICACEAE
SINAPIS TYPE
SPERGULA TYPE
CHENOPODIUM TYPE
FABACEAE UNDIFF.
TRIFOLIUM TYPE
ROSACEAE UNDIFF.
FILIPENDULA
APIACEAE
RUMEX
URTICA TYPE
ARMERIA 'B' LINE
PRIMULACEAE
SYMPHYTUM TYPE
MYOSOTIS
PLANTAGO MAJOR TYPE
PLANTAGO MARITIMA TYPE
VALERIANA OFFICINALIS
SCABIOSA TYPE
SUCCISA TYPE
BIDENS TYPE
ASTER TYPE
CIRSIUM TYPE
LACTUCOIDEAE
ORCHIDACEAE
POACEAE
LARGE POACEAE (Lisiera type)
UNIDENTIFIED/DEGRADED
CALTHA TYPE
LYTHRUM SALICARIA
cf. NYMPHAEA
POTAMOGETON
TYPHA LATIFOLIA TYPE
TYPHA ANGUSTIFOLIA TYPE
CYPERACEAE
EQUISETUM
PTERIDIUM AQUILINUM
ORYOPTERIS TYPE
POLYPODIUM VULGARE
SPHAGNUM
PEDIASTRUM
HYSTRICOSPHERES
PRE-QUATERNARY PALYNOMORPHS
TREES/SHRUBS/HERBS
MARSH
SPORES
MISC.
POLLEN SUM

Figure S4

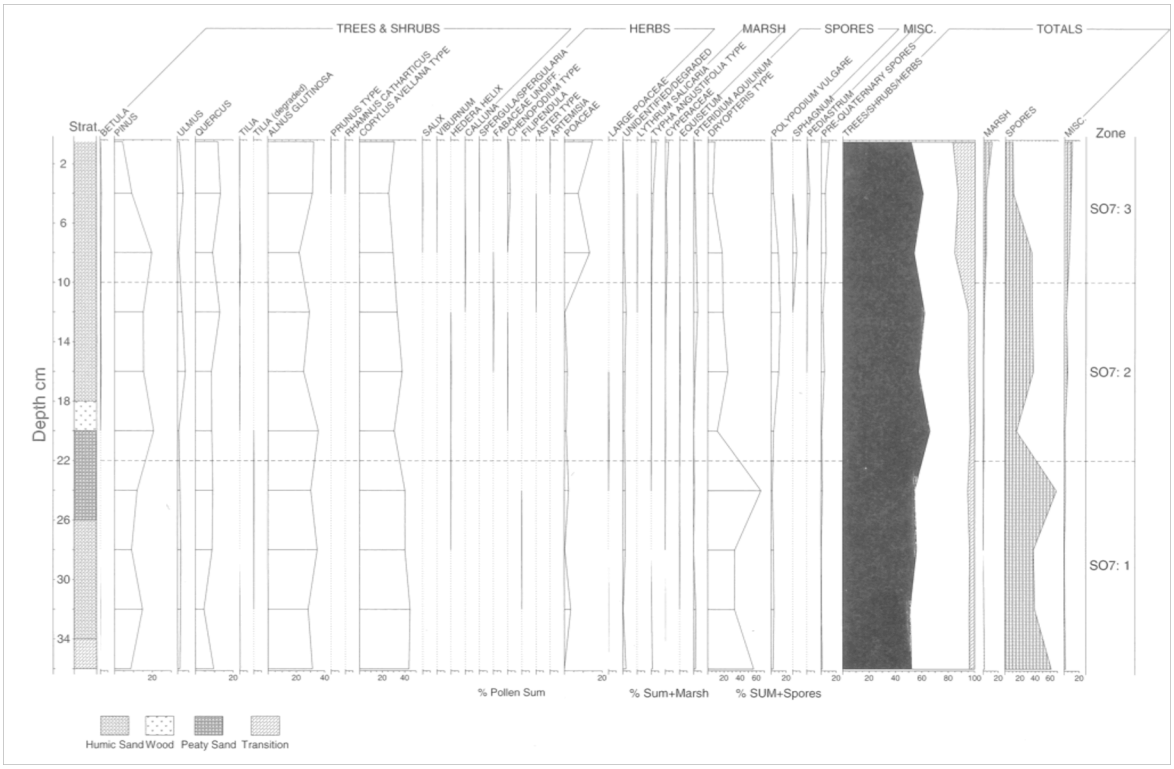
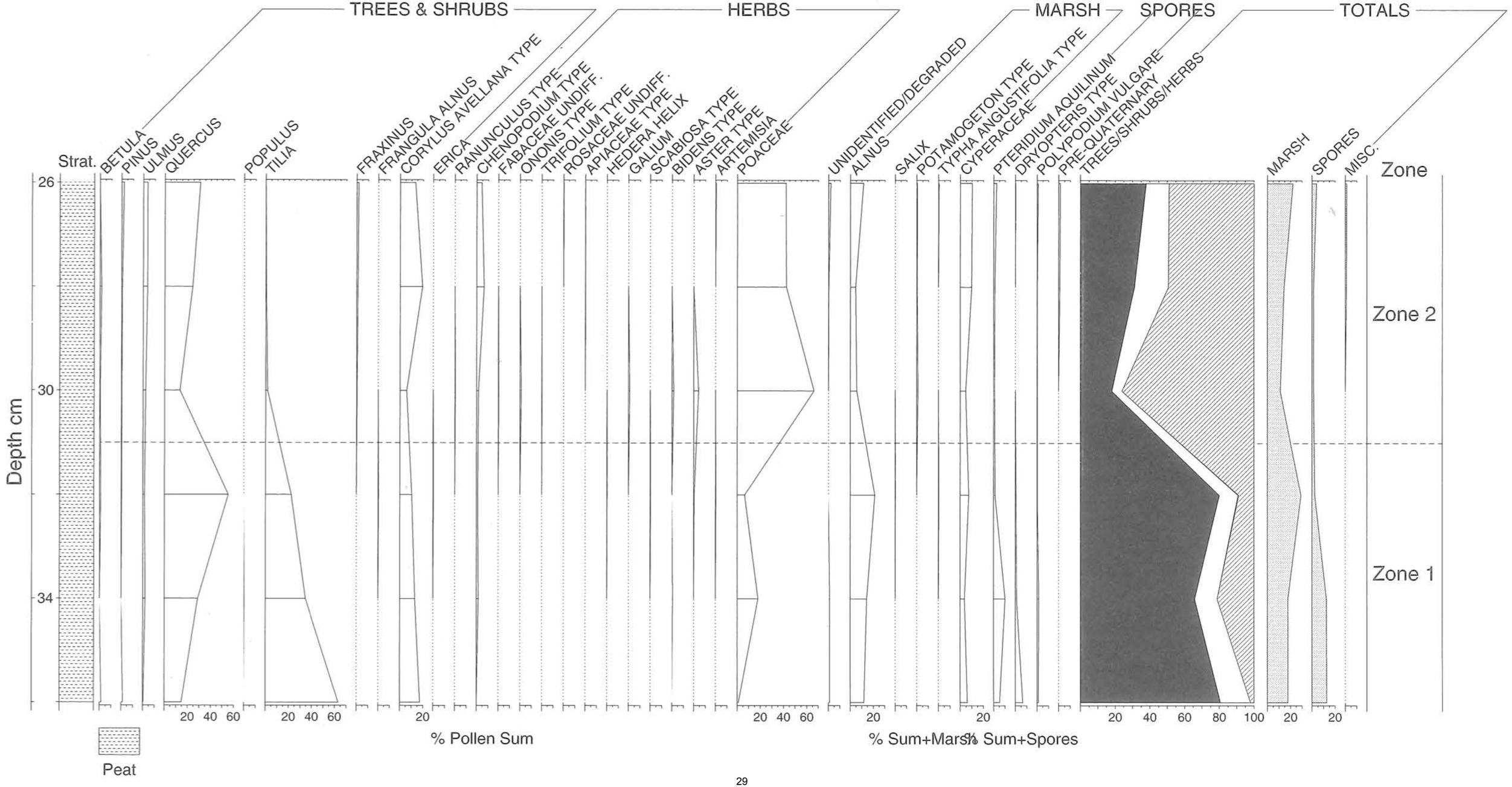
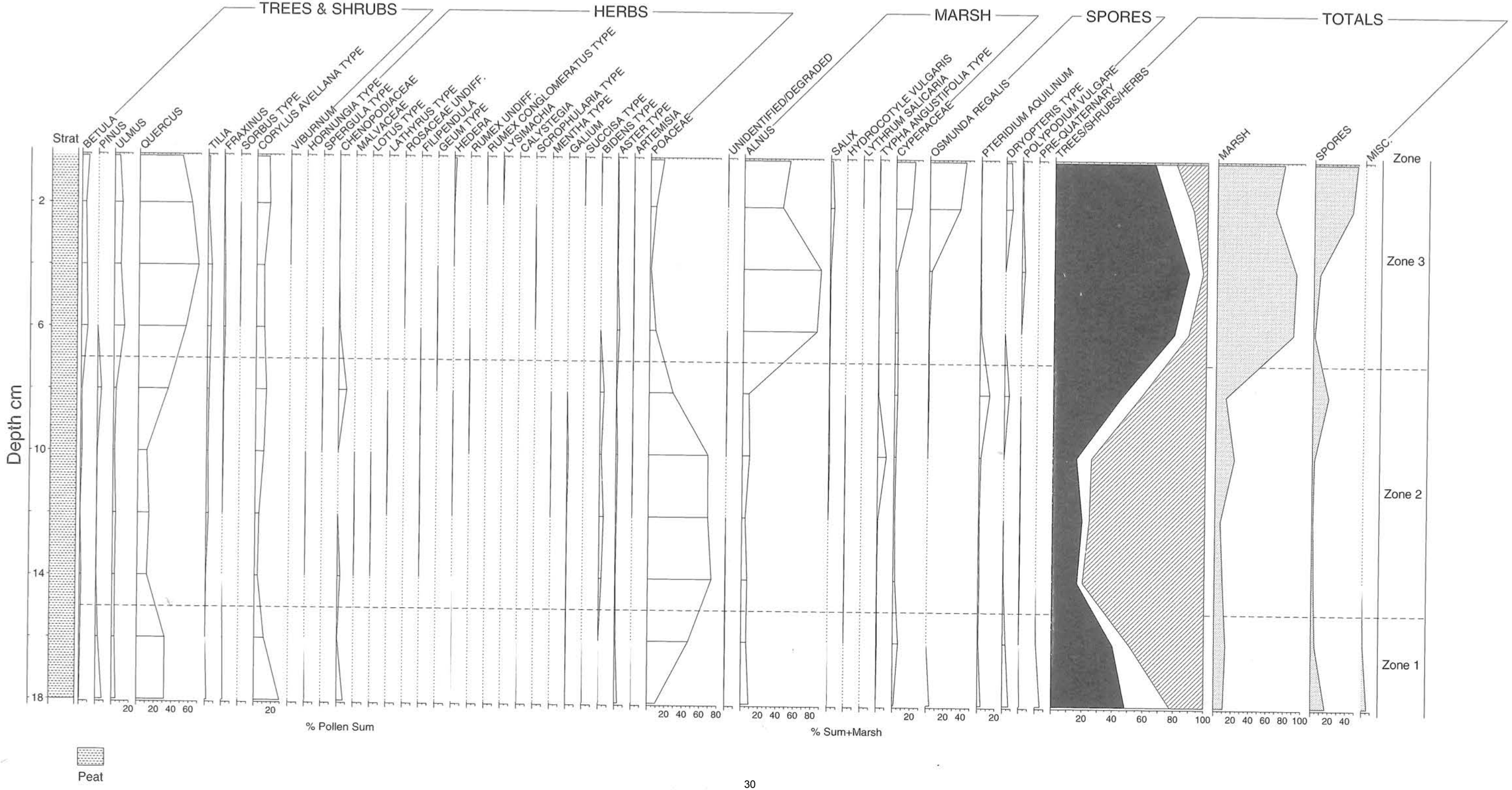
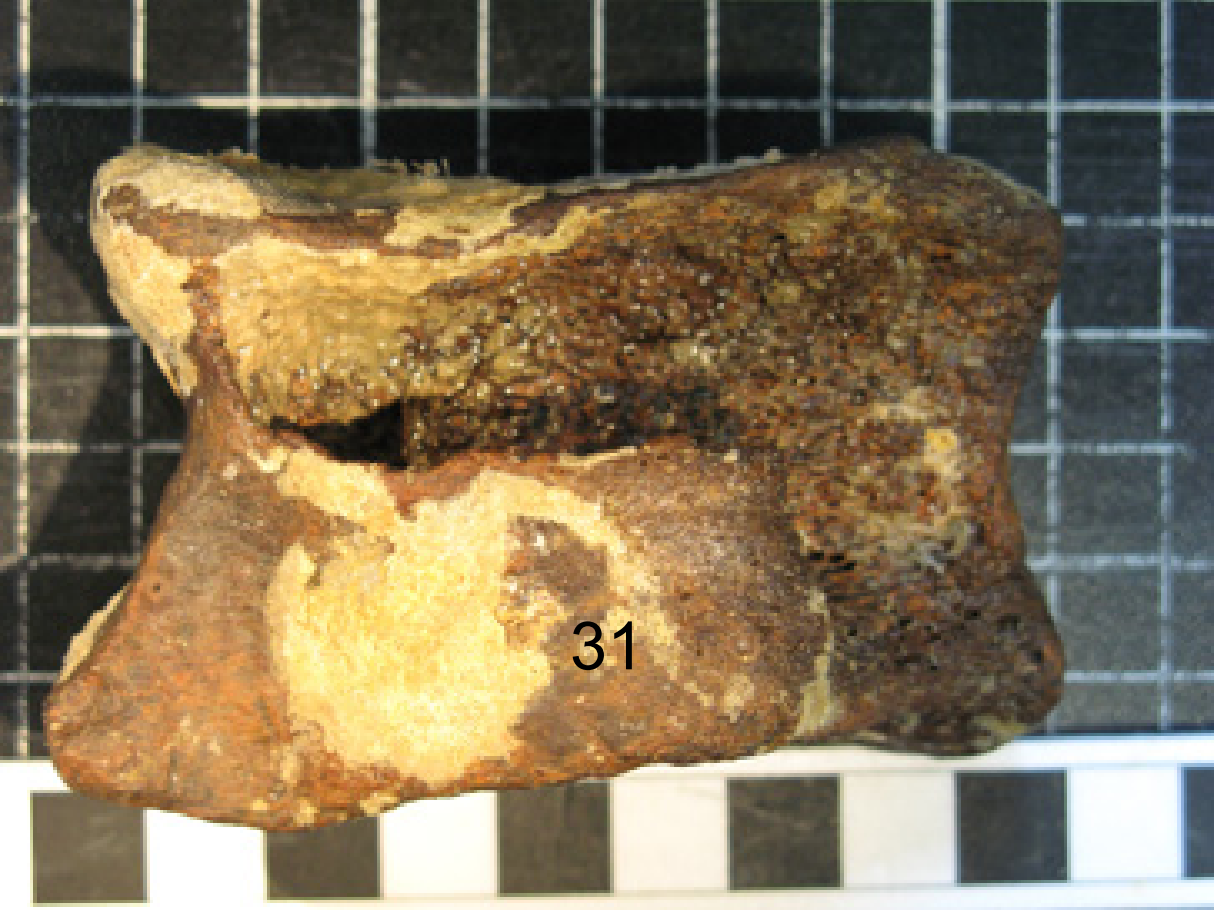
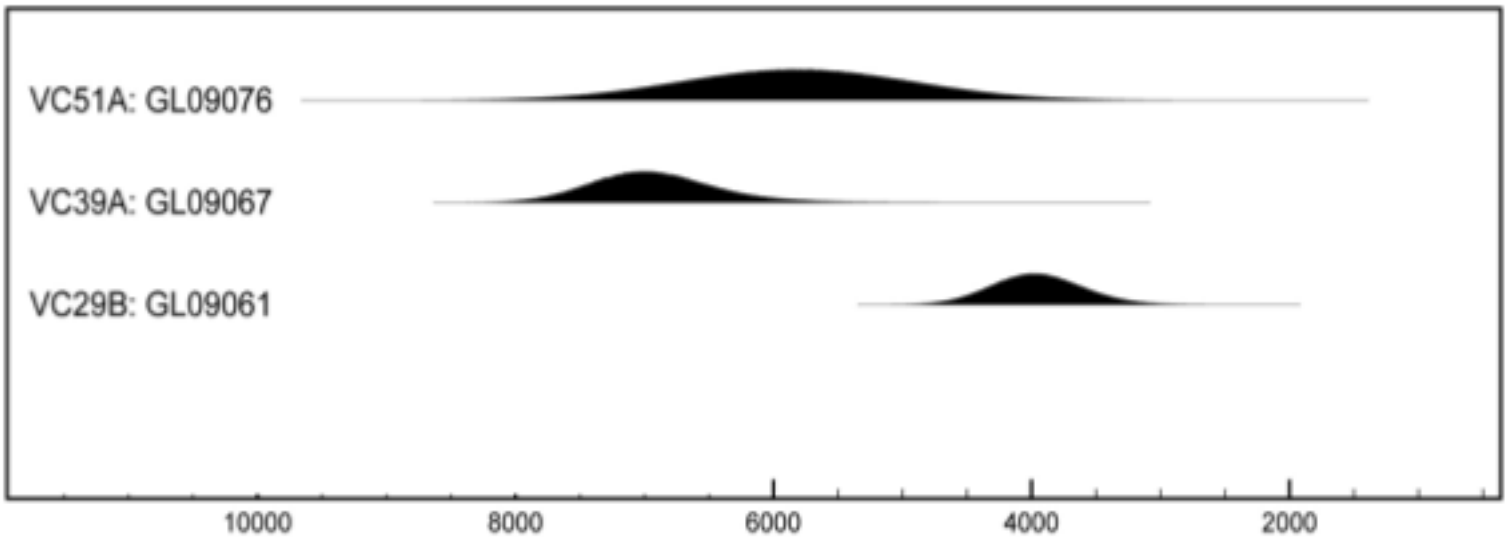

TREES & SHRUBS
HERBS
MARSH
SPORES
MISC.
TOTALS
Strat.
BETULA
PINUS
ULMUS
QUERCUS
TILIA
TILIA (degraded)
ALNUS GLUTINOSA
PRUNUS TYPE
RHAMNUS CATHARTICUS
CORYLUS AVELLANA TYPE
SALIX
VIBURNUM
HEDERA HELIX
CALLUNA
SPERGULA-SPERGULARIA
FABACEAE UNDIFF. TYPE
CHENOPODIUM
FILIPENDULA
ASTER TYPE
ARTEMISIA
POACEAE
LARGE POACEAE
UNIDENTIFIED/DEGRADED
LYTHRUM SALICARIA
TYPHA ANGUSTIFOLIA
CYPERACEAE
EQUISETUM
PTERIDIUM AQUILINUM
DRYOPTERIS TYPE
POLYPODIUM VULGARE
SPHAGNUM
PEDIASTRUM
PRE-QUATERNARY SPORES
TREES/SHRUBS/HERBS
MARSH
SPORES
MISC.
Zone
Depth cm
2
6
10
14
18
22
26
30
34
SO7: 3
SO7: 2
SO7: 1
% Pollen Sum
% Sum+Marsh
% SUM+Spores
Humic Sand
Wood
Peaty Sand
Transition

Figure S5


TREES & SHRUBS
HERBS
MARSH
SPORES
TOTALS
Strat.
BETULA
PINUS
ULMUS
QUERCUS
POPULUS
TILIA
FRAXINUS
FRANGULA ALNUS
CORYLUS AVELLANA TYPE
ERICA
RANUNCULUS TYPE
CHENOPODIUM TYPE
FABACEAE UNDIFF.
ONONIS TYPE
TRIFOLIUM TYPE
ROSACEAE UNDIFF.
APIACEAE TYPE
HEDERA HELIX
GALIUM
SCABIOSA TYPE
BIDENS TYPE
ASTER TYPE
ARTEMISIA
POACEAE
UNIDENTIFIED/DEGRADED
ALNUS
SALIX
POTAMOGETON TYPE
TYPHA ANGUSTIFOLIA TYPE
CYPERACEAE
PTERIDIUM AQUILINUM
DRYOPTERIS TYPE
POLYPODIUM VULGARE
PRE-QUATERNARY
TREES/SHRUBS/HERBS
MARSH
SPORES
MISC.
Zone
Zone 2
Zone 1
Depth cm
26
30
34
20 40 60
20 40 60
20
20 40 60
20
20
20 40 60 80 100
20
20
% Pollen Sum
% Sum+Marsh
% Sum+Spores
Peat

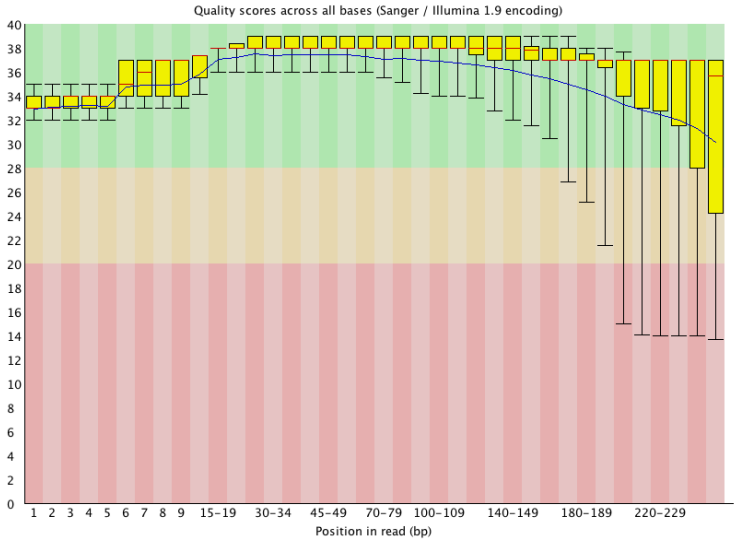

TREES & SHRUBS
HERBS
MARSH
SPORES
TOTALS
Strat
BETULA
PINUS
ULMUS
QUERCUS
TILIA
FRAXINUS
SORBUS TYPE
CORYLUS AVELLANA TYPE
VIBURNUM
HORNUNGIA TYPE
SPERGULA TYPE
CHENOPODIACEAE
MALVACEAE
LOTUS TYPE
LATHYRUS TYPE
ROSACEAE UNDIFF.
FILIPENDULA
GEUM TYPE
HEDERA
RUMEX UNDIFF.
RUMEX CONGLOMERATUS TYPE
LYSIMACHIA
CALYSTEGIA
SCROPHULARIA TYPE
MENTHA TYPE
GALIUM
SUCCISA TYPE
BIDENS TYPE
ASTER TYPE
ARTEMISIA
POACEAE
UNIDENTIFIED/DEGRADED
ALNUS
SALIX
HYDROCOTYLE VULGARIS
LYTHRUM SALICARIA
TYPHA ANGUSTIFOLIA TYPE
CYPERACEAE
OSMUNDA REGALIS
PTERIDIUM AQUILINUM
DRYOPTERIS TYPE
POLYPODIUM VULGARE
PRE-QUATERNARY
TREES/SHRUBS/HERBS
MARSH
SPORES
MISC.
Zone
Zone 3
Zone 2
Zone 1
Depth cm
2
6
10
14
18
20 20 40 60
20
20 40 60 80
20 40 60 80
20
20 40
20
20 40 60 80 100
20 40 60 80 100
20 40
% Pollen Sum
% Sum+Marsh
Peat


31

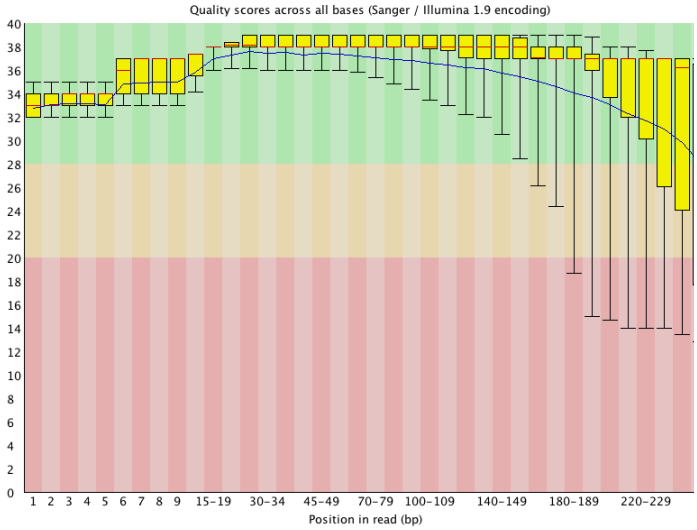

VC51A: GL09076
VC39A: GL09067
VC29B: GL09061
10000
8000
6000
4000
2000

Years cal. BC

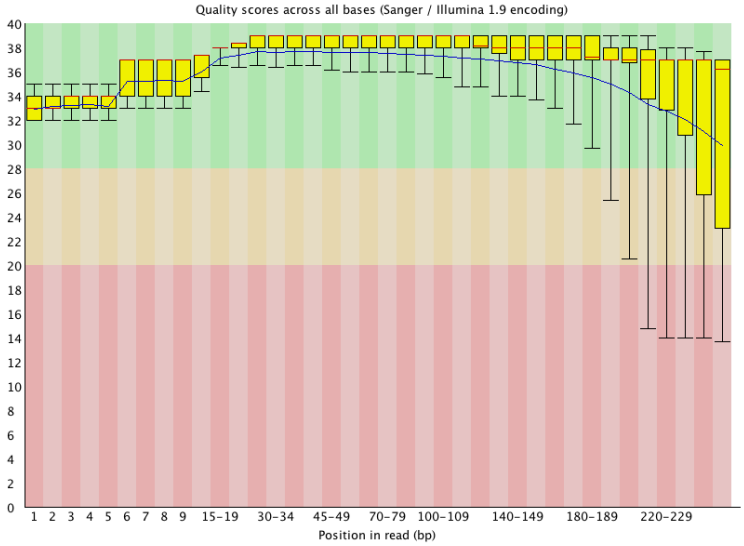
A


Quality scores across all bases (Sanger / Illumina 1.9 encoding)
Position in read (bp)

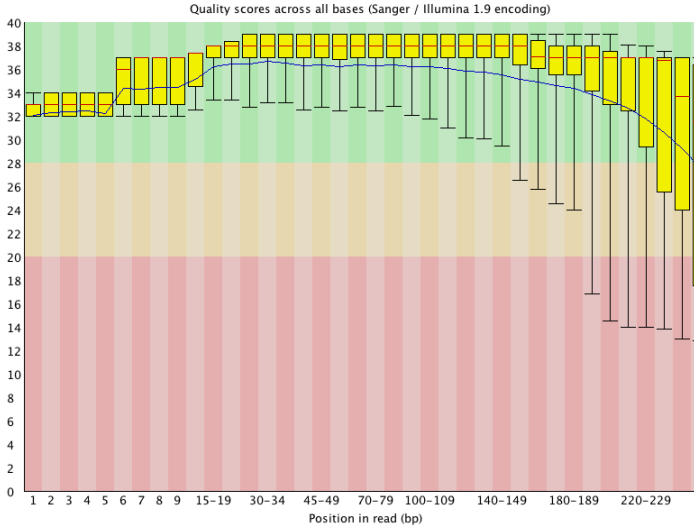
B

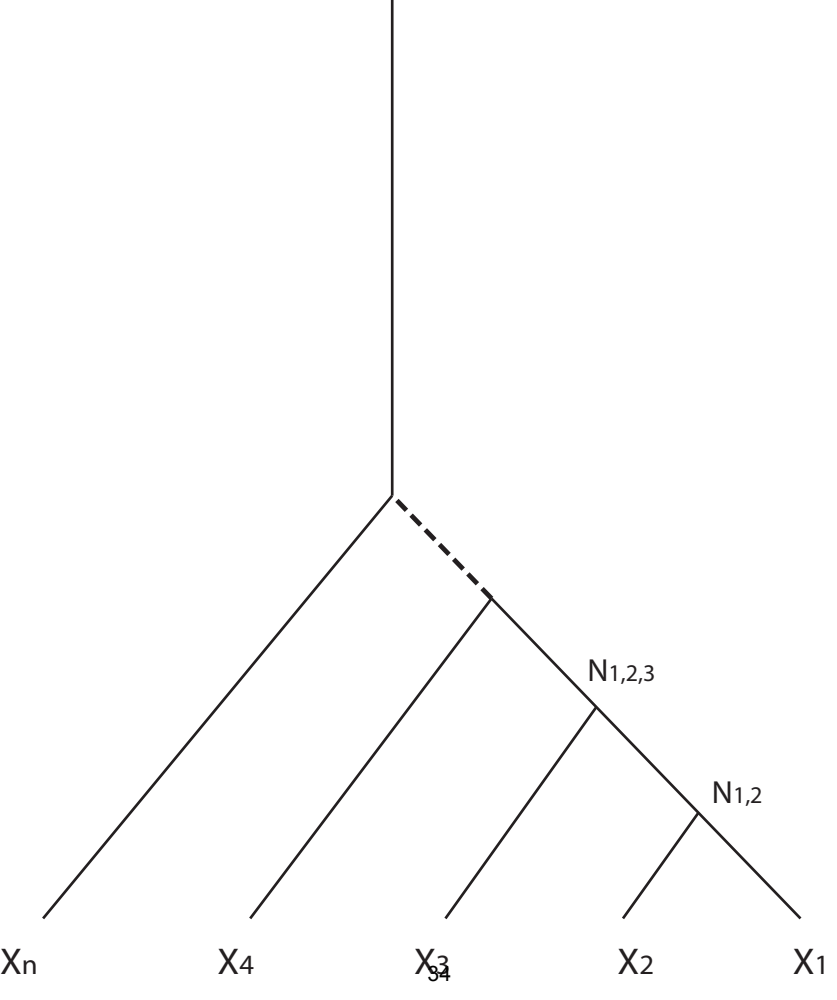

Quality scores across all bases (Sanger / Illumina 1.9 encoding)
Position in read (bp)

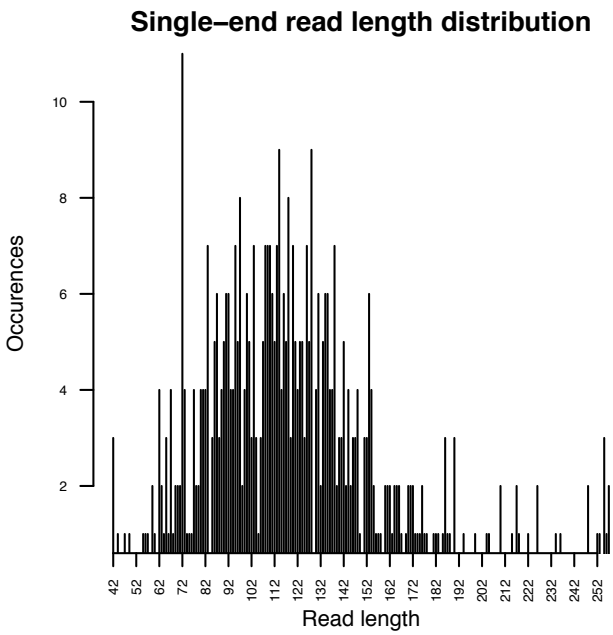
C


Quality scores across all bases (Sanger / Illumina 1.9 encoding)
Position in read (bp)

D


Quality scores across all bases (Sanger / Illumina 1.9 encoding)
Position in read (bp)


N1,2,3
N1,2
Xn
X4
X3
X2
X1


Single–end read length distribution
Occurences
Read length

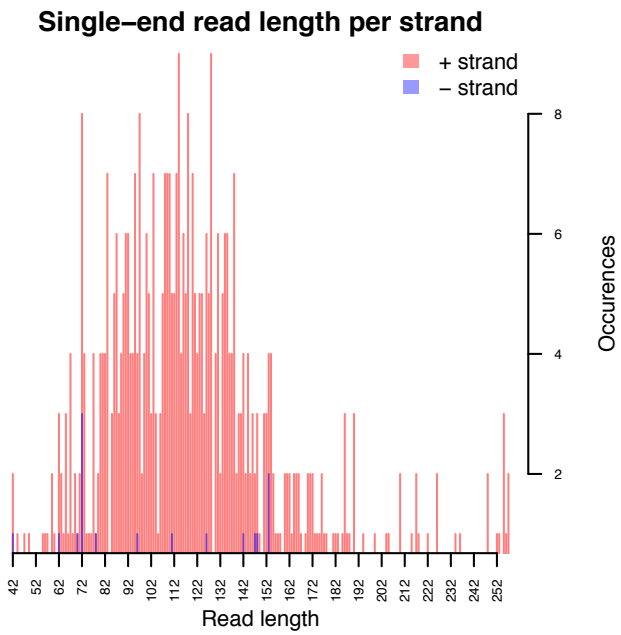

Single–end read length per strand
+ strand
– strand
Occurences
Read length

Figure S12