



OPEN

Lake microbiome and trophic fluctuations of the ancient hemp retting

Olga Iwańska^{1,2,9}, Przemysław Latoch^{1,3,9}, Magdalena Suchora^{4,5}, Irena A. Pidek⁴, Miłosz Huber⁴, Iwona Bubak⁶, Natalia Kopik^{1,2}, Mariia Kovalenko^{1,2}, Michał Gąsiorowski⁷, Jean-Paul Armache⁸ & Agata L. Starosta^{1,2}✉

Lake sediments not only store the long-term ecological information including pollen and microfossils but are also a source of sedimentary DNA (*sedDNA*). Here, by the combination of traditional multi-proxy paleolimnological methods with the whole-metagenome shotgun-sequencing of *sedDNA* we were able to paint a comprehensive picture of the fluctuations in trophic and bacterial diversity and metabolism of a small temperate lake in response to hemp retting, across the past 2000 years. Hemp retting (HR), a key step in hemp fibre production, was historically carried out in freshwater reservoirs and had a negative impact on the lake ecosystems. In Lake Slone, we identified two HR events, during the late stage of the Roman and Early Medieval periods and correlated these to the increased trophic and imbalanced lake microbiome. The metagenomic analyses showed a higher abundance of Chloroflexi, Planctomycetes and Bacteroidetes and a functional shift towards anaerobic metabolism, including degradation of complex biopolymers such as pectin and cellulose, during HR episodes. The lake eutrophication during HR was linked to the allochthonous, rather than autochthonous carbon supply—hemp straws. We also showed that the identification of HR based on the palynological analysis of hemp pollen may be inconclusive and we suggest the employment of the fibre count analysis as an additional and independent proxy.

The lake ecosystems are sensitive to external impacts, such as climate change and anthropogenic factors¹. The responses to stressors applied either to the lake itself or to the catchment are recorded in lake sediments constituting well-established sources of historical archives, which can be investigated via a number of paleolimnological proxies^{2,3}. In addition to the classical paleolimnological analyses, the lake sediments have also recently become subjects of the next-generation sequencing (NGS) studies resulting in amplicon and shotgun metagenomic records⁴. Lake sediments constitute anaerobic and relatively inert conditions and are therefore fairly adequate for the preservation of the sedimentary DNA (*sedDNA*)⁴. The microbiome fluctuations over long periods of time in response to the environmental and anthropogenic stressors can thus be reconstructed from the lake sediments and can provide new and complementary information to the classical paleolimnological analyses⁵. As a result, the field of metagenomics in paleoecological research is new and rapidly developing with an increasing number of studies applying shotgun sequencing^{5,6} and combining multi-proxy paleo-environmental data with the NGS data^{1,7,8}.

The use and processing of hemp (*Cannabis sativa* L.) in Europe dates back to the Bronze age, where it spread during the Roman times and reached its peak in the early Middle Ages^{9,10}. Hemp retting (HR) is a central step of hemp fibre production which was historically carried out in freshwater reservoirs like ponds, lakes, or bogs (water-retting). HR on the fields (dew-retting) and in the artificial water reservoirs emerged later, probably as an attempt to avoid pollution of freshwater reservoirs¹¹. Indeed, HR has a high environmental impact due to the biological nature of the process. During water-retting the straws are submerged in water and soluble materials

¹ECOTECH-Complex and Institute of Biological Sciences, Maria Curie-Skłodowska University, Lublin, Poland. ²Institute of Biochemistry and Biophysics, Polish Academy of Science, Warsaw, Poland. ³Polish-Japanese Academy of Information Technology, Warsaw, Poland. ⁴Institute of Earth and Environmental Sciences, Maria Curie-Skłodowska University, Lublin, Poland. ⁵ECOTECH-Complex, Maria Curie-Skłodowska University, Lublin, Poland. ⁶Department of Hydrology, Institute of Geography, University of Gdansk, Gdansk, Poland. ⁷Institute of Geological Sciences Polish Academy of Sciences, Warsaw, Poland. ⁸Department of Biochemistry and Molecular Biology and the Huck Institutes of Life Sciences, Pennsylvania State University, University Park, PA, USA. ⁹These authors contributed equally: Olga Iwańska and Przemysław Latoch. ✉email: agata.starosta@gmail.com

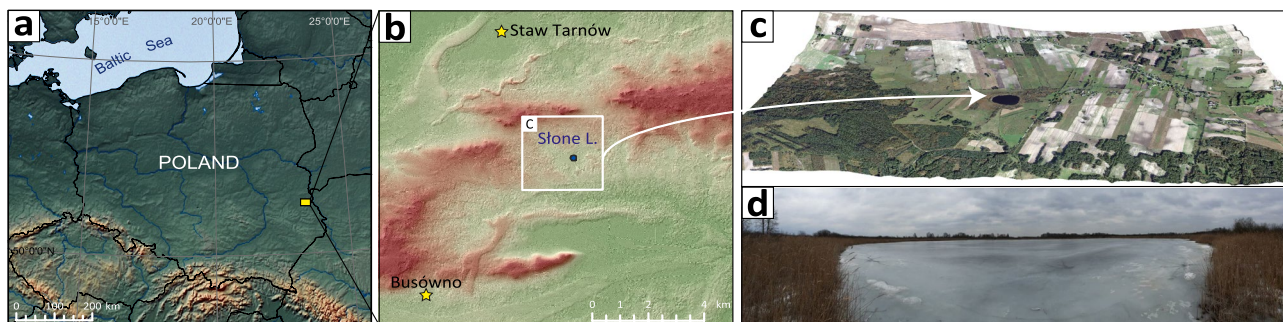


Figure 1. Localization of Lake Slone. (a) Lake Slone (51°18'16"N, 23°21'55"E) is located in SE Poland within the Chełm Hills Region, a part of Volhyn Polesie¹⁸. (b) Topography of the northern part of Chełm Hills Region showing the locations of the closest Early Medieval hillforts—Busówno and Staw Tarnów marked with yellow stars—in the vicinity of Lake Slone. In the Middle Ages, at least four settlement units functioned within the 10 km radius from the study site, with Busówno hillfort being the closest and the largest among them¹⁹. The results of geoarchaeological studies of the medieval settlements in the region have been well-described²⁰ however, earlier anthropogenic activity in SE Poland dating back to the Roman period and the Migration Period has not been fully characterised. During the Roman period, the Przeworsk and Wielbark cultures can be located to the region of SE Poland, which then migrated towards the south and west and were replaced by the first Slavs around the sixth century²¹. The first records of the settlement in Busówno date back to the seventh century and the archaeological evidence suggests that the hillfort functioned until the fourteenth century and was still used, but to a lesser extent, until the seventeenth century¹⁹. (c) Nowadays, Lake Slone is a small, shallow (3.4 ha, 8.0 m), eutrophic water body (Carlson Trophic State Index = 44–58) of high ecological status¹⁷. The lake has a catchment area of 5.5 km² with a predominant agricultural land use (arable lands, meadows and wetlands), whereas build-up areas within the catchment do not exceed 2%. (d) Ice-covered Lake Slone during sediment sampling, January 2019.

including polysaccharides dissolve, ensuring nutrients for the developing retting microorganisms which are critical to the process¹². The microbiome of a hemp rettery depends on the levels of dissolved oxygen, type of degraded biomass^{13,14}, and also on the geochemical and physical factors which influence the initial reservoir flora¹⁵.

Our previous investigations of Lake Slone (LS) (Fig. 1a), located in SE Poland, revealed that it was used as a hemp rettery in the early Middle Ages¹⁶, most probably by the nearby settlers (Fig. 1b). The relatively large catchment, together with the small lake area and volume results in a high eutrophication risk¹⁷ and as a consequence, LS reacts readily to the environmental changes in its catchment (Fig. 1c,d). The recurrent nature of HR events during the last 2000 years of LS history allowed us to investigate the long-term response of the lake microbiome to the changing anthropopressure, especially in the context of anthropogenically induced trophic fluctuations.

In this highly interdisciplinary study, we used multi-proxy paleolimnological data and combined it with the metagenome shotgun sequencing of the 1 m long LS sediment core collected in 2019 (LS-C19) to identify the historical periods of HR in the lake and examine their effects on lake trophic and bacterial microbiome fluctuations, in the context of environmental changes. Our study shed light onto the microbiome of a hemp rettery versus a recovered lake as a consequence of varying anthropopressure during the preindustrial period, by providing a detailed analysis of the microbial taxonomy and functionality based on the *sedDNA* analysis.

Results

Paleo-environmental analysis. The LS-C19 core comprises gyttja type of sediments with varied colouring and a distinguishable 5.5 cm thick laminated layer (62.5–68 cm from the sediment surface), containing macroscopically visible plant fibres, which were shown to be hemp fibres—remains of HR (Fig. 2).

We used Cladocera and diatoms together with the loss on ignition analyses to investigate lake trophic, and palynological analysis to evaluate the intensity of anthropogenic activity. This, together with the composite age-depth model (Supp Figs. 1–4; Supp Tables 1, 2), allowed us to distinguish six phases in the lake development, correlated to the archaeological periods: (I) the Roman Period; (II) the Migration Period; (III) the Early Middle Ages; (IV) Settlement Relocation; (V) the Early Modern Period, and (VI) the Contemporary Times (Fig. 3). The Contemporary Time period was not a part of this study as it falls into the Great Acceleration period²³.

During the Roman Period (93–80 cm, I–III CE) LS was relatively deep and alkaline, with a well-developed open-water zone (Fig. 3a,c), as shown by the presence of planktonic taxa of Cladocera and diatoms (*Lindavia* and *Cyclotella*), and dominance of alkaliphilous (*Lindavia ocellata* and *L. radiosa*), alkalibionthic (*Stephanodiscus parvus*) and neutrophilic (*Lindavia comensis* and *Cyclotella cyclopuncta*) groups of diatoms (Supp Figs. 3, 4; Supp Tables 1, 2). The paleoecological proxies imply the oligo-/mesotrophy level, however, at the depth 86–81 cm a slight elevation of the trophic was noticed, recorded as the increase and composition change of Cladocera (Fig. 3b, Supp Fig. 3, Supp Table 1). The palynological analysis revealed a minor anthropogenic influence involving deforestation (altered ratio of the arboreal pollen, AP, to non-arboreal pollen, NAP), presence of segetal weeds and agricultural plants, and sporadic occurrence of wheat (*Triticum t.*) and cereals (*Cerealia undiff.*) (Fig. 3f,h,i). In samples LS-85 and LS-80, corresponding to depths of 85 and 80 cm, we identified hemp pollen—0.69% (1.32%

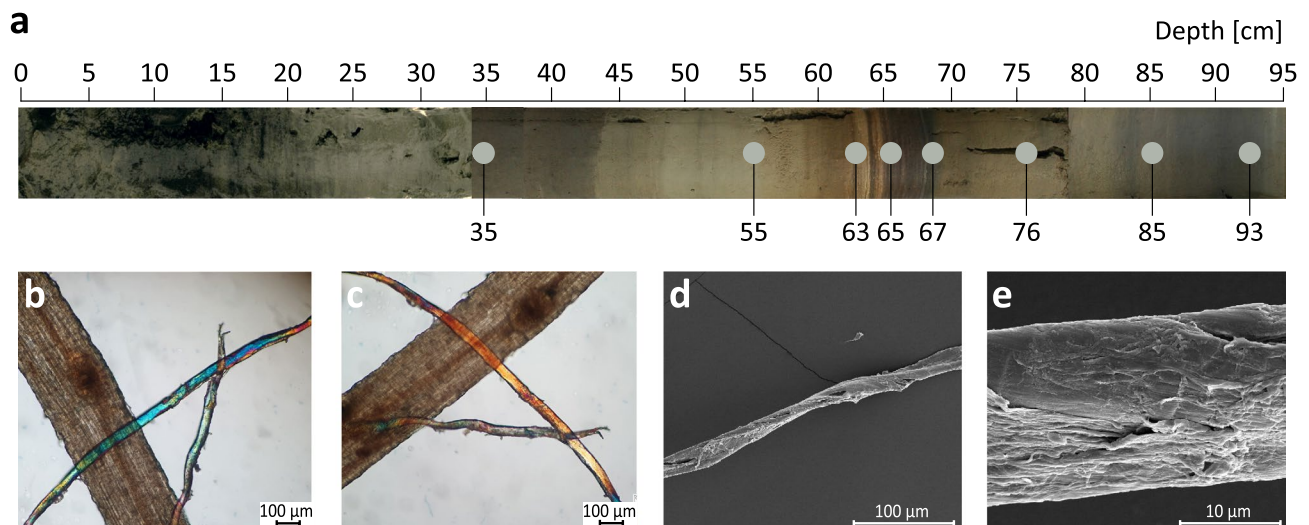


Figure 2. Lake Slone sediment core and hemp fibres. (a) Photo of the LS-0 sediment core (composite core sampled parallel to LS-C19 core with Instorf-type of sediment sampler), exhibiting distinct lithological changes of the sediment with an approximate location of the layers selected for metagenomic DNA isolation (material taken from LS-C19 core). (b,c) Polarised light microscopy images of hemp fibres extracted from a layer at 65 cm depth of the LS-C19 core. Results of the modified Herzog test²² to identify hemp bast fibres show the characteristic colour change from blue to yellow upon sample stage rotation by 90° which would be expected of hemp fibres, (d) Scanning Electron Microscopy (SEM) image of a hemp bast fibre collected from the layer at 65 cm depth from LS-C19 at the magnification $\times 500$, (e) SEM image of a hemp bast fibre at the magnification $\times 5000$.

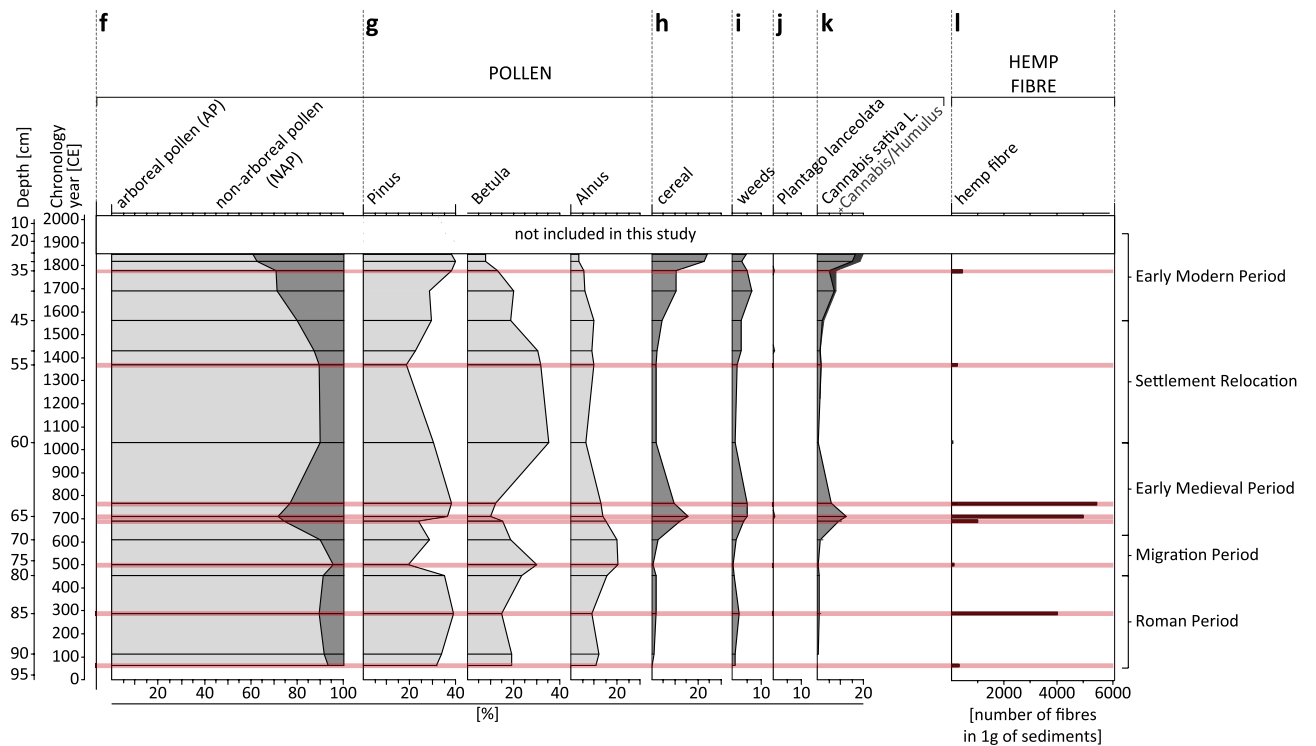
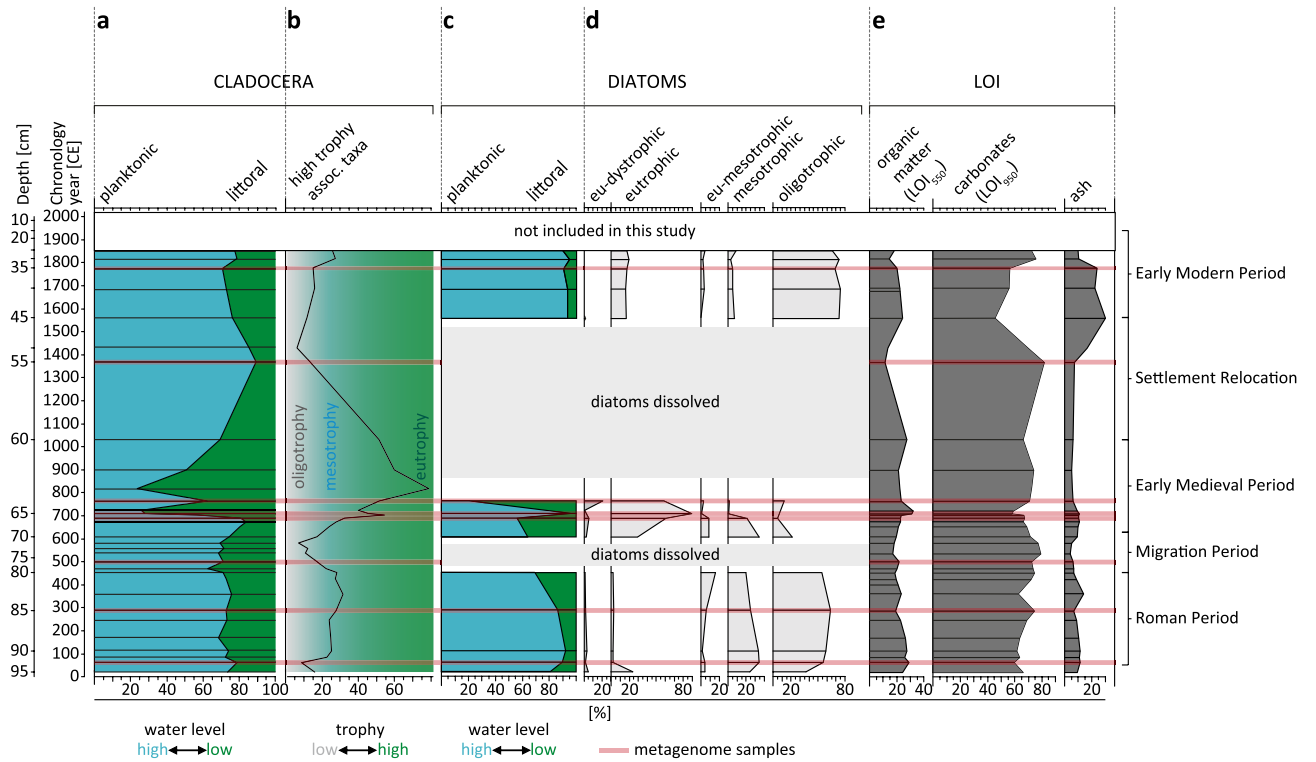
including *Cannabis/Humulus*) and 0.66% (0.86%) of the total pollen sum respectively—together with hemp fibres, suggesting LS was occasionally used as a hemp rettery as early as in the Roman Period (Fig. 3k,l).

The Migration Period (80–69 cm, IV–VII CE) was characterised by the relieved anthropogenic influences and forest regeneration (Fig. 3f,g), accompanied by a decrease in depth and lake trophy. The increase in benthic diatoms points to an extension of the euphotic zone (Fig. 3d). The improvement of benthic oxygen availability is evidenced by an increase in the abundance of Cladocera benthic taxa and a decrease in the littoral species (Fig. 3a,b; Supp Figs. 3, 4; Supp Tables 1, 2). A decrease in the NAP values including all cereals and *C. sativa*, and an increase in the AP values, further indicate a diminished anthropogenic influence (Fig. 3f,h,i,k). The fibre count in layers corresponding to this period was very low suggesting no HR practices (Fig. 3l).

The Early Middle Ages (69–60 cm, VII–XI CE) were the most distinguishable in terms of the ecological changes phase of the lake development with the corresponding layers covering the lamination zone (62.5–68.0 cm) (Fig. 2). The planktonic Cladocera and diatoms, dominant in the layers 69–67 cm, were gradually being replaced with the benthic species indicative of trophy and pH rise, also shown by the dominance of alkaliphilous (*Ulnaria contracta*, *Stephanodiscus medius*, *Gomphonema truncatum* and *G. olivaceum*) and alkalibiontic (*Stephanodiscus parvus*) diatoms (Supp Fig. 4; Supp Table 2). A drastic decrease in the Cladocera planktonic taxa proportion in the layers 66–63 cm suggests lake shallowing, whereas the increase of *Alona guttata*, *Alonella excisa* and *Kurzia laticornis* points to an elevated supply of humic substances (Fig. 3b). In the top layers (62–60 cm) the lake trophy was still high, however, HR activity probably ceased, as suggested by the palynological data and very low fibre count (50 fibres/g of sediment) (Fig. 3b,l; Supp Figs. 3, 4; Supp Tables 1, 2). The high anthropogenic activity in the area during the Early Middle Ages is evidenced by high (11.5–30.2%) NAP values indicative of arable fields, pastures and a significant degree of deforestation, as well as intensive agriculture of cereals (*Secale cereale* and *Triticum t.*) and hemp (Fig. 3f,h,i,j,k). The fibre count in layers 63 and 65 cm was the highest (up to 5500 fibres/g) suggesting the most intensive HR practices in LS in the Early Middle Ages (Fig. 3l).

During the Settlement Relocation (60–45 cm, XII–XVII CE), the NAP values consisting of both segetal weeds and crops decreased significantly (Fig. 3h,i). Such decrease also included hemp pollen (Fig. 3k). Together with the low values of fibre count, these reflect cessation of *Cannabis* cultivation and retting in the area (Fig. 3l) which strongly suggests settlement relocation, as supported by the archeological literature¹⁹. The absence of anthropopressure resulted in a clear oligotrophication and an increase in the water level indicated by the Cladocera-based inference (Fig. 3a,b). At 60–55 cm diatoms frustules were dissolved, however, after their reappearance at 45 cm, diatom composition confirmed the interpretation of Cladocera data suggesting oligotrophy and increased water level (Fig. 3c,d). Moreover, diatom composition at 45 cm points to a drop in the pH value to ≤ 7 (Supp Fig. 4).

The Early Modern Period (45–25 cm, XVIII–XIX CE) of the LS history is characterised by the returning anthropopressure evidenced by deforestation (decrease in AP values) and the increasing NAP values (crops, segetal and ruderal weeds) (Fig. 3f,h,i,j). Interestingly, we found high levels of hemp pollen, however, low fibre count suggests hemp cultivation but no HR in the lake (Fig. 3k,l). This is reflected in the lake trophy which remained at the oligo-mesotrophy with the elevated water level, as evidenced by the higher abundance of planktonic Cladocera and diatoms, compared to the benthic taxa (Fig. 3b,d).



◀ **Figure 3.** Paleo-environmental indicators. The age-depth correlation of the LS-C19 with paleo-environmental indicators: (a) Cladocera-based water level inference based on the proportion of planktonic to littoral taxa (Supp Table 1, Supp Fig. 3) in LS-C19 sediments; (b) Cladocera-based trophic inference based on the proportion of high trophic indicator taxa in Chydoridae family (Supp Table 1, Supp Fig. 3); (c) Diatom-based water level inference based on the proportion of planktonic to littoral taxa in LS-C19 sediments (Supp Table 2, Supp Fig. 4); (d) Diatom-based trophic inference shown as proportion of diatom taxa indicative to oligo-, meso-, eu- and oligotrophic status of water (Supp Table 2, Supp Fig. 4); (e) proportion of the sediment components based on loss on ignition (LOI) measurements at different temperatures—organic matter (550 °C), carbonate (950 °C) and ash (non-carbonate mineral residue) in LS-C19 sediments. Sediment composition of LS-C19 is dominated by the carbonate component (46–85%) with organic matter ranging from 11 to 32%, whereas the non-carbonate mineral matter ranges from 3 to 30% (Supp Data 1); (f) proportion of arboreal pollen (tree sum) to non-arboreal plants pollen, indicator of the landscape openness; (g) selected tree pollen curves; (h) sum of the cereal pollen (includes the pollen of *Secale cereale*, *Triticum* sp., *Linum usitatissimum*, *Cerealia* undiff., *Fagopyrum* sp., and *Cannabis sativa* L.); (i) sum of the weeds pollen (includes segetal and ruderal weeds *Centaurea cyanus*, *Convolvulus arvensis*, *Artemisia* sp., *Chenopodiaceae*, *Urtica*, *Brassicaceae*, *Rumex acetosella* t., *Spergula arvensis*, *Scleranthus annuus*); (j) *Plantago lanceolata* pollen sum (indicator of pastoral activity); (k) *Cannabis sativa* L. pollen and *Cannabis sativa*/*Humulus* type pollen sum (when unequivocal distinction of the grain type was impossible); (l) fibre count (absolute number) per 1 g of sediments, with threshold level for hemp retting practices (400 fibres/1 g); layers which *sedDNA* was extracted from are marked in red.

Shotgun metagenomics. Based on the paleo-environmental analysis, we selected eight layers for shotgun NGS. Samples were prepared in duplicates and originated from depths of 93 and 85 cm (Roman period), 76 cm (Migration period), 67, 65 and 63 cm (Early Middle Ages), 55 cm (Settlement Relocation) and 35 cm (Early Modern period). Samples nomenclature corresponds to the lake's name (LS), extraction depth (as listed above) and duplicate number (1 or 2).

Taxonomic diversity. The taxonomical classification of the 16 samples resulted in 70% of the sequences assigned to taxonomic ranks (Supp Fig. 5a). The most dominant group was Bacteria (77–84%), followed by Archaea (13–20%), Eukaryota (0.5–3%) and Viruses (1–2%) (Supp Fig. 5b). The most abundant bacterial phyla in all samples were *Proteobacteria* (26.4–38.4%) and *Chloroflexi* (23.5–33.9%), followed by *Planctomycetes* (8.8–16.5%), *Bacteroidetes* (2.9–7.0%), *Actinobacteria* (3.5–8.7%), *Acidobacteria* (3.8–5.2%), *Spirochaetes* (1.6–3.4%), *Firmicutes* (1.6–2.2%) and *Verrucomicrobia* (0.6–1.4%) which in total account for 94.2–97.0% of the total abundance in each sample (Fig. 4a). The remaining ten phyla range between 0.0–1.0% per sample and account for the total of 3.0–5.8% abundance of bacterial communities.

Based on the hemp retting activity in LS we categorized the samples into the 'no hemp retting' group ('nHR' including 8 samples: LS-93-1/2, LS-76-1/2, LS-55-1/2, and LS-35-1/2) and 'hemp retting' group ('HR' including 8 samples: LS-85-1/2, LS-67-1/2, LS-65-1/2, LS-63-1/2) (Fig. 4b). The PCA revealed a clear grouping of the two groups: samples LS-65-1/2 and LS-63-1/2 (intensive lake usage as a retery) clustered together and opposite to the LS-76-1/2 and LS-55-1/2 (settlement emptiness). LS-85-1/2 and LS-67-1/2 (moderate levels of HR) clustered together and in the middle section of the PCA plot and could thus represent the transitional state of the lake microbiome. Interestingly, LS-93-1/2 and LS-35-1/2—no HR but known anthropogenic presence—were also located in the middle section of the PCA plot. This may suggest shifts in the lake microbiome depending on the degree of anthropopressure. Likewise, in the UPGMA tree the samples originating from periods of no anthropogenic influence (LS-76-1/2, LS-55-1/2) clustered together and distant to the remaining layers. Samples corresponding to the intensive anthropogenic impact (LS-65-1/2, LS-63-1/2) were grouped together however, constituted a part of a more complex cluster including LS-35-1/2, LS-67-1/2, LS-85-1/2 and LS-93-1/2 (Fig. 4c). PERMANOVA analysis indeed showed a difference between the two groups of microbial communities ('HR' vs 'nHR', $F = 6.2209$, $p < 0.01$).

We further examined the correlation between the microbial communities and the investigated environmental factors (Supp Table 3). Four factors affected the microbiome the most and accounted for 77% of community dissimilarities. These included fibre count, forestation, carbonate content and trophic (p < 0.01) (Fig. 4c). The first two describe the anthropogenic impact on the lake microbiome—fibre count serves as a proxy for the intensity of HR whereas forestation (AP values) describes forest regeneration and the absence of human activities. Carbonate content and lake trophic independently describe physico-chemical characteristics including pH, oxygen concentration of the benthic zone and amount of organic carbon.

The taxonomic composition of the microbial communities varied notably depending on the usage of LS as a hemp retery. We found significant differences in the abundance of five bacterial phyla (Fig. 4d). In the 'HR' group, three phyla were more abundant: *Chloroflexi* with the average abundance at $30.5 \pm 1.5\%$ ($p < 0.01$) and a difference in mean proportions (DM) between the two groups at -3.9% , followed by *Planctomycetes* ($13.8 \pm 1.6\%$, $p < 0.05$, DM -2.4%) and *Bacteroidetes* ($5.5 \pm 1.2\%$, $p < 0.05$, DM -1.4%) (Supp Data 3a). In the 'nHR' group we identified two significantly overrepresented bacterial phyla, including *Proteobacteria* ($38.0 \pm 2.0\%$, $p < 0.01$, DM 5.6%) and *Actinobacteria* ($6.4 \pm 1.5\%$, $p < 0.05$, DM 2.0%). To further investigate the taxonomic changes, we identified significant differences in the relative abundances in 8 classes/orders in the 'nHR' group and 6 classes/orders in the 'HR' group (Supp Fig. 6). DM values were small (DM < 1%) for all groups, except *Dehalococcoidia* (DM -1.05% ; Supp Data 3b), demonstrating that bacterial community changes were subtle yet consistent. This suggests fluency in the process of adaptation to the new growth conditions rather than radical changes to the community composition.

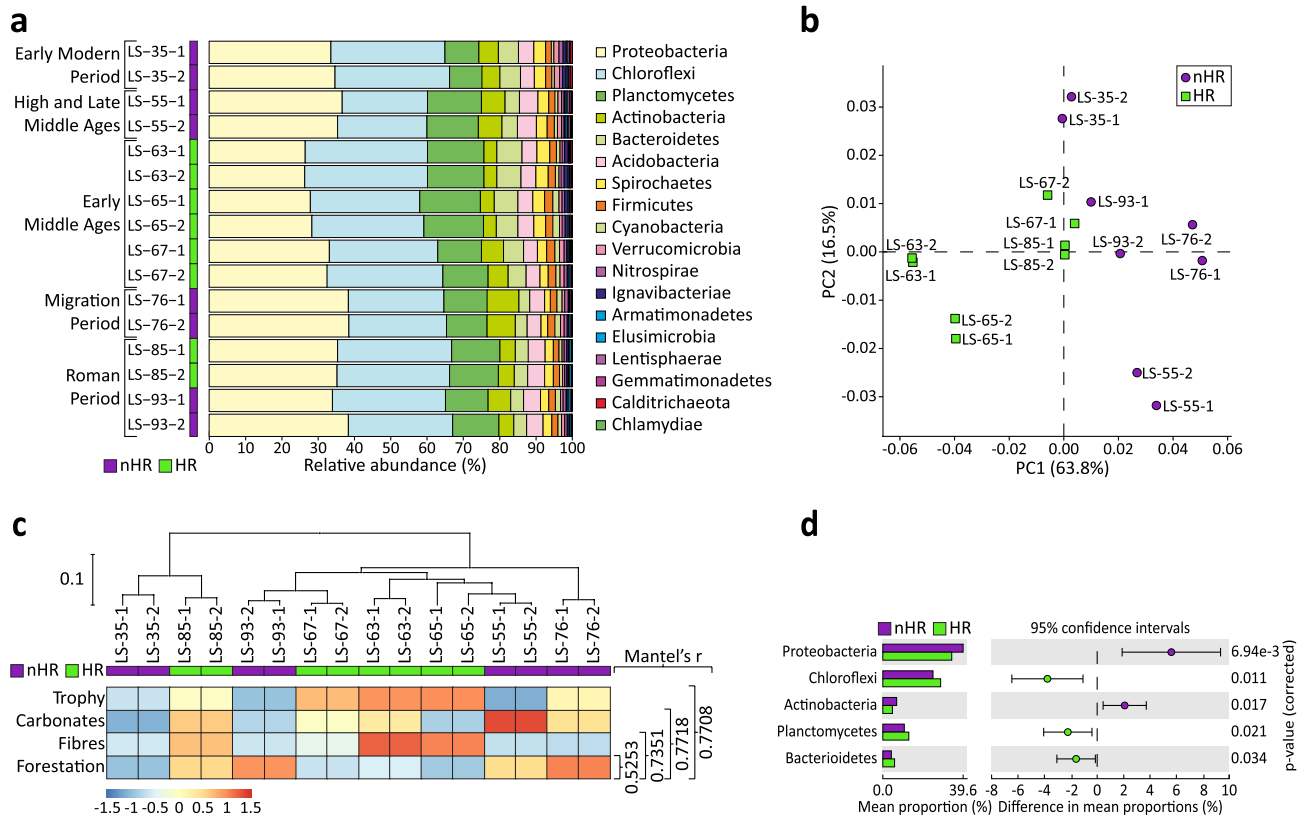


Figure 4. Taxonomic analysis of the microbiomes in the hemp retting ('HR') vs no hemp retting ('nHR') groups. **(a)** Taxonomical classification of the bacterial communities from Lake Slone. Relative abundances are presented at the phylum level. Samples from the periods of hemp retting ('HR') are coded in green, samples from the no hemp retting ('nHR') group are in purple. **(b)** Principal Component Analysis plot of the taxonomical structures of all samples in two groups—'HR' (green) vs 'nHR' (purple); **(c)** Hierarchical clustering (unweighted pair group method with arithmetic mean, UPGMA) based on taxonomy Bray–Curtis dissimilarity matrix (MEGAN). Samples from the two groups are colour coded—the 'HR' group in green and 'nHR' in purple. Heat map of the normalised values of the selected environmental factors showing the highest correlation to the taxonomic composition of the microbial communities, calculated by Mantel test (bioenv and mantel functions in vegan package), values of r -statistic indicated on the right, $p < 0.01$; **(d)** Extended error bar plot showing significant differences ($p < 0.05$) in the taxa abundances at the phylum level for 'HR' (green) vs 'nHR' (purple) groups, calculated with two-sided Welch's t test (STAMP).

Functional analysis. Data assembly and coding sequence prediction were followed by annotation and KEGG Orthology (KO) assignment (Supp Data 2b,c). The PCA performed on all KOs showed a similar ordination of the functional profiles compared to that of the taxonomical groups (Fig. 5a). Samples corresponding to the extensive HR activity (LS-63-1/2, LS-65-1/2) clustered together and distal to the centre of the PCA plot, and LS-67-1/2 (beginning of HR in the Early Middle Ages) clustered close by. Samples LS-85-1/2 (Roman episode of HR) were positioned closer to the middle part of the PCA plot. In the 'nHR' group, LS-35-1/2, LS-55-1/2 and LS-93-1/2 clustered together, and samples LS-76-1/2 (Migration Period) further away, closer to the middle section of the PCA plot. Interestingly, LS-35-1/2 and LS-55-1/2 clustered together on the functional PCA plot, in contrast to the taxonomy PCA plot, showing a degree of functional redundancy. A similar grouping was revealed by the UPGMA tree (Fig. 5b). Here, the functional structures of the microbiomes from LS-67, LS-65 and LS-63 (most intense HR) were grouped together and separate to the other samples. LS-76-1/2, LS-55-1/2, LS-35-1/2 clustered together and samples corresponding to the Roman Period (LS-93-1/2, LS-85-1/2) created a separate group, however, situated in a more complex cluster including samples from the 'nHR' group. PERMANOVA analysis of the functional profiles including all KOs revealed a statistically significant difference between the two groups ('HR' vs 'nHR', $F = 6.0242$, $p < 0.01$) and the functional and taxonomic compositions showed 76% relatedness (Mantel test, $R^2 = 0.7632$, $p < 0.01$). The functional diversity of all samples (all identified KOs) was best described by four independent environmental factors that together contributed to 78% data similarity ($p < 0.01$). These included trophic, agriculture of cereals, fibre count and organic carbon (Fig. 5b). The cereal agriculture and fibre count describe the anthropogenic impact on the lake, both direct (HR) and indirect (cereal cultivation in the lake vicinity). Estimation of the organic carbon content (LOI_{550}) provides information on both autochthonous and allochthonous carbon that has not been mineralized during sedimentation. Therefore, it is a proxy of lake productivity, catchment supply and mineralization conditions²⁴.

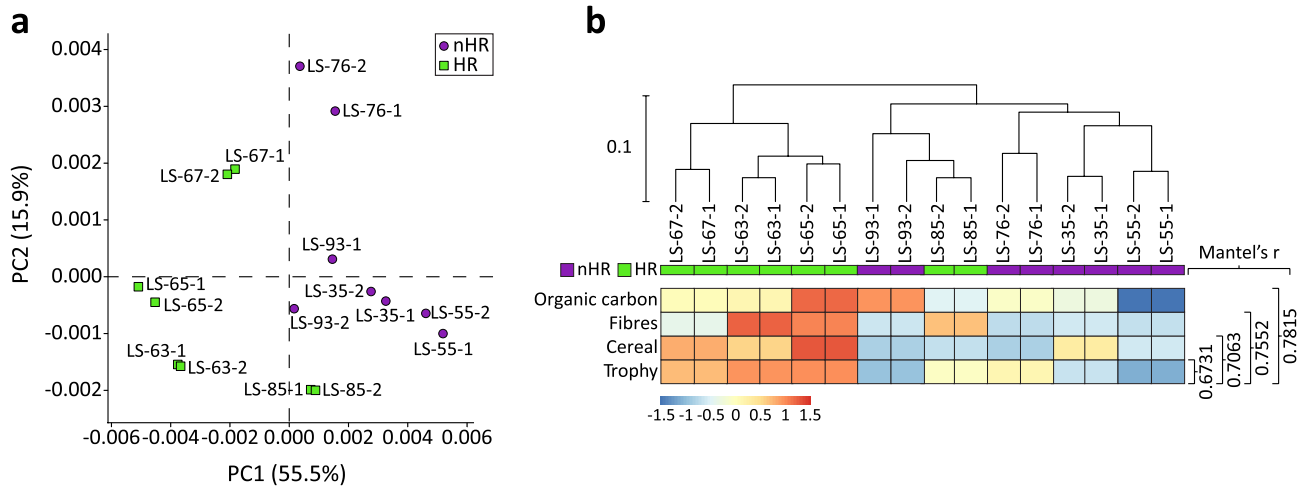


Figure 5. Functional analysis of the microbiomes in the hemp retting ('HR') vs no hemp retting ('nHR') groups. **(a)** Principal Component Analysis plot based on complete functional profiles of the samples in two groups—'HR' (green) vs 'nHR' (purple); **(b)** Hierarchical clustering (UPGMA) based on Bray–Curtis dissimilarity matrix of the complete functional profiles of all samples in the 'HR' (green) vs 'nHR' (purple) groups. Heat map of the normalised values of the selected environmental factors showing best correlation with the functional profiles, calculated by Mantel test (bioenv and mantel functions in vegan package), values of r-statistic indicated on the right, $p < 0.01$.

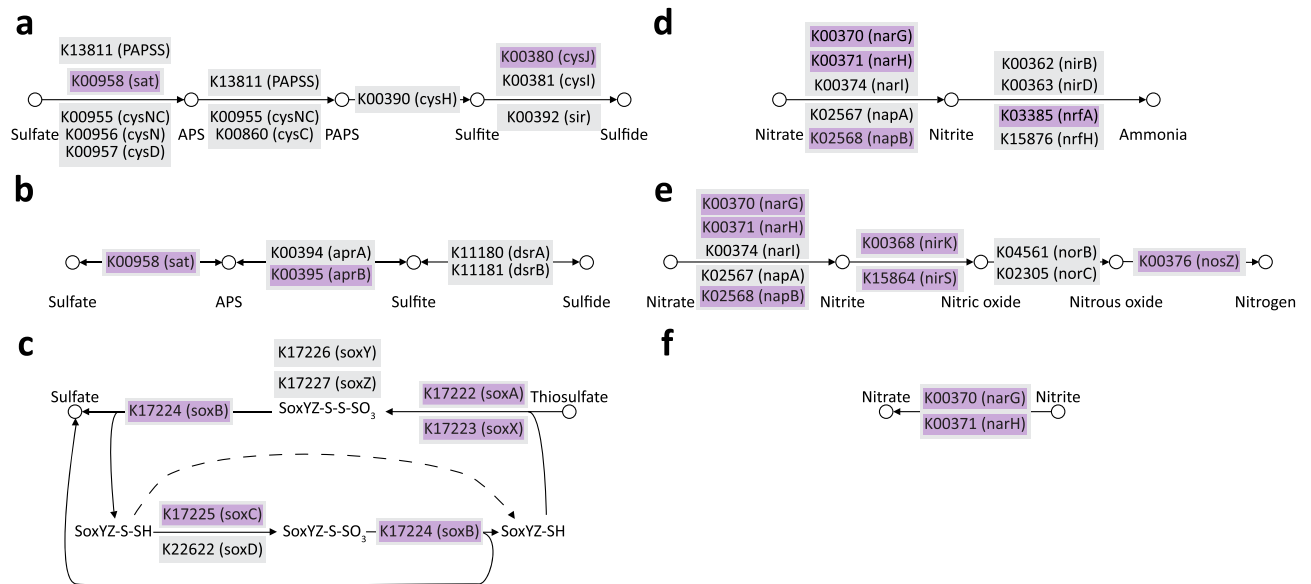


Figure 6. Schematics of the selected sulfur and nitrogen KEGG metabolic pathways. Pathways were inferred using the MinPath tool based on the overrepresented KEGG Orthologs (KOs), which were identified in the two-sided Welch's t test ($p < 0.05$) in STAMP. KOs significantly overrepresented ($p < 0.05$) in the no hemp retting ('nHR') group are shown in purple boxes. KOs included in the KEGG pathway, but not significantly overrepresented ($p > 0.05$) are shown in grey boxes. **(a)** Assimilatory sulfate reduction, **(b)** Dissimilatory sulfate reduction and oxidation, **(c)** SOX system, **(d)** Dissimilatory nitrate reduction, **(e)** Denitrification, **(f)** Nitrification.

Metabolic potential of the sediment microbiome. In order to determine differences in bacterial metabolism, we reconstructed the KEGG pathways from differentially abundant KOs using the MinPath tool²⁵. 17.6% of all KOs were found in different proportions between the two groups ($p < 0.05$) (Supp Data 4), which resulted in 67 differentially abundant KEGG_{level3} pathways in the 'HR' group, and 91 pathways in the 'nHR' group (Supp Data 5). PERMANOVA analysis of the KOs classified to metabolism showed a moderate difference between the two groups ('HR' vs 'nHR', $F = 5.7496$, $p < 0.01$). Using the pathway reconstruction analysis, we further investigated nitrogen, sulphur and carbon metabolic pathways. We also calculated the genetic potential of the microbial communities towards different types of carbon metabolism using selected marker genes²⁶ (Supp Table 4).

Sulfur metabolism was more pronounced in the 'nHR' group. Both dissimilatory and assimilatory sulfate reduction pathways were partially reconstructed, together with an almost complete sulfur oxidation pathway

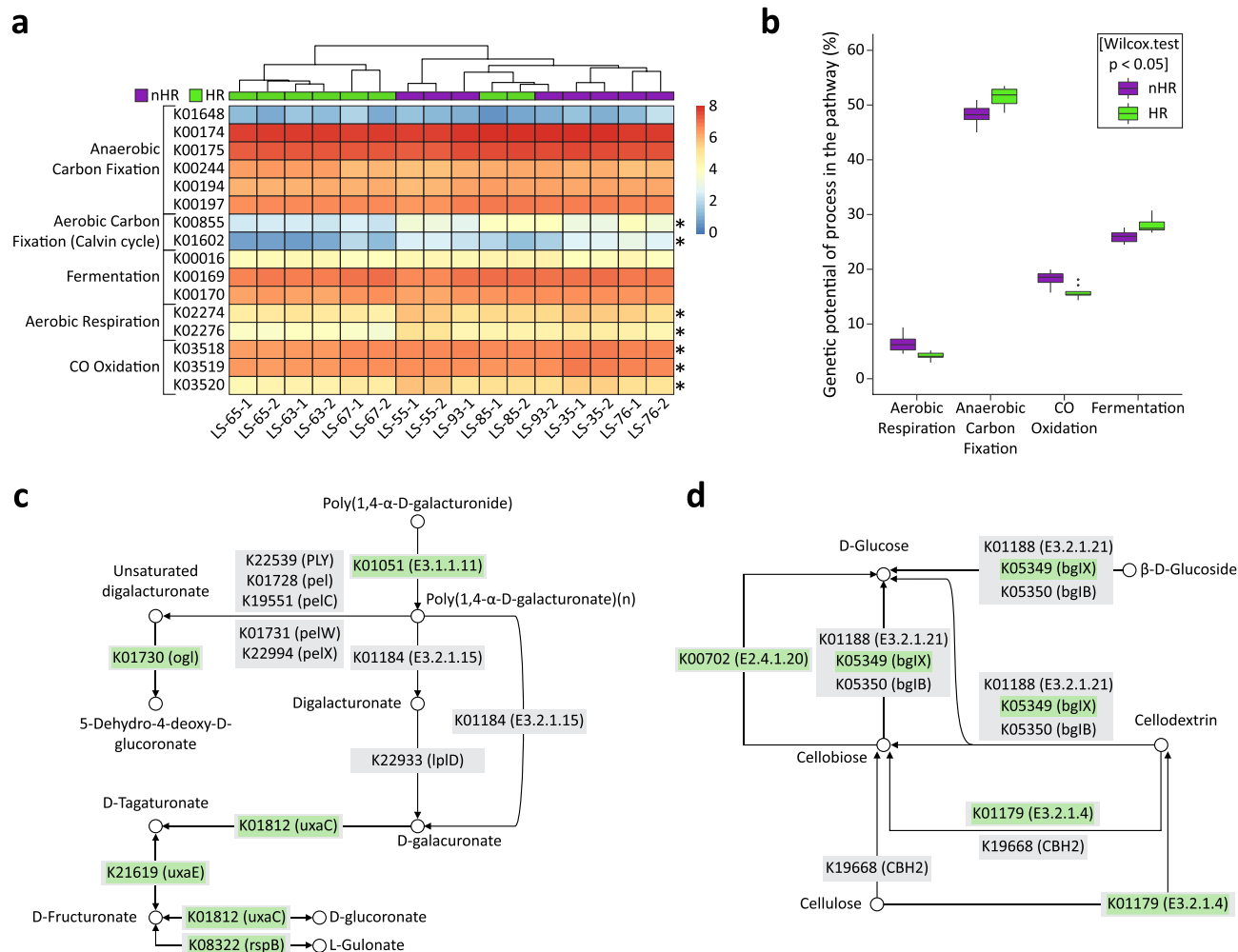


Figure 7. Carbon metabolism potential of the microbial communities in the ‘HR’ vs ‘nHR’ groups. **(a)** Heat map representing functional clustering of the selected KOs involved in different pathways of the carbon cycle²⁶ (Supp Table 4) based on normalised relative abundances ($\log(\text{TPM}+1)$) for all samples in the ‘HR’ (green) and ‘nHR’ (purple) groups. Differentially abundant KOs are marked with asterisks (Boruta, $\text{maxRuns} = 9999$, $\text{mean importance} > 4$). Hierarchical clustering (UPGMA) is based on the Bray–Curtis dissimilarity matrix for the selected KOs. **(b)** Box-and-whiskers plot of carbon metabolism potential calculated according to Lauro et al.²⁶ based on relative abundances of the selected KOs in the ‘HR’ group (green) and ‘nHR’ group (purple): anaerobic carbon fixation— $(\text{K01648} + \text{K00174} + \text{K00175} + \text{K00244} + \text{K00194} + \text{K00197})/6$, aerobic carbon fixation— $(\text{K00855} + \text{K01602})/2$, fermentation— $(\text{K00016} + \text{K00169} + \text{K00170})/3$, aerobic respiration— $(\text{K02274} + \text{K02276})/2$, CO oxidation— $(\text{K03518} + \text{K03519} + \text{K03520})/3$. Significance of differences between group potentials tested with Wilcoxon test ($p < 0.05$). **(c)** Schematic of the inferred pectin hydrolysis pathway (fragment of KEGG 00040 Pentose and glucuronate interconversions pathway) based on the overrepresented KOs identified in two-sided Welch’s t test ($p < 0.05$). KOs overrepresented in the ‘HR’ group are shown in green boxes, KOs involved in the pathways, but not significantly overrepresented ($p > 0.05$) are shown in grey boxes. **(d)** Schematic of the inferred cellulose hydrolysis pathway (fragment of KEGG 00500 Starch and sucrose metabolism pathway) based on the overrepresented KOs identified in two-sided Welch’s t test ($p < 0.05$). KOs overexpressed in the ‘HR’ group are shown in green boxes. KOs involved in the pathways, but not significantly overrepresented ($p > 0.05$) are shown in grey boxes.

(Fig. 6a–c). No sulfur metabolism pathway was reconstructed based on the overrepresented reads originating from the ‘HR’ group. The inferred sulfur oxidation pathway contained reads classified mainly to Betaproteobacteria and Actinomycetes taxa ($p < 0.05$), while Deltaproteobacteria, Gammaproteobacteria and Betaproteobacteria were mostly responsible for sulfate reduction ($p < 0.05$). Likewise, reads classified to KOs describing nitrogen metabolism were more abundant in the ‘nHR’ group. Both dissimilatory nitrogen reduction and denitrification pathways were almost completely reconstructed. This was due to the overrepresentation of nitrate and nitrite reductases together with nitrous oxide reductase in the ‘nHR’ group. The nitrification pathway was partially reconstructed due to the higher abundance of nitrite reductases in the ‘nHR’ group (Fig. 6d–f). These were shown to be carried out mostly by Betaproteobacteria, Deltaproteobacteria and Gammaproteobacteria ($p < 0.05$).

The most differentially abundant metabolic pathways in the 'HR' group were classified to carbon metabolism. Based on selected marker genes²⁶ (Supp Table 4) we constructed the UPGMA tree which showed two distinct clusters. The first cluster included samples corresponding to the most intense HR activity in LS (LS-67-1/2, LS-65-1/2 and LS-63-1/2) and the second included samples from the 'nHR' group, as well as LS-85-1/2. We also identified marker genes which were differentially abundant between the two groups ($p < 0.05$) (Fig. 7a). Based on the calculated metabolic potential²⁶, the most prevalent type of carbon metabolism in both groups was anaerobic carbon fixation, whereas aerobic respiration showed the lowest metabolic potential. Still, anaerobic carbon metabolism including fermentation and anaerobic carbon fixation was more pronounced in the 'HR' group (Wilcoxon test, $p < 0.05$) (Fig. 7b). Indeed, upon pathway reconstruction analysis we found the reductive Acetyl-CoA pathway and reductive citric acid cycle (TCA) to be partially inferred based on the overrepresented KOs in the 'HR' group. In the 'nHR' group the microbial community showed higher potential for CO oxidation and aerobic respiration (Fig. 7b). The marker genes for CO oxidation are not included in KEGG pathways which rendered pathway reconstruction analysis unfeasible. However, upon pathway reconstruction for aerobic respiration, several bacterial cytochrome complexes were shown to be highly abundant in the 'nHR' group. As the main source of organic carbon during HR episodes originated from hemp stems, we further investigated complex plant polysaccharides degradation pathways. Pectin and cellulose hydrolysis pathways were nearly completely reconstructed based on the differentially abundant reads (Fig. 7c,d). Pectin hydrolysing enzymes including pectin esterases (K01051) and pectate lyases (K01728, K19551 and K01731) were more abundant in the 'HR' group, however, for the latter the difference was not statistically significant ($p < 0.1$) (Fig. 7c). Interestingly, the number of reads mapped to endo-polygalacturonases (K01184), another important pectin-hydrolysing enzyme, was rather low and very similar in both groups. Also, no reads were classified to exo-polygalacturonases (K22933). This may suggest that hydrolysis of 1,4 glycosidic bonds between galacturonic acid residues could have been carried out by fungal enzymes and bacteria employed an alternative mechanism of pectin breakdown—eliminative cleavage of 1,4- α -D-galacturonan. The overrepresented pectin hydrolysing enzymes in the 'HR' group were classified mostly to Planctomycetes, Bacteroidetes, Chloroflexi and Clostridia. The remaining enzymes involved in pectin hydrolysis as well as the related conversions of galacturonate and glucuronate including glucuronate isomerase (K01812), tagaturonate epimerase (K21619), L-gulonate 5-dehydrogenase (K08322) and oligogalacturonide lyase (K01730) were also significantly more abundant in the 'HR' group (Fig. 7c). Although during HR cellulose hydrolysis is an unwanted by-product of microbial activity, a vast number of microorganisms are capable of cellulose degradation. We inferred a complete cellulose hydrolysis pathway based on the overrepresented reads from the 'HR' group (Fig. 7d), including endoglucanases (K01179), β -glucosidases (K05349) and cellobiose phosphorylases (K00702). Interestingly, as was the case for pectin hydrolysis, the number of reads assigned to exoglucanases (K19668) was relatively low and similar in both groups. Again, this may suggest participation of fungal enzymes in the cellulolytic processes. Based on the taxonomic classification of the overrepresented reads, cellulose degradation in the 'HR' group was carried out predominantly by Chloroflexi, Planctomycetes, Bacteroidetes, Acidobacteria, and to a smaller extent by Clostridia.

Discussion

Based on the paleo-environmental data we identified increased anthropogenic activity in the lake vicinity during the late stage of the Roman Period, Early Middle Ages and Early Modern Period, and a markedly decreased anthropogenic influence during the Migration Period and High and Late Middle Ages. The cannabis type pollen records aligned with other palynological indicators of anthropopressure, however, did not correspond fully with the fibre count data. Typically, the determination of hemp water-retting has been based on the pollen records^{10,27,28}. However, the use of cannabis type pollen percentages as determinants of HR has been discussed and different percentage cut-offs have been suggested^{11,28}. Additional proxies for confirming water-retting of hemp have been proposed, including lithological changes, diatoms or algae^{29,30}. The presence of phytophilic Chironomid taxa as potential indicators of HR in stagnant water was also discussed¹¹. Hemp fibres were previously identified in the laminated lake sediments originating from the seasonal changes in biomass production and mineralisation and were reported as remains of hemp soaking³¹. Since lamination layers containing fibrous material were present in LS-C19, we employed fibre count as a proxy for HR and an estimate of the process intensity. Application of such approach revealed discrepancies between the hemp pollen and fibre count—we identified relatively low levels of *Cannabis* type pollen (1.3%) and high values of fibre count (4000 fibres/g) in the samples corresponding to the late Roman Period (85 cm) and elevated pollen values (7.9%) in the XVIII century (35 cm) but essentially no hemp fibres in the sediment. This suggests cultivation of hemp in the lake vicinity in the Early Modern Period however, no HR. Therefore, we propose the fibre count analysis as a significant additional proxy for investigating HR, especially in the lake sediments with low cannabis type pollen percentages.

The composition of bacterial communities of a hemp rettery was significantly different to the microbiome of the recovered lake. We observed enrichment in the abundance of Chloroflexi, Planctomycetes and Bacteroidetes in the 'HR' group, and Proteobacteria and Actinobacteria in the 'nHR' group. The increase in Actinobacteria abundance has been associated with the meso-/oligotrophic lakes, suggesting the recovery of LS after the episodes of HR³². Most Actinobacteria are aerobic and can form dormant spores under low nutrient conditions. Although many taxa are capable of plant polysaccharides degradation, the abundance of Actinomycetes usually decreases with decreasing oxygen levels³³, as observed in LS. Proteobacteria are generally abundant in lakes and lake sediments and range from 20% to even 80%³⁴. They were reported to be characteristic of both oligotrophic^{32,35} and eutrophic lakes³⁴ which is most probably associated with a great ecological richness within this phylum³³. Proteobacteria, including Betaproteobacteria, Deltaproteobacteria and Gammaproteobacteria, were more predominant in the 'nHR' group. These taxa consist of diverse, mostly aerobic bacteria presenting rich catabolic capabilities and important for the biogeochemical cycling of elements³⁶. Especially the Betaproteobacteria orders play an

important role in nitrification³⁷, denitrification³⁸ and nitrogen fixing³⁹. Although many different proteobacteria are capable of anaerobic organic matter utilisation, also of plant polysaccharide origin³⁶, upon the conditions of hypoxia and high trophic associated with HR, the proteobacteria classes did not shift towards anaerobic and/or fermentative orders but instead, became less abundant.

Chloroflexi, Planctomycetes and Bacteroidetes are gram-negative nonproteobacteria capable of anaerobic metabolism. Chloroflexi, the second most represented bacterial phylum in LS, became enriched during HR. This group is relatively understudied, with Dehalococcoidia's role in halogen cycling and dechlorination being best understood⁴⁰. As more metagenome-assembled genomes (MAGs) of Chloroflexi become available, their role in sediment carbon cycling including fermentation and anaerobic carbon fixation is being recognised. For example, the reductive Acetyl-CoA pathway recently described in the uncultured sediment-associated Chloroflexi⁴¹ was partially inferred in the 'HR' group possibly suggesting their functional role. Planctomycetes constitute another understudied bacterial taxon and have been shown to be more abundant in high trophic lakes^{42,43}. Like proteobacteria, Planctomycetes exhibit very diversified metabolism including chemoheterotrophic aerobes and facultative anaerobes and are capable of anammox type reactions, as well as complex carbohydrate fermentation^{32,33}. Bacteroidetes were identified to be associated with the dew-retting of hemp¹⁴. They take part in the degradation of complex biopolymers, including cellulose and pectin, and increase abundance in response to elevated dissolved organic carbon and after cyanobacterial blooms³³. The role of Chloroflexi, Planctomycetes and Bacteroidetes in the HR process was further reiterated as they were identified to be involved in the inferred cellulose and pectin hydrolysis pathways. Such shifts in bacterial communities towards anaerobic or facultatively anaerobic phyla capable of plant polysaccharides utilisation indicate retting-induced hypoxia of the lake and show bacterial adaptations to the conditions of low oxygen and high plant organic matter. This was supported by the set of environmental factors which best explained the distribution of taxonomic data and included HR, forestation, fibre count and carbonates. Hence, the overrepresented taxa in the 'HR' group indeed encompassed more polysaccharides utilising bacteria with no or low oxygen requirements while the overrepresented taxa in the 'nHR' group constituted a balanced ecological system.

During the episodes of HR, the sources and ratios of the allochthonous vs autochthonous organic matter were biased. The main source of autochthonous organic matter in lakes is usually phytoplankton and aquatic macrophytes while allochthonous organic matter originates from the catchment, especially terrestrial plants. These origin differences affect the sediment carbon to nitrogen ratio (C/N) as terrestrial plants are rich in polysaccharides and phytoplankton exhibits higher protein content⁴⁴. During HR, an additional source of allochthonous organic matter was introduced into the lake in the form of hemp straws disturbing the C/N which affected the microbiome metabolism. In the 'nHR' bacterial community the overrepresented reads were classified to diverse types of metabolism including different nitrogen and sulfur metabolic pathways, together with carbon metabolism, while the functional profile of the 'HR' group was dominated by the carbon cycle, especially the anaerobic pathways.

The functional structures of the two groups were significantly different, however, characteristic of low oxygen conditions of the benthic environment. LS is a dimictic, thermally stratified lake exhibiting temperature and density dependent water gradient. Consequently, in the summer and winter months, the deepest strata may become oxygen poor⁴⁵. In LS, under the conditions of oligotrophy and in the absence of HR, light could reach deeper layers of the lake (increased euphotic zone) allowing growth of anoxygenic phototrophic bacteria capable of sulfur metabolism, as shown by the taxonomy and metabolic potential analyses. The dissimilatory sulfate reduction and oxidation, as well as SOX metabolic pathways, which depend on light and inorganic sulfur compounds availability, were overrepresented in the 'nHR' group and classified predominantly to different Proteobacteria⁴⁶. During the episodes of HR, trophic increased significantly resulting in the decreased euphotic zone, likely rendering the phototrophic sulfur bacteria dependent metabolism very limited or even impossible. As dissimilatory sulfate reduction is the major anaerobic process of organic carbon mineralisation⁴⁷, disturbed sulfur cycling can affect microbial metabolism. Indeed, eutrophication and lake shallowing were shown to influence microbial taxonomy and functionality and thus, redox chemistry and sulfur metabolism in lake sediments⁴⁸. During the episodes of reduced anthropopressure, the sediment microbiome also exhibited increased potential for dissimilatory nitrate reduction and denitrification as well as, to some degree, nitrification. The first two processes are anaerobic and characteristic of the long-term anaerobic habitats such as freshwater sediments (dissimilatory nitrate reduction) or habitats of transient oxygen levels such as oligotrophic sediments or soils (denitrification). Both pathways take part in the electron transport in the absence of oxygen⁴⁹, however, were not reconstructed in the 'HR' group, most probably because they depend on available nitrate as an electron acceptor, which is depleted in shallow eutrophic lakes⁴⁸. The availability of sulfate and nitrate is also dependent on pH and amount of organic carbon⁵⁰, both of which varied upon HR.

Anaerobic carbon fixation was by far the most abundant pathway of carbon metabolism in both groups which is usual for the oxygen low conditions of the lake sediments. Under the conditions of an increased trophic, the microbiome of a hemp rettery showed higher potential for anaerobic carbon fixation and fermentation, while the microbiomes in the 'nHR' group were more abundant in marker genes of the aerobic pathways (aerobic carbon fixation and CO oxidation). The utilisation of complex polysaccharides of the allochthonous origin has been linked to the oligo-mesotrophic lakes in which the autochthonous carbon sources are limited⁴². In LS however, the increase of trophic status during HR was a result of introducing large amounts of allochthonous carbon—hemp stems rich in complex plant polysaccharides. The microbiome of the rettery was capable of the utilisation of this carbon source as near complete cellulose and pectin degradation pathways were overrepresented in this group. This functional characteristic of the rettery microbiome was also shown experimentally⁵¹. Thus, it can be concluded that over the centuries the source and amount of organic carbon in LS were one of the drivers of the microbiome fluctuations, both taxonomic and functional. In contrast to nitrogen and sulfur metabolism, complex plant polysaccharides were degraded predominantly by Chloroflexi, Planctomycetes, Bacteroidetes

and Clostridia demonstrating the considerable metabolic potential and environmental significance of these still understudied bacterial taxa³³.

Conclusions

Using the NGS and paleo-ecological data we were able to examine the microbiome structure and function, as well as trophic changes in LS, the site of an ancient hemp rettery. This provided an excellent opportunity to investigate the microbial response to a well-defined stressor over several centuries. To the authors' best knowledge, the lake trophic and microbiome of a historical hemp retting site have not been described before. We showed that the increased allochthonous carbon supply (hemp stems) to the lake correlated with the conditions of elevated trophic and low dissolved oxygen, as well as shifts in the microbial taxonomy and functionality. During HR, the lake microbiome was dominated by the complex polysaccharide utilising bacteria capable of anaerobic and/or facultatively anaerobic growth, while upon cessation of HR practices both lake trophic and microbiome recovered.

Materials and methods

Lake bottom core collection and sub-sampling. The LS-C19 sediment core of 98 cm was sampled in January 2019 from the central part of the frozen lake (51°18'17.06"N, 23°21'56.22"E) at a water depth of 8.00 m, with Ø90 mm UWITEC gravity corer. Immediately after collection the core was wrapped in non-transparent black foil and transported for sub-sampling to a laboratory where no previous work with environmental samples had been performed. 1 cm thick samples in the form of disks were collected resulting in a total of 98 samples. The layers were divided with a stainless steel cutting blade and samples for DNA extraction were collected from the middle of each cut disk with a stainless-steel spoon, ensuring an approximately 5 cm margin to the edge of the disk. All stainless-steel utensils were cleaned with distilled water and 70% ethanol between uses. Samples for DNA extraction were immediately transferred into plastic bags, flash frozen in liquid nitrogen and stored at -80 °C. The remaining material was transferred into plastic bags and stored at 4 °C until further processing.

Radiocarbon dating and age-depth modelling. The composite age-depth model was based upon ²¹⁰Pb and radiocarbon AMS dating (Supp Table 5). Data was obtained from 3 parallel profiles: JS-A (²¹⁰Pb, unpublished), JS-C¹⁶ (1 AMS ¹⁴C) and LS-C19 (3 AMS ¹⁴C), which were precisely correlated based on palynological, LOI₅₅₀, LOI₉₅₀ and subfossil Cladocera data (Supp Fig. 3, Supp Table 1).

The ¹⁴C measurements were performed at the Poznań Radiocarbon Laboratory using the Accelerator Mass Spectrometry. The ¹⁴C dates were calibrated using OxCal 4.4.2 software⁵² with the application of the IntCal20 calibration curve⁵³. For age-depth modelling the P_Sequence algorithm of OxCal was used^{52,54} with the parameters: $k_0 = 0.2$, $\log_{10}(k/k_0) = 2$, and interpolation = 1 cm.

The sediment sequence of JS-A core was dated by ²¹⁰Pb method, at the Laboratory of Institute of Geological Sciences, Polish Academy of Science in Warsaw. For ²¹⁰Pb analysis, 3 cm³ of sediment was taken from each level. For all sediment samples, bulk density and water content were determined. The samples were dried at 105 °C to constant weight and homogenized in an agate mortar. The ²¹⁰Pb activity was determined indirectly by alpha-spectrometry measurement of ²¹⁰Po ($E_\alpha = 5.31$ MeV, $T_{1/2} = 138$ days) activity in 0.1–0.9 g sub-samples⁵⁵. Polonium was extracted from the samples using concentrated hydrochloric and nitric acids and deposited on silver disks⁵⁵. For the complete removal of organic matter, 30% perhydrol was used. The activity of ²¹⁰Po and ²⁰⁸Po was measured via an alfa OCTETE PC (EG&G ORTEC) in 12 samples. A known amount of ²⁰⁸Po ($E_\alpha = 5.11$ MeV) was added to the weighted sample as an internal control, and the constant rate of unsupported ²¹⁰Pb supply model (CRS) was used to calculate the sediment age⁵⁶. Supported ²¹⁰Pb was determined by measurements of old sediments (older than 150–200 years) that contain no allochthonous ²¹⁰Pb, assuming constant activity of authigenic ²¹⁰Pb along the sediment column. For samples over the extent of the dating method, age was determined by extrapolation of the sedimentation rate of the lowermost samples. An age-depth function was calculated using the randomization method and fitted by means of the LOESS procedure⁵⁷.

Loss on ignition (LOI). Contents of the organic matter and carbonates in the sediments were determined by a sequential loss-on-ignition method based on combustion of dry, homogenized sediment in 550 °C/4 h (organic matter) and 950 °C/2 h (carbonates)⁵⁸ (Supp Data 1).

Palynological analysis. The selected samples (1 cm³ of sediment) were macerated using the standard method of Erdman's acetolysis. Carbohydrates were removed with 10% HCl and the mineral fraction with 40% HF⁵⁹. Pollen spectra were counted on at least two microscopic slides 18 × 18 mm. From 400 up to 500 grains were counted per single sample. Pollen concentration in all examined samples was high (362,000–1,298,000 per cm³). The cumulative sum used for percentage calculations was the sum of tree and shrub pollen (arboreal pollen—AP), and of herb pollen (non-arboreal pollen—NAP), excluding the pollen of aquatic and reed-swamp plants, *Pteridophyta* and *Bryophyta* spores, as well as *Pediastrum* and *Botryococcus* algae colonies. The percentage of AP was used as a forestation proxy, whereas the NAP was further divided into proxies describing types of anthropogenic influence⁶⁰. These included pollens of different cereals (*Secale cereale*, *Triticum* sp., *Fagopyrum* sp., *Linum usitatissimum* and others assigned as *Cerealia* undiff.) and hemp (*Cannabis* sp.) describing cultivation of these crops. Pollen of *Plantago lanceolata* was used here as a proxy for pasture and pollens of segetal and ruderal weeds including *Centaurea cyanus*, *Convolvulus arvensis*, *Spergula arvensis*, *Scleranthus annuus*, *Artemisia* sp., *Chenopodiaceae*, *Brassicaceae*, *Urtica* and *Rumex acetosella* were used as indicators of anthropogenic impact (Supp Fig. 2, Supp Table 3).

Plant macrofossil material (hemp fibres) identification and quantification. The count of plant macrofossil material (hemp fibres) was estimated. 100 μL of wet sediment from each layer was weighed and diluted tenfold with distilled water. Thirty aliquots of 10 μL per sample were prepared and diluted further to obtain sufficient visibility of the individual sediment components in a petri dish. The quantity of fibres was counted in each aliquot and the number of fibres per sample was expressed as an arithmetic mean ($n = 30$). This was then converted into a number of fibres per 1 g of wet sediment.

The identification of hemp fibres in the 8 samples (from the depths 35, 55, 63, 65, 67, 76, 85 and 93) was performed by application of polarised light microscopy (PLM) with a modified Herzog test²².

Diatoms. Diatom samples were prepared using standard techniques introduced by Battarbee⁶¹. In total, 18 samples were treated with 10% HCl to remove carbonates, washed with distilled water, and treated with 30% H_2O_2 in a water bath to remove organic matter. Diatoms were mounted in Naphrax on permanent microscopic slides. Species identification was based on literature^{62–68}. In each sample, about 500 diatom frustules were counted to estimate the relative abundance of individual taxa. The taxonomic nomenclature was based on AlgalBase⁶⁹. The list of dominant diatom taxa and diatoms percentage diagrams of LS-C19 core are provided in Supp Table 2 and Supp Fig. 4 respectively.

Cladocera. Samples of 1 cm^3 were prepared according to a slightly modified standard procedure⁷⁰. Each sample was heated in 10% KOH until boiling, with a subsequent manual stirring. The residue was washed with distilled water through a 33- μm mesh sieve, treated with 10% HCl, washed again, transferred into a scaled tube and topped up to 10 ml volume. Prior to counting, the remains were stained with safranin. The samples were analysed under a compound light microscope with a 100–400 \times magnification. All the remains from each sample were enumerated. A minimum of 200 remains of Cladocera were counted in each sample. The most abundant body part of each taxon was chosen to represent the number of individuals. The percentage of planktonic and littoral taxa (P/L ratio) was used as a lake water level inference. Cladocera based trophic inference was estimated based on the ecological preferences of the examined taxa^{71,72} (Supp Table 1, Supp Fig. 3).

DNA isolation and sequencing. A total of sixteen samples originating from eight layers (in duplicates), with a mass of two to three grams of frozen sediments were processed for *sedDNA* extraction. The samples were ground in liquid nitrogen using pestle and mortar as well as aluminium oxide as an abrasive and lysis buffer (6 M guanidine hydrochloride, 1% *N*-lauroylsarcosine sodium salt, 50 mM Tris-HCl pH 7.6, 10 mM EDTA, 1% β -mercaptoethanol). Once thawed, the supernatants were collected by centrifugation and incubated with 6 mg proteinase K for 2 h at 55 °C. The samples were further purified using the phenol:chloroform extraction method, followed by overnight ethanol precipitation at –20 °C. DNA samples were then purified with a commercially available Anti-Inhibitor Kit (A&A Biotechnology) to remove PCR inhibitors and were then treated with additional proteinase K digestion (0,6 mg for 20 min at 55 °C). Next, the samples were purified using Zymo DNA Clean and Concentrator-25 kit and stored at –20 °C. After that, we performed a DNA repair step using a NEBNext FFPE DNA Repair Mix, followed by an additional clean-up step (Zymo DNA Clean and Concentrator-25 kit) and storage at –80 °C. This resulted in the absence of visible degradation, as well as improved 260/280 and 260/230 ratios (as measured by NanoDrop). In order to minimise contamination risk, all utensils including pestles, mortars, pipettes, as well as laboratory benches were routinely cleaned with laboratory grade disinfectant removing DNases, RNases, residual nucleic acids and inactivating viruses and bacteria. A designated laminar flow hood for PCR work was sterilised with UV light before each use. Filtered pipette tips were used for *sedDNA* work. Personal protective equipment was used at all times.

Library preparation and sequencing was performed by Genomed S.A. (Warsaw, Poland). For library preparation NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs, E7645L) was used, according to the manufacturer's guidelines. Sequencing was performed using Illumina HiSeq 4000 PE150.

Taxonomic and functional assignment of metagenomes. Sequencing of 16 samples obtained from eight sediment layers resulted in an average of 134.37 million high-quality reads. A total of ~314 Gb data was obtained with 147 bp of mean read size distributions for all paired-end fastq files (Supp Data 2a). The quality of data received from the sequencing company was checked with FastQC⁷³ software. All samples were pre-filtered and adapters were removed by the sequencing company. FastQC did not report any file with adapters content greater than 0.1%. The samples also met the quality criteria (mean quality score > 35). Filtering was applied using reformat.sh⁷⁴ tool to retain only sequences over 100 nucleotides. The high-quality filtered paired-end reads were assembled into contigs using megahit v1.2.9⁷⁵ with 'meta-large' preset which resulted in a total of 2,368,565 to 4,560,710 contigs with 28,732,753,935 total number of bases (N50 values of 556–650 bp). The resulting files were mapped to the entire non-redundant protein sequence database (NCBI NR, accessed on 20.01.2021)⁷⁶ using DIAMOND v. 2.0.6.144⁷⁷. The following parameters were used: --long-reads, block size 5 (-b 5), index chunks 1 (-c 1), the number of CPU threads 28 (-p 28), and the output file as DAA (--outfmt 100), the other parameters remained at default settings. The resulting DAA files were then meganized using the daa-meganizer⁷⁸ tool with the latest available mapping file for NCBI-nr accessions to taxonomic and functional classes (NCBI NR⁷⁶, eggNOG⁷⁹, SEED⁸⁰)—megan-map-Jul2020-2.db. Data from all samples prepared in this way were loaded into the MEGAN v. 6.20.16⁷⁸ software. Using a graphical interface, plots showing the taxonomic composition of individual samples, tables containing all detected taxonomies, unweighted pair group method with arithmetic mean dendrograms (UPGMA), as well as Bray–Curtis dissimilarity matrices based on the data assigned to the bacteria, were obtained.

The SqueezeMeta v. 1.3.0⁸¹ pipeline in sequential mode was used for functional classification. Contigs obtained at an earlier stage were used (-extassembly option). The key steps of the pipeline consisted of gene prediction using Prodigal⁸² and similarity search to the GenBank⁸³, eggNOG⁷⁹ and KEGG⁸⁴ databases using DIAMOND⁷⁷. Prediction of ORF resulted in 3.24–6.04 million of CDS per sample and between 29.73% and 37.85% of CDS per sample were assigned to KEGG Orthology (KO) (Supp Data 2b,c). The data obtained in this way was then loaded into R v. 4.0.4⁸⁵ and, using the SQMtools⁸⁶ library (function loadSQM), only data taxonomically assigned to the bacterial group was selected for further analysis (function subsetTax). Tables containing the abundance and TPM (transcript per million) of each KEGG Orthologs (KOs) for each sample were exported. The TPM of carbon metabolic marker genes (KOs accession numbers) identified in previous studies^{26,32} (Supp Table 4) were compared and presented as a heatmap with pheatmap function of the pheatmap⁸⁷ R package.

Statistical analysis. Data normalisation for all detected taxonomies was performed in MEGAN⁷⁸ in sample comparison mode (normalisation to the smallest sample size). The Bray–Curtis dissimilarity matrix of the taxonomic data was also constructed in MEGAN. For all samples, bacterial KOs abundances were normalised to TPM values by SqueezeMeta library and were exported using SQMtools library in R. Bray–Curtis dissimilarity matrix was calculated with vegdist function from the obtained normalised bacterial KOs. The normalisation of the environmental data was performed with a function scale from the R base package. The Euclidean distance matrix was calculated for the normalised environmental data using the dist function from the stats R package. All this data was used for the following statistical analyses. Unweighted pair group method with arithmetic mean (UPGMA) clustering analysis and principal component analysis (PCA) were used to display and compare the patterns of taxonomic structure and functional characteristics of the bacterial communities. PCA plots were computed in Statistical Analysis of Metagenomic Profiles (STAMP)⁸⁸ for taxonomy data and all bacterial KOs. UPGMA dendrogram for the taxonomy data was prepared in MEGAN, and for the functional data (all bacterial KOs) was plotted via hclust function from the stats package in R. The significant variance ($p < 0.01$) between ‘HR’ and ‘nHR’ groups was assessed by permutational multivariate analysis of variance (PERMANOVA) using adonis2 function in the vegan⁸⁹ package. Two group statistical testing was carried out using two-sided Welch’s *t* test in STAMP with Welch’s inverted confidence (0.95) interval method. A percentile bootstrapping method (1000 replications) was used to estimate confidence intervals, and the false discovery rate (FDR) in multiple testing was corrected with the Benjamini–Hochberg FDR method. Differentially abundant KOs in the ‘HR’ and ‘nHR’ groups, identified by STAMP (corrected *p*-value < 0.05) were used in MinPath (Minimal set of Pathways)²⁵ to reconstruct pathways. The correlation between the overall taxonomic and functional composition was calculated by the Mantel test (9999 permutations, Pearson’s correlation method). Correlation analyses between all environmental factors and the abundances of the taxonomies or functional categories were performed using the bioenv function from the vegan package using Pearson’s correlation method. The mantel function was used to calculate empirical *p*-values, which were corrected for multiple testing using the p.adjust function from the R stats library (Benjamini–Hochberg FDR, BH).

Carbon metabolic marker genes KOs that are differentially segregated across the two groups (‘HR’ vs ‘nHR’) were identified by random forest analysis with Boruta⁹⁰ feature selection (Boruta function in R package Boruta, maxRuns = 9999, mean importance > 4). To infer the genetic potential of carbon metabolic processes, the relative abundance of selected marker genes was used as described elsewhere^{26,32}. The individual processes were then normalised to 100% of the total carbon metabolism and their statistical significance was checked using the Wilcoxon test in R (function wilcox.test in stats package). The data was plotted with a box and whisker plot using the ggplot function of ggplot2⁹¹ library in R. All the above statistical analyses were performed using R version 4.0.4 and STAMP version v2.1.3.

Data availability

The cleaned data was deposited in the European Nucleotide Archive (ENA: PRJEB36143, PRJEB39351).

Received: 29 November 2021; Accepted: 16 May 2022

Published online: 25 May 2022

References

- Parducci, L. *et al.* Shotgun environmental DNA, pollen, and macrofossil analysis of lateglacial lake sediments from southern Sweden. *Front. Ecol. Evol.* **7**, 189 (2019).
- Slowinski, M. *et al.* The role of Medieval road operation on cultural landscape transformation. *Sci. Rep.* **11**, 1–9 (2021).
- Hargan, K. E. *et al.* Post-glacial lake development and paleoclimate in the central Hudson Bay Lowlands inferred from sediment records. *J. Paleolimnol.* **64**, 25–46 (2020).
- Domaizon, I., Winegardner, A., Capo, E., Gauthier, J. & Gregory-Eaves, I. DNA-based methods in paleolimnology: New opportunities for investigating long-term dynamics of lacustrine biodiversity. *J. Paleolimnol.* **58**, 1–21 (2017).
- Garner, R. E., Gregory-Eaves, I. & Walsh, D. A. Sediment metagenomes as time capsules of lake microbiomes. *mSphere* **5**, e00512–20 (2020).
- Dommain, R. *et al.* The challenges of reconstructing tropical biodiversity with sedimentary ancient DNA: A 2200-year-long metagenomic record from Bwindi Impenetrable Forest, Uganda. *Front. Ecol. Evol.* **8**, 218 (2020).
- Pedersen, M. W. *et al.* A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. *Quatern. Sci. Rev.* **75**, 161–168 (2013).
- Moguel, B. *et al.* Holocene life and microbiome profiling in ancient tropical Lake Chalco, Mexico. *Sci. Rep.* **11**, 1–15 (2021).
- McPartland, J. M., Guy, G. W. & Hegman, W. Cannabis is indigenous to Europe and cultivation began during the Copper or Bronze age: A probabilistic synthesis of fossil pollen studies. *Veg. Hist. Archaeobot.* **27**, 635–648 (2018).
- Mercuri, A. M., Accorsi, C. A. & Bandini Mazzanti, M. The long history of *Cannabis* and its cultivation by the Romans in central Italy, shown by pollen records from Lago Albano and Lago di Nemi. *Veg. Hist. Archaeobot.* **11**, 263–276 (2002).

11. Kittel, P. *et al.* A multi-proxy reconstruction from Lutomięsk–Koziówki, Central Poland, in the context of early modern hemp and flax processing. *J. Archaeol. Sci.* **50**, 318–337 (2014).
12. Di Candilo, M. *et al.* Effects of selected pectinolytic bacterial strains on water-retting of hemp and fibre properties. *J. Appl. Microbiol.* **108**, 194–203 (2010).
13. Tamburini, E., León, A. G., Perito, B. & Mastromei, G. Characterization of bacterial pectinolytic strains involved in the water retting process. *Environ. Microbiol.* **5**, 730–736 (2003).
14. Ribeiro, A. *et al.* Microbial diversity observed during hemp retting. *Appl. Microbiol. Biotechnol.* **99**, 4471–4484 (2015).
15. Ventorino, V. *et al.* Exploring the microbiota dynamics related to vegetable biomasses degradation and study of lignocellulose-degrading bacteria for industrial biotechnological application. *Sci. Rep.* **5**, 1–13 (2015).
16. Kulesza, P., Suchora, M., Pidek, I. A., Dobrowolski, R. & Alexandrowicz, W. P. The Holocene palaeoenvironmental changes reflected in the multi-proxy studies of Lake Słone sediments (SE Poland). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **363–364**, 79–98 (2012).
17. Pęczuła, W., Suchora, M. & Toporowska, M. Ecological and trophic status of two small hardwater lakes of the Lublin Upland with agricultural catchments. *Teka Kom. Ochr. i Kształtowania Środowiska Przyr.* **11**, 146–155 (2014).
18. Solon, J. *et al.* Physico-geographical mesoregions of Poland: Verification and adjustment of boundaries on the basis of contemporary spatial data. *Geogr. Pol.* **91**, 143–170 (2018).
19. Buko, A., Dzieńkowski, T. & Kusiak, J. Dating early Medieval ceramics by the thermoluminescence method: Busówno site case study. *Archeol. Pol.* **1**, 25–49 (2008).
20. Dobrowolski, R. *et al.* Environmental conditions of settlement in the vicinity of the mediaeval capital of the Cherven Towns (Czermno site, Hrubieszów Basin, Eastern Poland). *Quatern. Int.* **493**, 258–273 (2018).
21. Bursche, A., Kowalski, K., Rogalski, B., Kinecka, A. & Witek, M. *Barbarian Tsunami: Migration Period Between the Odra and the Vistula* (University of Warsaw; National Museum in Szczecin, 2017).
22. Haugan, E. & Holst, B. Determining the fibrillar orientation of bast fibres with polarized light microscopy: The modified Herzog test (red plate test) explained. *J. Microsc.* **252**, 159–168 (2013).
23. Steffen, W., Broadgate, W., Deutsch, L., Gaffney, O. & Ludwig, C. The trajectory of the anthropocene: The great acceleration. *Anthr. Rev.* **2**, 81–98 (2015).
24. Meyers, P. A. & Teranes, J. L. Sediment organic matter. in *Tracking Environmental Change Using Lake Sediments* 239–269. https://doi.org/10.1007/0-306-47670-3_9 (2005).
25. Ye, Y. & Doak, T. G. A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput. Biol.* **5**, e1000465 (2009).
26. Lauro, F. M. *et al.* An integrative study of a meromictic lake ecosystem in Antarctica. *ISME J.* **5**, 879–895 (2011).
27. Nakagawa, T., De Beaulieu, J. L. & Kitagawa, H. Pollen-derived history of timber exploitation from the Roman period onwards in the Romanche valley, central French Alps. *Veg. Hist. Archaeobot.* **9**, 85–89 (2000).
28. Schofield, J. E. & Waller, M. P. A pollen analytical record for hemp retting from Dungeness Foreland, UK. *J. Archaeol. Sci.* **32**, 715–726 (2005).
29. Cox, M., Chandler, J., Cox, C., Jones, J. & Tinsley, H. The archaeological significance of patterns of anomalous vegetation on a raised mire in the Solway Estuary and the processes involved in their formation. *J. Archaeol. Sci.* **28**, 1–18 (2001).
30. Riera, S., López-Sález, J. A. & Julià, R. Lake responses to historical land use changes in northern Spain: The contribution of non-pollen palynomorphs in a multiproxy study. *Rev. Palaeobot. Palynol.* **141**, 127–137 (2006).
31. Saarnisto, M., Huttunen, P. & Tolonen, K. Annual lamination of sediments in Lake Lovojärvi, southern Finland, during the past 600 years. *Ann. Bot. Fenn.* **14**, 35–45 (1977).
32. Shen, M. *et al.* Trophic status is associated with community structure and metabolic potential of planktonic microbiota in plateau lakes. *Front. Microbiol.* **10**, 2560 (2019).
33. Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D. & Bertilsson, S. A guide to the natural history of freshwater lake bacteria. *Microbiol. Mol. Biol. Rev.* **75**, 14–49 (2011).
34. Huang, W., Chen, X., Jiang, X. & Zheng, B. Characterization of sediment bacterial communities in plain lakes with different trophic statuses. *Microbiol. Open* **6**, e00503 (2017).
35. Ji, B. *et al.* Bacterial communities of four adjacent fresh lakes at different trophic status. *Ecotoxicol. Environ. Saf.* **157**, 388–394 (2018).
36. Kersters, K. *et al.* Introduction to the proteobacteria. *Prokaryotes* **5**, 3–37 (2006).
37. Dang, H. *et al.* Diversity, abundance, and spatial distribution of sediment ammonia-oxidizing Betaproteobacteria in response to environmental gradients and coastal eutrophication in Jiaozhou Bay, China. *Appl. Environ. Microbiol.* **76**, 4691–4702 (2020).
38. Prosser, J. I., Head, I. M. & Stein, L. Y. The family Nitrosomonadaceae. In *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* Vol. 9783642301 (eds Rosenberg, E. *et al.*) 901–918 (Springer, 2014).
39. Pérez-Pantoja, D. *et al.* Genomic analysis of the potential for aromatic compounds biodegradation in Burkholderiales. *Environ. Microbiol.* **14**, 1091–1117 (2012).
40. Krzmarzick, M. J., Mcnamara, P. J., Crary, B. B. & Novak, P. J. Abundance and diversity of organohalide-respiring bacteria in lake sediments across a geographical sulfur gradient. *FEMS Microbiol. Ecol.* **84**, 248–258 (2013).
41. Hug, L. A. *et al.* Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome* **1**, 1–17 (2013).
42. Kiersztyn, B. *et al.* Structural and functional microbial diversity along a eutrophication gradient of interconnected lakes undergoing anthropopressure. *Sci. Rep.* **9**, 11144 (2019).
43. Dedysh, S. N. *et al.* Wide distribution of Phycisphaera-like planctomycetes from WD2101 soil group in peatlands and genome analysis of the first cultivated representative. *Environ. Microbiol.* **23**, 1510–1526 (2021).
44. Fan, J. *et al.* Carbon and nitrogen signatures of sedimentary organic matter from Dali Lake in Inner Mongolia: Implications for Holocene hydrological and ecological variations in the East Asian summer monsoon margin. *Quatern. Int.* **452**, 65–78 (2017).
45. Zhu, K., Lauro, F. M. & Su, H. Stratification modelling of key bacterial taxa driven by metabolic dynamics in meromictic lakes. *Sci. Rep.* **8**, 1–7 (2018).
46. Imhoff, J. F., Hiraishi, A. & Süling, J. Anoxygenic phototrophic purple bacteria. *Bergey's Man. Syst. Archaea Bact.* 1–23. <https://doi.org/10.1002/9781118960608.bm00002> (2015).
47. Watanabe, T., Kojima, H., Takano, Y. & Fukui, M. Diversity of sulfur-cycle prokaryotes in freshwater lake sediments investigated using aprA as the functional marker gene. *Syst. Appl. Microbiol.* **36**, 436–443 (2013).
48. Han, X., Schubert, C. J., Fiskal, A., Dubois, N. & Lever, M. A. Eutrophication as a driver of microbial community structure in lake sediments. *Environ. Microbiol.* **22**, 3446–3462 (2020).
49. Tiedje, J. M., Sørensen, J. & Chang, Y.-Y.L. Assimilatory and dissimilatory nitrate reduction: Perspectives and methodology for simultaneous measurement of several nitrogen cycle processes. *Ecol. Bull.* **33**, 331–342 (1981).
50. Burgin, A. J., Yang, W. H., Hamilton, S. K. & Silver, W. L. Beyond carbon and nitrogen: How the microbial energy economy couples elemental cycles in diverse ecosystems. *Front. Ecol. Environ.* **9**, 44–52 (2011).
51. Kalisz, G. *et al.* Vibrational spectroscopic analyses and imaging of the early middle ages hemp bast fibres recovered from lake sediments. *Molecules* **26**, 1314 (2021).
52. Ramsey, C. B. & Lee, S. Recent and planned developments of the program OxCal. *Radiocarbon* **55**, 720–730 (2013).

53. Reimer, P. J. *et al.* The IntCal20 northern hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* **62**, 725–757 (2020).
54. Bronk Ramsey, C. Radiocarbon dating: Revolutions in understanding. *Archaeometry* **50**, 249–275 (2008).
55. Flynn, W. W. The determination of low levels of polonium-210 in environmental materials. *Anal. Chim. Acta* **43**, 221–227 (1968).
56. Appleby, P. G. Chronostratigraphic Techniques in Recent Sediments. in *Tracking Environmental Change Using Lake Sediments* 171–203 (Kluwer Academic Publishers, 2002). https://doi.org/10.1007/0-306-47669-X_9
57. Cleveland, W. S. & Devlin, S. J. Locally weighted regression: An approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* **83**, 596 (1988).
58. Heiri, O., Lotter, A. F. & Lemcke, G. Loss on ignition as a method for estimating organic and carbonate content in sediments: Reproducibility and comparability of results. *J. Paleolimnol.* **25**, 101–110 (2001).
59. Berglund, B. E. & Ralska-Jasiewiczowa, M. *Handbook of Holocene Palaeoecology and Palaeohydrology* (Wiley, 1986). <https://doi.org/10.2307/2260436>.
60. Behre, K.-E. *Anthropogenic Indicators in Pollen Diagrams* (Balkema, 1986).
61. Battarbee, R. W. Diatom analysis. in *Handbook of Holocene Palaeoecology and Palaeohydrology* 527–570 (Wiley, 1986).
62. Krammer, K. The genus *Pinnularia*. in *Diatoms of Europe: Diatoms of the European Inland Waters and Comparable Habitats 1* 1–703 (A. R. G. Gantner Verlag K. G., 2000).
63. Krammer, K. The genus *Cymbella*. in *Diatoms of Europe. Diatoms of the European Inland Waters and Comparable Habitats 3* 1–584 (A. R. G. Gantner Verlag K. G., 2002).
64. Krammer, K. & Lange-Bertalot, H. Bacillariophyceae. 1. In *Süßwasserflora von Mitteleuropa* Vol. 876 (eds Ettl, H. *et al.*) (Gustav Fisher Verlag, 1986).
65. Krammer, K. & Lange-Bertalot, H. Bacillariophyceae. 2. In *Süßwasserflora von Mitteleuropa* (eds Ettl, H. *et al.*) (Gustav Fisher Verlag, 1988).
66. Krammer, K. & Lange-Bertalot, H. Bacillariophyceae. 3. In *Süßwasserflora von Mitteleuropa* (eds Ettl, H. *et al.*) 1–576 (Gustav Fisher Verlag, 1991).
67. Lange-Bertalot, H. & Krammer, K. Morphology and taxonomy of *Surirella ovalis* and related taxa. *Diatom. Res.* **2**, 77–95 (1987).
68. Van Dam, H., Mertens, A. & Sinkeldam, J. A coded checklist and ecological indicator values of freshwater diatoms from The Netherlands. *Netherlands J. Aquat. Ecol.* **28**, 117–133 (1994).
69. Guiry, M. D. & Guiry, G. M. AlgaeBase. World-wide electronic publication. <https://www.algaebase.org/> (2020).
70. Frey, B. S. & Pamini, P. Cladocera analysis. In *Handbook of Holocene Palaeoecology and Palaeohydrology* (ed. Berglund, B. E.) 677–692 (Wiley, 1986).
71. Bjerring, R. *et al.* Subfossil Cladocera in relation to contemporary environmental variables in 54 Pan-European lakes. *Freshw. Biol.* **54**, 2401–2417 (2009).
72. Chen, G., Dalton, C. & Taylor, D. Cladocera as indicators of trophic state in Irish lakes. *J. Paleolimnol.* **44**, 465–481 (2010).
73. Andrews, S., Krueger, F., Seconds-Pichon, A., Biggins, F. & Wingett, S. FastQC. A quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
74. Bushnell, B., Rood, J. & Singer, E. BBTools Software Package. <https://sourceforge.net/projects/bbmap/> (2017).
75. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
76. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, D501–D504 (2005).
77. Buchfink, B., Reuter, K. & Drost, H. G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).
78. Huson, D. H. *et al.* MEGAN Community Edition—Interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* **12**, e1004957 (2016).
79. Huerta-Cepas, J. *et al.* EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
80. Overbeek, R. *et al.* The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33**, 5691–5702 (2005).
81. Tamames, J. & Puente-Sánchez, F. SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Front. Microbiol.* **10**, 3349 (2019).
82. Hyatt, D. *et al.* Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 1–11 (2010).
83. Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **44**, D67–D72 (2016).
84. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
85. R Core Team. R: A language and environment for statistical computing. <http://www.r-project.org> (2021).
86. Puente-Sánchez, F., García-García, N. & Tamames, J. SQMtools: Automated processing and visual analysis of omics data with R and anvio. *BMC Bioinform.* **21**, 1–11 (2020).
87. Kolde, R. pheatmap: Pretty heatmaps. <https://cran.r-project.org/web/packages/pheatmap/> (2015).
88. Parks, D. H., Tyson, G. W., Hugenholtz, P. & Beiko, R. G. STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30**, 3123–3124 (2014).
89. Oksanen, J. *et al.* vegan: Community Ecology Package. R package version 2.5-7. <https://cran.r-project.org/web/packages/vegan/> (2020).
90. Kursu, M. B. & Rudnicki, W. R. Feature selection with the boruta package. *J. Stat. Softw.* **36**, 1–13 (2010).
91. Valero-Mora, P. M. ggplot2: Elegant graphics for data analysis. *J. Stat. Softw.* **35**, 1–3 (2010).

Acknowledgements

ALS would like to acknowledge the financial support of EMBO Installation Grants (IG project 3914), and POIR. 04.04.00-00-3E9C/17-00 carried out within the First TEAM programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund. The authors would like to acknowledge Aldona Nowicka for her help with SEM (Analytical Laboratory, Faculty of Chemistry, Maria Curie-Skłodowska University) and Magdalena Filoc for her assistance in palynological samples preparation (Faculty of Biology and Chemistry, University of Białystok).

Author contributions

O.I. wrote a publication; P.L. processed and analysed NGS data; O.I., P.L. analysed metagenome data, performed statistical tests and functional analyses; O.I., P.L. prepared and edited figures; A.L.S., O.I., M.S. conceptualised the project; A.L.S., M.S. co-wrote a publication; M.S. prepared Figs. 1 and 3; M.S. collected the core; M.S., O.I., A.L.S. subsampled the core; O.I. optimised *sedDNA* isolation protocol, O.I., M.S. isolated *sedDNA*, M.S. performed LOI and Cladocera analyses and constructed age-depth model; I.A.P. performed palynological analysis; I.B. performed

diatom analysis; M.S., M.H. performed identification of hemp fibres; M.G. performed 210Pb dating; N.K., M.K. performed fibre count analysis; J.P.A. provided computational support. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-12761-w>.

Correspondence and requests for materials should be addressed to A.L.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022