Trends in Microbiology

CellPress OPEN ACCESS

Review

Ancient environmental microbiomes and the cryosphere

Alexander D. Williams [1,2,*], Vivian W. Leung [2], Julian W. Tang [3,4], Nishimura Hidekazu[5], Nobuhiro Suzuki [6], Andrew C. Clarke [7], David A. Pearce [8,9,*], and Tommy Tsan-Yuk Lam [1,2,*]

In this review, we delineate the unique set of characteristics associated with cryosphere environments (namely, ice and permafrost) which present both challenges and opportunities for studying ancient environmental microbiomes (AEMs). In a field currently reliant on several assumptions, we discuss the theoretical and empirical feasibility of recovering microbial nucleic acids (NAs) from ice and permafrost with varying degrees of antiquity. We also summarize contamination control best practices and highlight considerations for the latest approaches, including shotgun metagenomics, and downstream bioinformatic authentication approaches. We review the adoption of existing software and provide an overview of more recently published programs, with reference to their suitability for AEM studies. Finally, we summarize outstanding challenges and likely future directions for AEM research.

Highlights

Ancient environmental microbiomes (AEMs) offer a window into past ecologies and pathogen evolution.

Perennially freezing environments provide some of the best preserved, oldest nucleic acid records.

In recent years, it has been possible to recover ancient DNA, and (to a lesser extent) ancient RNA from increasingly older sample material.

AEM studies are prone to contamination, and authentication steps raise many questions about robust dating of microbial signatures and the persistence of ancient, yet viable, microbes.

The cryosphere: a frontier for researching AEMs

The cryosphere collectively refers to environments where water exists in a frozen state, such as glaciers and ice sheets, seasonally frozen ground, and **permafrost** (see Glossary) (Figure 1). Comprising >9% of the Earth's surface (46–87 million km^2) year-round [1], the cryosphere is a vital part of the biosphere with far-reaching effects on climate change, and vice versa. However, it remains one of the most under-sampled and threatened components of the climate system [2]. For biological research, the inherent preservative properties of the cryosphere make it a uniquely valuable source of **ancient** nucleic acids (aNAs).

aNA research arose approximately 40 years ago, with efforts to retrieve ancient DNA (aDNA) sequences from diverse species, including extinct fauna dating back from thousands to tens of thousands of years ago [3,4]. Early aNA research also encompasses work on ancient RNA (aRNA) from plant material preserved for hundreds of years [5]. Studies in this field must contend with two major obstacles: **contamination** and sufficient target biomass. Methods have been developed to aid the **authentication** of aDNA, primarily, enhanced contamination control, analysis of postmortem **DNA damage patterns**, and phylogenetic placement (reviewed in [4,6]).

For microbiologists, advances in the field provided the opportunity to study the evolution of historical pathogens [7], including the genomic reconstruction of a medieval plague (*Yersinia pestis*) [8] and the 1918 pandemic influenza virus [9].

Cryosphere samples have enabled direct analysis of environmental antibiotic-resistance genes of ancient origin [10] and permafrost microbiome responses to thawing [11]. However, AEMs remain enigmatic, primarily due to challenges in (i) obtaining sufficient endogenous NA yield from low biomass environmental samples, (ii) authenticating complex communities where many members are poorly represented – or entirely absent – in **reference sequence** databases,

[1]Laboratory of Data Discovery for Health Limited (D^{2}4H), 12/F, Building 19W, 19 Science Park West Avenue, Hong Kong Science Park, Hong Kong Special Administrative Region of China
[2]State Key Laboratory of Emerging Infectious Diseases, School of Public Health, The University of Hong Kong, Hong Kong, SAR, China
[3]Respiratory Sciences, University of Leicester, Leicester, UK
[4]Clinical Microbiology, University Hospitals of Leicester, Leicester, UK
[5]Virus Research Center, Clinical Research Division, Sendai Medical Center, Sendai 983-8520, Japan
[6]Institute of Plant Science and Resources, Okayama University, Chuou 2-20-1, Kurashiki, Okayama 710-0046, Japan
[7]School of Biosciences, University of Nottingham, College Road, Sutton Bonington, LE12 5RD, UK
[8]Department of Applied Science, Faculty of Health and Life Sciences, Northumbria University at Newcastle, Newcastle, NE1 8ST, UK
[9]British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, UK

and (iii) identifying endogenous, yet viable, microbes that confound the use of DNA damage patterns as an absolute marker of recent *ex situ* contamination. Sample **cryostratigraphy** can provide insights into the relative age of entrapped biological materials, though the chronology of individual samples requires careful interpretation since the age of AEMs may not be the same as the surrounding material [12]. Despite these challenges, the cryosphere holds some of the best opportunities for studying AEMs.

Existing reviews on ancient microbes generally focus on pathogens and therefore often ignore the **environmental microbiome** and exclude such organisms as contaminants [7].

*Correspondence:
alexdwilliams@d24h.hk (A.D. Williams),
david.pearce@northumbria.ac.uk
(D.A. Pearce), and
ttylam@hku.hk (T.T.-Y. Lam).

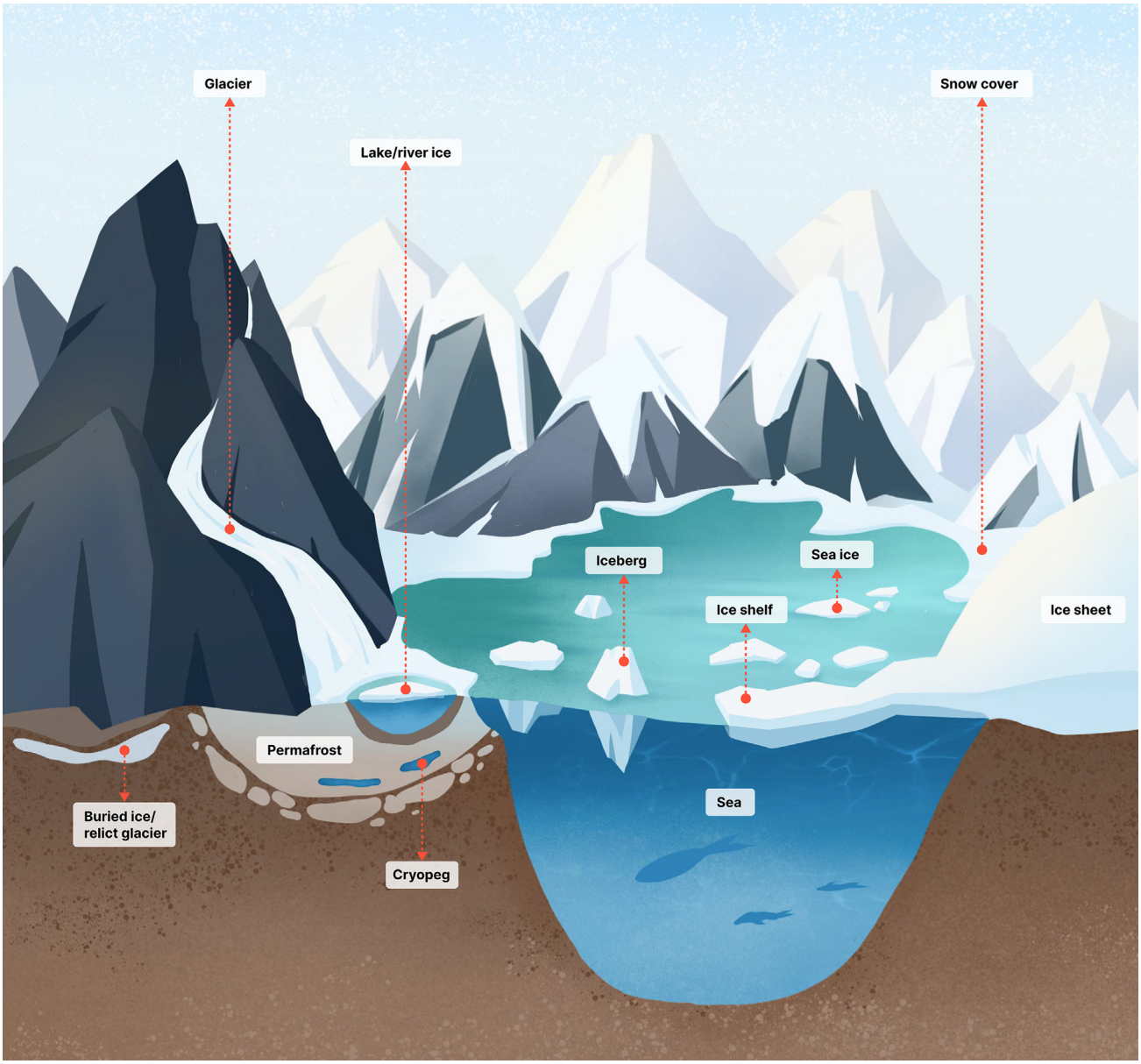

Figure 1. An overview of cryosphere landscape features. Figure adapted from Lemke *et al.* [109].

Although a recent paleogenomic review considered free-living environmental microbiomes, it was predominantly limited to sedimentary samples [13], and these can be retrieved from both cyotic and noncyrotic environments. We instead focus on the cryosphere, its significance for AEM research, specific contamination control procedures for the collection and processing of cryosphere samples, and the post-sequencing authentication of AEMs.

How old are cryosphere features?

While some cryospheric features – such as seasonal snow cover – are ephemeral, perennial ice formations and permafrost can record millennia. For example, in the Arctic, a continuous ice core was obtained from the Greenland glacier dating back 123 thousand years (kyr), with a depth of ~3 km [14].

Some of the oldest known ice on Earth is found in the Antarctic, where the buried remnants of ancient glacial ice are estimated to have been deposited over a million years ago [15–17]. However, buried glaciers do not provide a continuous archive; the oldest known chronologically continuous ice core was retrieved from Antarctica with a depth of 3.2 km and archived approximately 800 kyr [17]. Perennial ice formations can also occur in caves. The oldest known ice formation is the Scări oara Ice Cave, Romania, which dates back more than 1 kyr [18].

Distinct from other cryospheric features, permafrost is loosely defined as ground which has remained below 0°C for more than 2 years and is typically overlain by an active layer which undergoes seasonal cycles of freeze–thaw [12]. Permafrost is not a homogeneous landform. It can be sub-categorized based on a range of morphological characteristics, material composition (ice and organic matter content), process of formation, and distribution pattern (Figures 2 and 3). It may also contain cryotic water known as **cryopegs**. Depending on local conditions and depth, permafrost layers can be thousands to hundreds of thousands of years old (discussed by Abramov *et al.* [12]).

Environmental context must be carefully considered when estimating the age of AEMs since cryosphere samples can have complex histories. For example, permafrost may form through freezing at the time of deposition (i.e., permafrost age is equal to that of the incorporated material), or sometime afterwards (i.e., permafrost age is younger than that of the incorporated material). These scenarios are termed syngenetic and epigenetic, respectively [12] (note that the latter should not be confused with the unrelated concept of epigenetics in molecular biology). Cryostratigraphy is further complicated by thawing and re-freezing events (thaw unconformities) [19], as well as changes in deposition rates. Freeze–thaw events have direct implications for microbes: in every thaw, some may reproliferate and evolve more rapidly, and in every freeze, some may die, altering the microbiome composition. Likewise, temporal complexities can feature in ice (Figure 4). More generally, stratigraphy can be affected by post-depositional disturbance, a subject which is being addressed in the field of ancient sedimentary DNA using *in silico* compositional methods [20]. Researchers unfamiliar with these approaches may therefore misestimate the age of archived microbial communities. However, these complex chronologies are not merely nuisances but valuable information which helps to properly interpret and date AEMs. It is therefore advisable to consult geocryologists when investigating cryospheric AEMs.

The temporal limit of microbial biosignatures: how far back can we go?

The consistent, low temperatures of permafrost and perennial ice are thought to preserve some of the oldest recoverable NAs and viable microorganisms [21,22]. However, it is critical to consider the plausibility of microbial survival and NA preservation. Reports of amplifiable bacterial DNA and cultivable microbes from samples dated at more than 1 million years (myr) old have

Glossary

Ancient: the definition of 'ancient' is subjective and often context-dependent. For the purposes of this review, we use the term to refer to nucleic acids that have been archived (i.e., frozen in evolutionary time) in the environment for more than a century. These may be derived from dormant microbes, viruses, deceased cells still intact, free-floating genetic material from lysed organisms and virus particles, and other nucleic acids such as plasmids.

Ancient nucleic acid authentication: the processes by which nucleic acids are verified as being ancient and not derived from contaminants (modern or otherwise). Depending on sample context and study objectives, this may be comprehensively or partially possible (e.g., it may be possible to rule out modern contaminants, but not historical contaminants).

Aspartic racemization analysis: a nonenzymatic chemical reaction that causes an increase in the D:L aspartic acid enantiomer ratio over time, which can impair cellular viability. Metabolically active microbes can revert this process; hence a low D:L aspartic acid ratio can be used as a proxy for cellular activity [32].

Contamination: there are two main types of contamination: inherent and introduced. Inherent contamination depends on the study objective. For example, a study considering the microbiome of a permafrost-preserved cadaver would describe microorganisms from the surrounding permafrost microbiome as contaminants, whereas a study concerning AEMs would not. Introduced contamination includes nucleic acids and microorganisms that are introduced during the collection and processing of ancient samples.

Cryopeg: an unfrozen water feature in permafrost or ice that is perennially <0°C. Dissolved solids in the water result in a freezing point depression, enabling water to remain in liquid form. They are typically highly saline environments.

Cryostratigraphy: the study of frozen layers of the Earth, including the history of their formation.

DNA damage patterns: in the context of ancient nucleic acid research, these typically relate to hydrolytic depurination-induced fragmentation and cytosine deamination patterns. The accumulation of such damage is routinely used as an indicator of age.

been received with sustained skepticism due to validation shortcomings and irreproducibility [6,23–25].

Without repair, DNA accumulates damage over time due to hydrolysis, oxidation, and crosslinking [3,4]. Consequently, aDNA is often highly fragmented and <100 bps in length (reviewed in [4]). Until recently, the process of fragmentation was thought to present an approximate range for aDNA analysis (depending on preservation conditions) anywhere between 100 and 780 kyr [3,4]. However, some of the oldest permafrost-associated DNA (that is broadly accepted as authentic) originates from a mammoth dating back between 1 and 2 myr [26]. This was further eclipsed by ancient environmental DNA obtained from 2.4 myr old permafrost sediment collected from the Kap København Formation in Northern Greenland [27]. The authors suggested that the high DNA adsorption capacity of clay and smectite minerals found at the study site were likely responsible for the exceptional DNA preservation [27].

It is noteworthy that the findings of both studies focus on eukaryotes and megafauna which are either extinct or have well documented phylogenetic histories, and accordingly they negate some authentication issues associated with ancient microbial studies (e.g., exposure to pervasive contemporary microbial contaminants, even in dedicated aDNA laboratories). However, the team who worked on the deep- time Kap København metagenomes has also published a pre-print exploring the associated microbial and viral communities with accompanying DNA damage pattern analysis [28]. Additionally, Courtin *et al*. [29] presented several archaeal, bacterial, and fungal signatures showing aDNA damage patterns in permafrost sediments from the Batagay megaslump (22–650 kyr). Collectively, these data suggest that, given suitable conditions, it is feasible to recover and authenticate highly fragmented, postmortem DNA from at least 2.4 myr.

Old, but still kicking?

Unlike the case of mammoths, there is continuing debate surrounding the longevity of viable microorganisms, particularly **psychrophiles** endogenous to cryosphere samples. It is reasonable to assume that a viable microorganism has intact genomic material. This raises the question: does the retrieval of largely intact nucleic material from ancient samples exclusively correspond to microbial contamination? To be clear, evidence of viable microbes in deep-time [>1 million years ago (mya)] environmental samples remains a contentious subject for reasons which we now discuss.

Endospores are not metabolically active, and they lack active DNA repair mechanisms, yet they can be confidently revived after ≥100 years [30]. Although they have some resistance to DNA damage through small acid-soluble proteins [31], ancient endospores are otherwise exposed to DNA-degrading forces over longer timescales [32]. Claims of viable endospores retrieved from 250 myr old salt crystals [33] and 40 myr old amber [34], remain highly controversial due to contamination concerns [21]. Using *Geobacillus stearothermophilus* as a model organism, Liang *et al.* [32] conducted **aspartic racemization** experiments which indicated that the survival of permanently dormant endospores could be extrapolated up to only 4–11 kyr. The theoretical survival of ancient bacteria may thus be constrained by the necessity of DNA repair.

In an effort to address this question, Johnson *et al.* [22] employed a cultivation-free approach and strict contamination control procedures to evaluate bacterial survival in ancient permafrost samples dating from 7 thousand years ago (kya) to 1 mya. In this study, the authors used successful amplification of ≥4 kb of ribosomal bacterial DNA following uracil-N-glycosylase treatment to infer the survival of bacteria. This was based on the assumption that relict DNA from dead organisms of equivalent age would fail to produce such large amplifiable fragments due to any damaged

Environmental microbiome: for the purposes of this review, these are microbes detected in the substrate being studied (e.g., ice), rather than within host material (e.g., animal carcasses, bones), or waste materials (e.g., preserved feces). This is to reflect the distinct challenges related to the authentication of these microorganisms. However, we acknowledge that waste material and carcasses are still a component of the environment in a broader sense.

Permafrost: ground which is perennially <0°C. It may contain other cryotic features such as ice wedges and cryopegs. Note that dry permafrost, with little to no ice content, has been described in Antarctica.

Psychrophile: an organism with minimum, optimum, and maximum growth temperatures that are below 0°C, 15°C, and 20°C, respectively (Morita [108]). Psychrotolerant organisms represent those that can survive at low temperatures, but less successfully than psychrophiles.

Reference sequence: a nucleotide sequence selected as the closest available match to a set of closely related target sequences recovered from a sample. In the context of ancient nucleic acid research, a reference sequence is often required for authentication methods such as DNA damage pattern analysis. The reference sequence has a direct impact on the interpretability of authentication methods, so selecting a suitable reference is important for analysis.

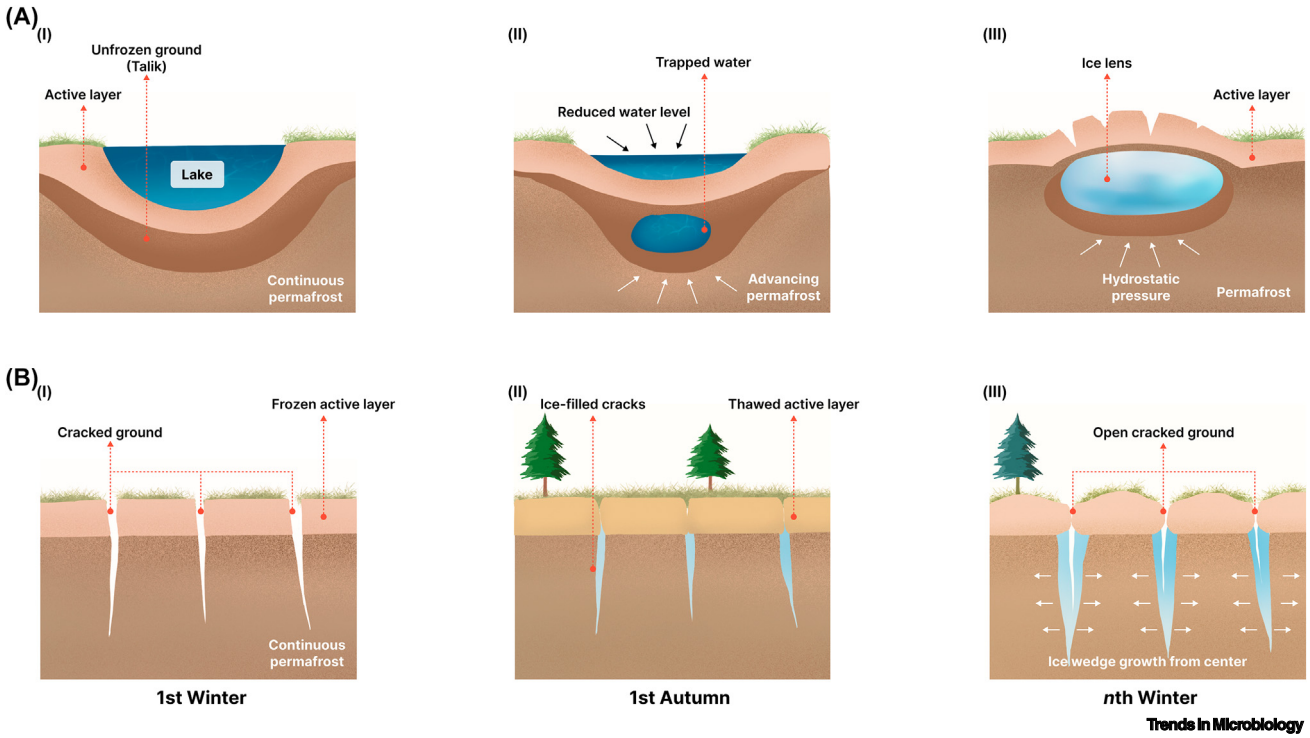

Figure 2. Overview of continuous permafrost feature development processes. (A) Closed-system pingo. (I) Continuous permafrost encloses unfrozen ground beneath a lake (talik). (II) Over time, the lake is filled in through sedimentation, and the permafrost encroaches into the talik; any residual water is subsequently forced upwards under hydrostatic pressure. (III) A blister develops, which subsequently freezes to form a mature pingo. Based on the description in Rowley *et al*. [110]. (B) Ice wedges or ice veins. (I) Ice accumulates in surface cracks formed by thermal contraction during annual active layer freezing. (II) Ice within the permafrost is protected from thawing during the summer and autumn. (III) The gradual lateral expansion of an active ice wedge is facilitated by successive cycles of freeze and thaw, with the youngest ice being present at its center. Figure adapted from Lachenbruch [111].

DNA being targeted by uracil-*N*-glycosylase. Using this method, Johnson *et al.* [22] provided evidence of bacterial DNA repair in samples up to ~600 kyr old. Although a promising finding, further work on this subject would be of great benefit to the field.

Johnson *et al.* [22] also showed that survival of low GC content, spore-forming Gram-positive bacteria was inferior to that of Gram-negative and high GC content, nonspore-forming Gram-positive bacteria. Spore-forming bacteria could not be recovered from samples dating back between 400 and 600 kya. By contrast, metagenomic studies by Mackelprang *et al*. [35] and Liang *et al*. [36] reported spore-forming bacteria were dominant in 33 kyr old Alaskan tunnel permafrost and ~1.1 myr old Yedoma permafrost, respectively. However, based on aspartic acid analyses, Liang *et al.* [36] concluded that the sporulating bacteria must have undergone periodic spurts of metabolic activity. While both studies employed live/dead cell-staining techniques [35,36], no attempt was made to cultivate, sequence, and phylogenetically place potential survivor microbes. The mechanisms and limit of endospore survival remain subjects of debate, although ongoing reporting from a 500-year endospore laboratory experiment aims to help clarify matters [37].

More recent studies have continued to suggest that viable bacteria can be cultivated from permafrost older than 3 myr [38], far exceeding the threshold implied by Johnson *et al.* [22]. Bidle *et al.* [15] aimed to revive ancient bacteria using ~1 l of nutrient-amended ice-melt from Mullins Valley (100 kyr) and Beacon Valley (purportedly 8 myr, although this has been questioned [39]). The ice underwent exterior decontamination with autoclaved nutrient-supplemented MilliQ water as

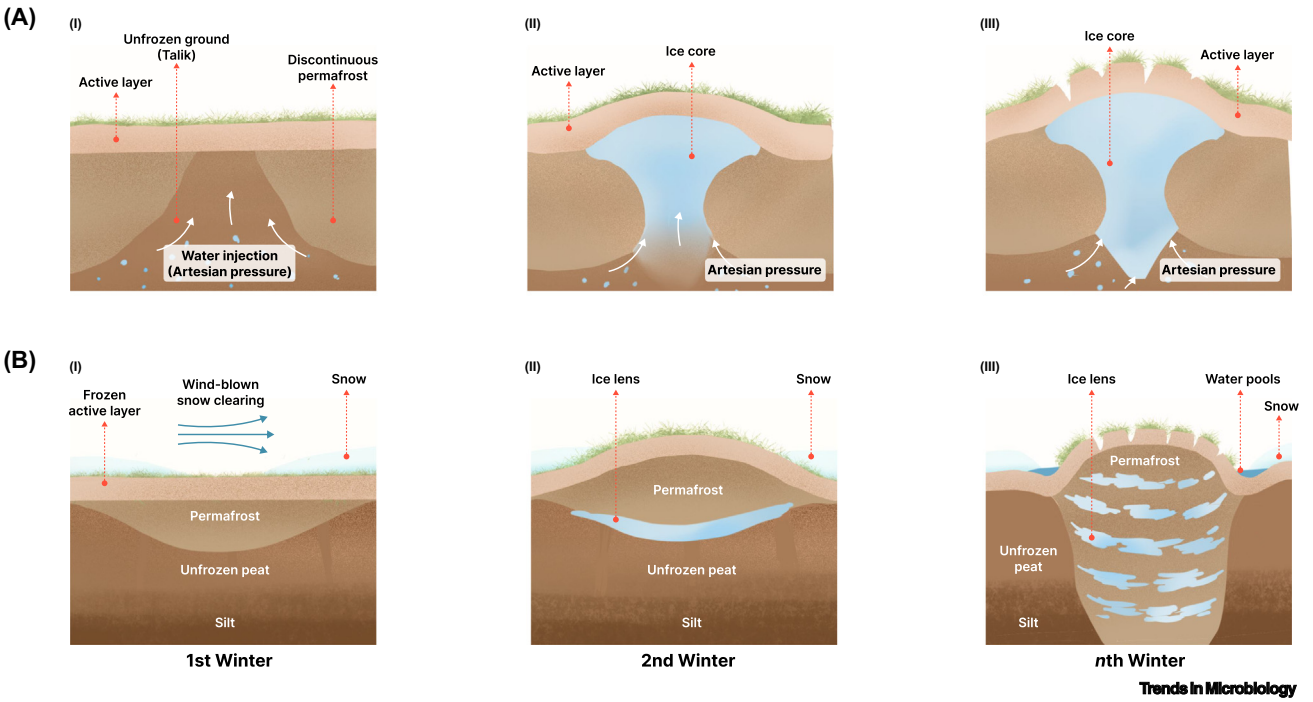

Figure 3. Overview of discontinuous permafrost feature development processes. (A) Open-system pingos. (I) Water is drawn up under artesian pressure from unfrozen ground (talik) lying beneath a thin layer of discontinuous permafrost. (II) Upon freezing, the water forms an ice lens which pushes up the active layer, creating a mound. (III) As the pingo matures, the ice core continues to grow, and the overlying active layer begins to crack. Based on the description in Rowley *et al.* [110]. (B) Palsa formation. (I) The removal of snow permits freezing to occur more deeply within the peat profile. Consequently, at the height of annual thawing, a thin layer of localized frozen ground remains intact. Water accumulates within the thawed peat beneath the frozen ground, which pushes the frozen ground upwards. (II) The layer of water then freezes during the following winter, further elevating the palsa. The thermal insulative properties of the raised, dried, peat protect the central ice core during summer. (III) The process of ice lens formation and palsa growth continues until the frozen core reaches silt, whereupon segregation ice is readily formed. As the palsa matures, cracks may form upon the surface, leading to the sloughing of surface peat into the surrounding bog. Figure adapted from Seppälä [112].

incubation controls. Interestingly, while both samples were able to generate growth, bacteria could be isolated only on solid medium using the ~100 kyr old sample. Bidle *et al.* [15] suggest that this is due to functional deterioration of the bacteria in the oldest sample. These findings have yet to be independently replicated, something regarded as a key criterion for such studies. Nonetheless, it has been acknowledged that a sparse distribution of cells may hinder independent validation [21]. An inability to recover microorganisms from samples of similar age may also be under-reported due to the lack of interest in negative results. Another issue which remains unresolved, is whether existing evidence of DNA repair in ancient samples primarily relates to long-dormant cells, replicating microbes, or actively growing microbes.

Although the accepted limits of preservation for aDNA that can be sequenced have undergone rapid and dramatic revision in recent years, robustly demonstrating the survival of ancient microbes is a separate and exceptionally challenging issue due to the possibility of contamination. Therefore, the subject of ancient, viable microbes must be treated with extreme caution until our understanding of microbial longevity is better grounded in foundational experimental research.

Recovery of ancient viruses, viable and trace genetic material

Amoeba-infecting giant DNA viruses reactivated from 30 kyr old Siberian permafrost purportedly stand among the oldest virions retaining infective capacity [40]. Assuming that samples were

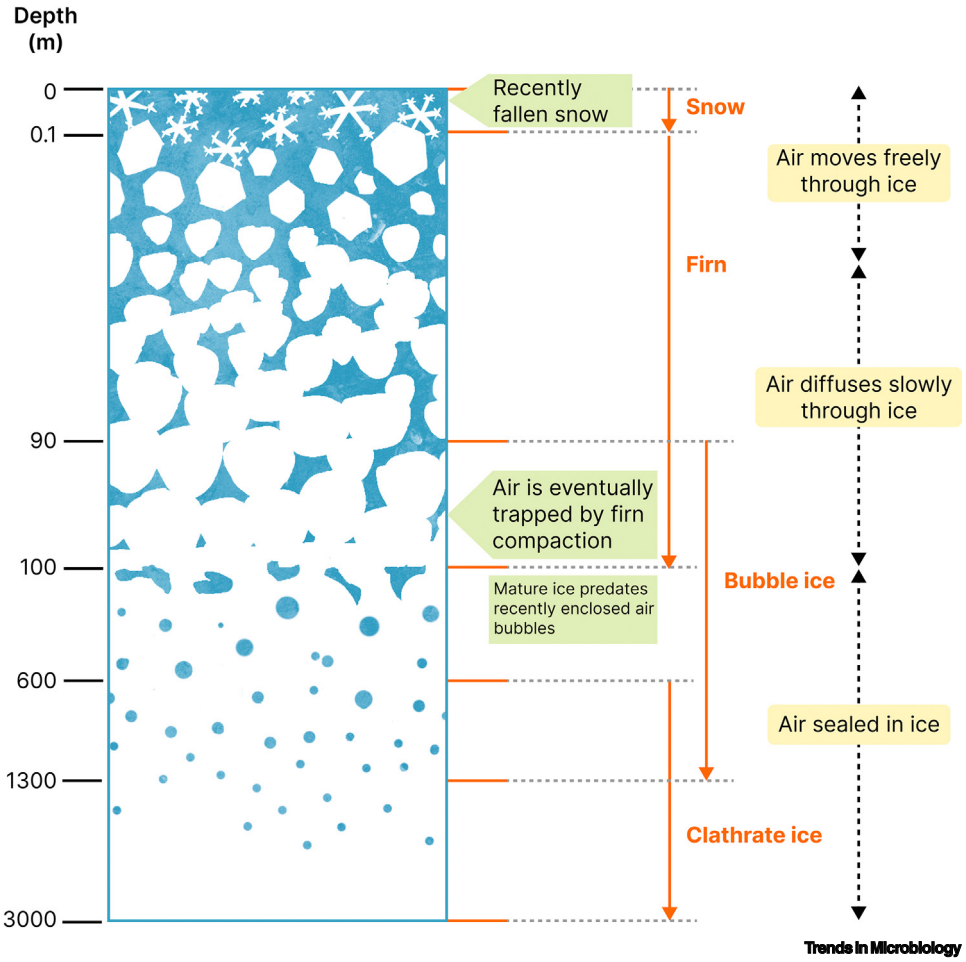

Figure 4. Cross-section of an ice core (not to scale), showing how – dependent on depth and structure – air trapped within ice may be of more recent origin than the surrounding ice. Accordingly, while the age of both ice and trapped material increases with depth (excepting unconformities) the depth of archived microbes may not directly correspond with the age of surrounding ice. Indicated depths are only general approximations. Figure adapted from Raynaud *et al*. [113].

appropriately handled, subsequent isolations performed by the same research group indicate that ancient permafrost as old as ~48 kyr is replete with such viruses [41]. In another study, Ng *et al*. [42] recovered a complete novel DNA virus genome from 700-year-old caribou feces preserved in permafrost and demonstrated its ability to infect plants. Holmes [43] suggested that Ng *et al.* [42] presented a more robust example of an ancient viral genome than Legendre *et al.* [40], in part, because the findings were replicated in another laboratory. Given that viral particles (virions) are inert outside of the host cell except for any innate protective properties of their particle structure (e.g., capsids), it might be expected that virions would be less resilient than bacterial endospores (~4–11 kyr without periods of reactivation), as they lack both DNA repair functionality and the protective binding proteins of endospores.

There is a paucity of thoroughly validated ancient viromes which approach the age of the oldest reported megafaunal sequences. However, Fernandez-Guerra *et al.* [28] have reported evidence (in pre-print) of viral sequences with DNA damage from ~2 million-year-old Kap København samples. Additionally, Zhong *et al*. [44] describe viromes from Tibetan glacial ice cores dating to a

much later period of >41 kya. The latter study employed a suite of negative control measures including sampling of the laboratory air, parallel processing of mock ice cores, as well as extraction blanks. The work was also carried out in a cold room dedicated to processing low-biomass samples and was UV irradiated before use. However, it does not appear that DNA damage patterns were evaluated.

If viable bacteria do persist for ~600 kyr they may possibly also serve as hosts for bacteriophages, thereby extending the viable recovery of bacteriophages. Moreover, giant DNA viruses have been shown to have genes associated with DNA repair [45] that could extend viability over much longer timescales, assuming that they were intermittently reactivated. Nonetheless, vestigial viral sequences in bacteria (relict prophage) and eukaryotes (e.g., endogenous retroviruses), could be misidentified as elements of contemporaneous viral genomes, particularly if the contextualizing flanking sequences were lost due to degradation.

Due to its comparative instability, RNA is assumed to degrade faster than DNA [21], and is therefore even more difficult to recover from ancient samples. While aRNA and aRNA viruses have received considerably less research attention [5,46] than aDNA, a recent commentary highlighted that the higher copy number of cellular RNA relative to DNA, and rapid inactivation of RNases by freezing, may enable preservation over thousands of years [47].

Perhaps the most notable example of historic viral RNA recovery was from human lung tissue obtained from a victim of the 1918–1919 flu pandemic interred in Alaskan permafrost [9]. RT-PCR had been previously used to amplify tobamoviruses (positive-sense single-stranded RNA viruses) from Greenland glacial ice dating back 14 kyr [48]. However, the findings are disputed as the study used a highly sensitive amplification technique and samples were not handled in a dedicated clean-room facility [5]. Nonetheless, a more recent study used extensive contamination control procedures to obtain an RNA transcriptome from a wolf preserved in 14 kyr old permafrost [46], suggesting that such a long period of preservation is plausible. To date, the oldest viral RNA genome belongs to a double-stranded (ds)RNA plant virus derived from 1 kyr old maize cobs [49]. It is notable that viral aRNA has most frequently been obtained from preserved plant and animal material [50]. Nevertheless, these studies indicate that RNA may be more persistent than generally assumed. To this end, Scheel *et al.* [51] described the recovery of total RNA from a 1 m deep permafrost core dating back 26.5 kyr; however the description of contamination-control procedures during core retrieval and sample processing is limited. More work needs to be done to uncover the dynamics of RNA survival, as well as the mechanisms which could aid the persistence of ancient viral RNA.

Contamination: reduction and control

Contaminants are a persistent challenge in molecular analyses, particularly for aNA-based research. Generic methods for preventing and removing contamination in environmental samples have been extensively reviewed [6,21]. We summarize in Table 1 some interventions adopted or recommended for ice and permafrost samples during collection, transport, and laboratory processing. While some studies focus only on post-collection decontamination [52], the wider aNA community typically emphasizes mitigating risk in all phases of sample handling, whether in the field or laboratory [21,53,54].

Authentication and bioinformatic advances

Although contamination control is critical to the successful recovery of authentic AEM, it remains impossible to guarantee complete elimination of contamination [6,55]. Additional *post hoc* steps

Table 1. Summarizing contamination-reduction and control measures

Work stage	Proposed intervention	Purpose	Refs and notes
Collection	Use of UV, acid, or bleach to clean drilling/collection equipment	Surface decontamination of collection equipment	[54]
	Swab sampling equipment	Monitoring of drilling equipment contaminants	[54]
	Wear nitrile gloves, coveralls and facemasks in field	Minimize sample exposure to human microbiota	[21,54]
	Tracer application to drilling equipment and/or sample exterior	Monitoring ingress of surface contaminants and validation of surface decontamination steps	Fluorescent beads [114–117] and Biological markers [65,115,117]. The deepest penetration of tracers reported in permafrost cores is 17 mm [115]
	Immediately seal in a clean bag – ideally NA free	Minimize environmental exposure during transit to laboratory	[36,115,118] We recommend gamma-irradiated sealable bags (ideally double-bagged)
Transport and storage	Packing to prevent cracking of cores	Minimize ingress of surface contaminants	Cracking is a particular concern for fragile ice cores [53,116]
	Maintain suitably cold temperatures during transport and storage (ideally –80°C)	Minimize contaminant growth and changes to community structure and aid preservation of aNAs	Cold-adapted bacteria are known to grow and show metabolic activity at temperatures as low as –20°C [119,120]
Laboratory	Physical removal of outer surfaces	Exterior decontamination of samples	Cutting and scraping [65,114,116–118].
	Chemical rinses of outer surfaces (often used in tandem with physical removal)		These have included ~5% sodium hypochlorite (dilute commercial bleach) [65,121], 95% ethanol [52,117,118,121] (though note this is not suitable for removal of NAs [21]), and sterile, ultrapure water [52,117,118,121]. Exterior surface melting has also been used for ice cores [117]
	Work from oldest to most recent sample sections. Thoroughly decontaminate workspace between samples	Prevent cross-contamination, with attention to degradation gradients	It has been shown that DNA yield generally decreases with depth [22,24,102] which is often, but not always, a proxy for age
	Removal of chemical contaminants	Minimize interference with workflow chemistry (e.g., sequencing)	Commercial kits and custom extraction methods designed for samples high in humic acids have been used to process peat permafrost [122,123]
	Processing samples in an aNA laboratory	High-level contamination control	Fulton [124] describes the requirements of an aNA laboratory in detail

are therefore recommended to detect and account for potential contamination as well as to authenticate putative ancient sequences.

DNA damage patterns and ancient metagenomic pipelines

The accumulation of postmortem DNA damage produces predictable time-correlative patterns that have become a central pillar of aDNA authentication methods [4,56–58]. Briefly, these DNA damage patterns are associated with hydrolytic depurination (fragmentation), resulting in increased deamination of exposed cytosine residues, which manifest as C→T misincorporations at the beginning of DNA read sequences, and G→A misincorporations at the end of reads in double-stranded libraries [4,59]. These miscoding lesions are the basis for most *in silico* aDNA damage authentication software, although other forms of DNA damage, such as cross-linking also occur [60]. The frequency of C→T misincorporations has been shown to correlate with age in plant DNA obtained from herbaria collections [61]. Similarly, Mitchell *et al*. [62] reported a correlation between age and fragmentation and crosslinking in total DNA from Siberian permafrost samples. Another study indicated C→T transitions are thermally correlated [63], which holds particular significance for cryosphere samples. To date, it appears that there has been no attempt to calibrate DNA damage over time for AEMs using bioinformatic techniques.

Calibrating a definitive DNA damage 'clock' is subject to several complications, including heterogeneous and nonlinear accumulation of damage, the presence of contaminants also with DNA damage, and damage caused during sample processing. Geographically and temporally specific environmental factors can alter rates of damage. Early-stage surface exposure to UV and freeze–thaw events may also accelerate damage. The presence of nuclease-inhibiting minerals (e.g., montmorillonite clays [64]) and the specific structural or metabolic properties of microbes themselves also affect damage accumulation rates.

As shown in Table 1, bleach is commonly used as a sample surface decontaminant and has been advocated in a recent comparison study concentrating on permafrost cores [65]. Some have raised concerns that bleach may generate damage patterns in modern contaminant DNA redolent of aDNA, potentially confounding damage analysis for authenticity [66], although this issue has yet to be more widely explored.

Despite the aforementioned caveats, DNA damage patterns remain a critical authentication criterion for aDNA. This has led to the development of protocols that selectively enrich for damaged DNA [67,68]. Such approaches discard longer NAs associated with endogenous, viable microbes; the applicability of this method therefore depends on the research objectives.

Several programs are currently available for the *in silico* detection of DNA damage patterns (Table 2). These authentication methods consider misincorporation patterns and require suitable reference sequences. The majority of these programs are genome-centric and have not been designed with complex mixed communities in mind.

Borry *et al*. [69] contended that programs operating at the read-level can impair assembly-led AEM investigations due to data loss. *PyDamage* was therefore specifically devised for authenticating metagenomic contigs [69], and the developers claim it is computationally faster at scale than other programs such as *DamageProfiler* [70] and *HOPS* (Heuristic Operations for Pathogen Screening) [71]. Although *mapDamage 2.0* can be applied directly to contigs, it also suffers from poor scalability with high throughout (HT) next generation sequencing (NGS) [69,70]. However,

Table 2. Summary of postmortem (DNA) damage pattern programs, packages, and authentication pipelines

Name	Capabilities	Refs
mapDamage 2.0	Damage pattern analysis for reads and contigs	[125]
AuthetiCT	Single-stranded DNA library-only compatible damage pattern analysis (model-based filtering)	[77]
aRchaic	Damage pattern analysis, with 'grade of membership' model, facilitating multisample comparisons. Not geared towards metagenomic analyses	[126]
DamageProfiler	Damage pattern and fragment length distribution analysis	[70]
MetaDamage	Carries out damage pattern analysis and is designed to automate identification of references for metagenomic reads using BLAST	[127]
PyDamage	Contig-based damage pattern profiling with log-likelihood ratio testing for authenticity	[69]
PMDtools	Likelihood framework using damage patterns, base quality scores and biological polymorphisms. Filter reads based on a postmortem damage score	[128]
MetaDMG	Damage pattern and fragmentation analysis for metagenomic samples. Adopts a beta-binomial damage model purportedly robust for low-coverage data	[76]
HOPS	Metagenomic, pathogen-centric damage pattern, edit distance, coverage evenness and fragment length distribution analysis	[75]
aMeta	Metagenomic pipeline, incorporating many features of HOPS, adapted to enhance scalability and application to environmental microorganisms	[71]

contig-based approaches rely on obtaining good-quality assemblies, something that may be challenging for low biomass samples containing complex communities.

HOPS incorporates aDNA authentication steps within a larger pipeline which also classifies and validates bacterial taxonomy using *MaltExtract*, an adaptation of the alignment-based taxonomic classifier *MALT* [71]. Eisenhofer and Weyrich [72] showed that *MALT* is reasonably effective at handling AEMs, although Cárdenas *et al.* [73] reported errors when using viral databases and instead recommended *Centrifuge* [74] for taxonomic assignment. It is unclear whether these issues have been resolved. *HOPS* employs additional edit distance calculations and reference coverage distribution to validate taxonomic assignments, in line with previous recommendations for analyzing ancient microbiomes [57,58]. However, *HOPS* targets only user-specified taxa, limiting applicability in whole community AEM studies where community composition is often unknown in advance, and this is often the case with cryosphere samples.

aMeta seeks to address the limitations of HOPS by combining *k*mer-based (*KrakenUniq*) and alignment-based (*MALT-MaltExtract*) tools [75]. Briefly, *aMeta* uses *KrakenUniq* to perform initial classification and pre-filtering, enabling the construction of a sample-specific *MALT* database which can then be used to carry out *MALT* alignments and appropriate aDNA authentication via *MaltExtract* [75]. Unlike *PyDamage*, *aMeta* does not require an assembly-based approach. *MetaDMG* is another program developed to assess DNA damage patterns in metagenomic data. It uses a beta-binomial damage model which the developers assert remains robust even with low coverage data typical to AEM studies [76].

AuthentiCT is designed to process single-stranded libraries, which exhibit different properties that provide some advantages over traditional double-stranded libraries [77]. Notably, single-strand based approaches can increase library yields when processing highly degraded samples since; unlike double-stranded methods, they are capable of capturing information from DNA molecules with breaks on both strands [78]. Moreover, single-stranded libraries provide full-length sequences and preserve strand orientation [79]. Although most damage-mapping software is compatible with single-stranded libraries, the recovery of full-length sequences underpins a unique feature of *AuthentiCT*: deamination is considered at all positions, rather than at only sequence termini. However, it has been noted that single-stranded library preparations are most effective at targeting shorter fragments (<120 bp); given the exceptional preservation afforded by cryosphere environments, double-stranded libraries may remain a more suitable method for permafrost and ice core samples [78,79].

Reliance on reference sequences for damage pattern analysis can be problematic for AEMs, since environmental microbiomes often harbor many uncharacterized taxa [80,81]. Gancz *et al.* [82] introduced a reference-free damage pattern algorithm called *ChangePoint*, which was used to authenticate ancient dental calculi microbiomes alongside *MapDamage 2.0*. *ChangePoint* has also been used by Pérez *et al.* [56] to evaluate soil core microbiomes. However, the source code is not currently publicly available, and the only descriptions of the software are found in a dissertation (Y. Liu, PhD thesis, University of Adelaide, 2019).

Damage-based authentication of RNA has yet to be fully explored, possibly due to its relative instability [5], compounded by difficulties in recovering aNA in general. Among the few studies focused on recovering aRNA, Smith *et al*. [83] showed systematic damage patterns in barley stripe mosaic virus sequences obtained from a ~750-year-old barley grain. Differing degradation dynamics of DNA and of RNA may yield differing damage patterns. Consequently, damage-based authentication methods tailored to aDNA may not be suitable for aRNA. Smith *et al.* [46]

describe a 14.3 kyr old Pleistocene liver tissue transcriptome obtained from a canid preserved in permafrost but were unable to detect typical aDNA damage patterns. However, they were able to detect classical damage patterns in recovered DNA from the same tissue and, critically, showed that the transcriptomic data from the liver tissue was similar to tissue-specific modern canid transcriptomes. Ng *et al.* [42] did not demonstrate NA damage in RNA and DNA viruses detected in ice-bound caribou feces; however, their findings were independently replicated by another laboratory. Holmes [43] suggests that rapid freezing and the protective properties of viral capsids may explain the lack of NA damage detection; however, this hypothesis has yet to be tested.

Phylogenetic placement and molecular dating

Although damage patterns are a critical metric in authenticating aDNA, they do not, on their own, necessarily demonstrate the ancient origin of microbes [57]. For example, differences in cell structure/sporulation could affect observed rates of NA degradation. Phylogenetic placement and evolutionary rate testing of putatively ancient sequences have been proposed as additional authentication tools for ancient microbes and viruses [6,7,57,84]. The recovery of viable bacteria from ancient material is often contentious, especially when sequences show minimal divergence from modern reference sequences. For example, Eisenhofer *et al.* [23] challenged the authenticity of an *Enterococcus faecium* isolate obtained from 28 kyr old woolly mammoth intestinal tissue [85]. They contend that the reported findings are significantly undermined by the absence of aDNA contamination control procedures and highlight the genomic sequence has suspiciously high multilocus sequence typing (MLST)-based similarity to modern *E. faecium* strains [23]. When screening samples for clinically relevant species, it is also important to use competitive mapping to avoid erroneously assigning pathogenic taxa to closely related but nonpathogenic and environmentally abundant counterparts [58]. More recently, van Bergeijk *et al*. [86] reported actinomycete isolates from the same mammoth carcass, and while the authors showed that sequences are divergent from modern strains, description of contamination controls is limited.

Unfortunately, significant divergence from known strains is no guarantee of antiquity since it may simply reflect deficiencies in our knowledge of contemporary organisms. For example, little is known about giant viruses, whose existence was recognized only relatively recently [87]. The *Pithovirus* first described by Legendre *et al.* [40] was thought to represent an extinct viral lineage until Levasseur *et al*. [88] found a similar virus in wastewater. Duchêne and Holmes [89] also underscored the difficulty in estimating rates of evolutionary change for giant viruses for which few genomes are available.

Reviewing obstacles to dating viral sequences, de-Dios *et al*. [90] caution that purifying selection and site saturation can confound divergence estimations in temporal signal and root-to-tip analyses as previously described in Duchêne *et al*. [91] and Duchêne *et al.* [7]. These factors can undermine strict molecular clock-based analyses that assume a fixed rate of change over time, though they may be mitigated by 'relaxed clocks'. However, Duchêne *et al.* [92] show that while many bacteria exhibit acceptably robust genome-wide molecular clocks when provided with data spanning enough time to overcome rate variation, others do not, despite the inclusion of ancient samples. For example, *Y. pestis* has been shown to have an unstable molecular clock [93].

Much attention has been placed on obligate pathogens, which are expected to exhibit divergence from their modern counterparts [23]. However, this assumption of divergence may not be universally applicable. For instance, 'ultra-conserved' bacteriophage have recently been described in 1.3 kyr old human paleofeces [94]. AEMs also comprise many free-living organisms which are

not compelled to coevolve with a host. Indeed, Vreeland and Rosenzweig [95] invoke this notion to argue that not all authentic AEM sequences may diverge greatly from contemporary equivalents, particularly in largely stable environments.

Viable environmental microbes endogenous to ancient samples may be either dormant (nonreplicating) or actively replicating, even if slowly. Investigations into deep-subsurface microbes have demonstrated remarkably long generation times and indicated generally slow rates of evolution [96,97]. Nonetheless, replicating microbes are subject to evolutionary processes; therefore, their molecular age may be considerably less than indicated by the surrounding environmental material.

However, in cases where viable, nonreplicating microbes have been maintained for millennia, genomic material may exhibit relatively little or no damage, incongruous with the axiom that aDNA exhibit damage patterns, and are consequently dismissed by current authentication methods.

In silico contamination removal

Contamination is inevitable, even with rigorous decontamination techniques and dedicated aNA laboratories [6,58,98]. Laboratory reagents themselves have been shown to be a potential source of contaminants [99], sometimes referred to as the 'kitome'. It is therefore essential that appropriate negative controls are not only collected but also analyzed and published alongside AEM findings [58,84]. Efforts have also been made to maintain lists of common laboratory contaminants [55,98]. Software such as *SourceTracker* [100] can be used to identify the relative contribution of negative control sequences to sample libraries and has been used extensively in aDNA studies [4,55,57,98]. One conservative approach is to subtract all organisms present within negative controls. This is perhaps more suitable for studies investigating obligate pathogens than AEMs, since documented or close relatives of contaminant organisms may be truly intrinsic to the sampled environment [55], and removing those results can artificially reduce the observed biodiversity. Knowlton *et al*. [101] used this approach with Greenland ice cores and retained only 33 unique sequences. In the case of ancient, deceased microorganisms, DNA damage, fragment size, and edit distance may be alternatively used to screen sequences. (Note that this excludes only modern contamination; historical contamination may also exhibit damage patterns.) Finally, Key *et al.* [58] underscored that the effectiveness of negative control libraries depends on having sufficient sequencing depth to amplify most fragments within, such that even trace contaminants can be identified. Compared with those of biological samples, negative control libraries should also exhibit reduced complexity [58].

The adoption of authentication steps remains inconsistent across AEM studies. Some studies do not explicitly state use of controls [51], do not always attempt sequencing of negative controls [102], or do not actively report their contents in publications [103]; others incorporate negative controls in analyses [36].

Considerations for authenticating amplicon libraries

When working with low-biomass samples, it may be necessary to resort to amplification-based target enrichment prior to NGS. While powerful, amplification steps impart limitations. First, PCR can amplify trace contaminants that overwhelm endogenous signals [21,54]. It is therefore essential to run no-template controls and carry out amplification steps in a separate laboratory (indeed, it is a strict requirement that aNA laboratories are amplicon-free workspaces). Second, damage indicative of aDNA is predisposed to chimera formation during amplification [104]. Third, uneven template amplification – due either to randomness (e.g., jackpotting) or template

composition (e.g., GC content) – can distort the observed relative composition of microbial communities [105,106] and reduce detectable biodiversity. This amplification bias increases with more PCR cycles. To mitigate bias if extensive amplification is needed, unique molecular identifiers (UMIs) and double-indexing can be incorporated in multiple library preparation steps.

It has been suggested that authentic aDNA amplification products should show an inverse relationship between amplification strength and length, since short fragments should predominate [21,54]. Willerslev *et al.* [21] state this should be true even when viable endogenous microbes are present, since many more dead cells of the same type should also be present.

Finally, 16S rRNA amplicon libraries are not regarded as suitable for AEM analyses due to the highly fragmentated nature of aDNA and considerably long polymorphisms in several regions of the 16S gene [107].

Concluding remarks and future perspectives

Considerable scope exists for investigating the microbiomes of increasingly ancient samples. Over time, field, laboratory, and bioinformatic methods have refined the retrieval and authentication of AEMs. However, to date, approaches in aNA research have largely been informed by aDNA studies that do not address the nuances of environmental microbiomes (see Outstanding questions). So much remains unknown about cryospheric microbes, including their capacity for survival, and true diversity, that the study of AEMs remains challenging and exciting. While the study of ancient, deceased microbes has developed appreciably – largely thanks to damage pattern analysis – confident characterization of ancient, viable microorganisms continues to present complex authentication issues. We therefore stress that more work is needed before consensus on this subject can be achieved across both the microbiology and wider ancient biomolecules community. We highlight that dating AEMs requires robust knowledge of sample geology and microbiology. As for aRNA, validated authentication methods are lacking, and the field of aRNA would undoubtedly benefit from greater research attention.

Outstanding questions

How do we distinguish between viable, nonreplicating cells and those which are occasionally replicating? This is an important question, as it is critical to understanding their age relative to the sample material and their positioning in evolutionary history.

What is the best approach for accurately dating AEM community members? This is a serious challenge given that archived microbial communities could contain cells that may have been sporadically active, possibly even replicating for varying periods of time. Simply relying on the age of the surrounding earth or ice may not be an appropriate proxy for microbial age.

How do we authenticate aDNA for entire microbial communities which often contain poorly characterized organisms? Although pipelines exist, they are not straightforward to implement, require substantial computing resources, and rely on reference sequences.

Can we consistently recover and authenticate aRNA from AEM? There is growing interest in investigating this possibility. This allows the study of aRNA viruses and possibly ancient microbial expression profiles.

Is the distribution of cryosphere samples suitable for ancient microbial community analyses broadly reflective of environmental heterogeneity? Due to the difficulty of obtaining cryosphere samples, there is a tendency to sample similar, well known, or well-resourced locations.

Acknowledgments

This work was supported by AIR@InnoHK administered by Innovation and Technology Commission. We are also very grateful to Jean Zhao for producing the figures accompanying this review article.

Declaration of interests

The authors declare no competing interests.

References

1. Peng, X. *et al.* (2021) A holistic assessment of 1979–2016 global cryospheric extent. *Earth's Future* 9, e2020EF001969
2. Key, J.R. *et al.* (2020) Snow and ice products from ABI on the GOES-R Series. In *The GOES-R Series*, pp. 165–177, Elsevier
3. Dalén, L. *et al.* (2023) Deep-time paleogenomics and the limits of DNA survival. *Science* 382, 48–53
4. Orlando, L. *et al.* (2021) Ancient DNA analysis. *Nat. Rev. Methods Primers* 1, 14
5. Smith, O. and Gilbert, M.T.P. (2019) *Ancient RNA*. Paleogenomics: genome-scale analysis of ancient DNA pp. 53–74
6. Hebsgaard, M.B. *et al.* (2005) Geologically ancient DNA: fact or artefact? *Trends Microbiol.* 13, 212–220
7. Duchêne, S. *et al.* (2020) The recovery, interpretation and use of ancient pathogen genomes. *Curr. Biol.* 30, R1215–R1231
8. Feldman, M. *et al.* (2016) A high-coverage *Yersinia pestis* genome from a sixth-century Justinianic Plague victim. *Mol. Biol. Evol.* 33, 2911–2923
9. Taubenberger, J.K. *et al.* (2007) Discovery and characterization of the 1918 pandemic influenza virus in historical context. *Antivir. Ther.* 12, 581–591
10. D'Costa, V.M. *et al.* (2011) Antibiotic resistance is ancient. *Nature* 477, 457–461
11. Mackelprang, R. *et al.* (2011) Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* 480, 368–371
12. Abramov, A. *et al.* (2021) Are permafrost microorganisms as old as permafrost? *FEMS Microbiol. Ecol.* 97, fiaa260
13. Grasso, G. *et al.* (2024) Paleomicrobiology: tracking the past microbial life from single species to entire microbial communities. *Microb. Biotechnol.* 17, e14390
14. Andersen, K. *et al.* (2004) High-resolution record of Northern Hemisphere climate extending into the last interglacial period. *Nature* 431, 147–151
15. Bidle, K.D. *et al.* (2007) Fossil genes and microbes in the oldest ice on Earth. *P Natl Acad. Sci. USA* 104, 13455–13460

16. Bergelin, M. *et al.* (2022) Cosmogenic nuclide dating of two stacked ice masses: Ong Valley, Antarctica. *Cryosphere* 16, 2793–2817
17. Bibby, T. *et al.* (2016) Million year old ice found under meter thick debris layer in Antarctica. *Geophys. Res. Lett.* 43, 6995–7001
18. Paun, V.I. *et al.* (2021) First report on antibiotic resistance and antimicrobial activity of bacterial isolates from 13,000-year old cave ice core. *Sci. Rep.* 11, 514
19. Gilbert, G.L. *et al.* (2016) Recent advances (2008–2015) in the study of ground ice and cryostratigraphy. *Permafr. Periglac. Process.* 27, 377–389
20. Allaby, R. *et al.* (2023) Pleistocene-Holocene sedaDNA reconstruction of Southern Doggerland reveals early colonization before inundation consistent with northern refugia. *Res. Square*, Published online September 21, 2023. https://doi.org/10.21203/rs.3.rs-3296992/v1
21. Willerslev, E. *et al.* (2004) Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends Ecol. Evol.* 19, 141–147
22. Johnson, S.S. *et al.* (2007) Ancient bacteria show evidence of DNA repair. *P Natl Acad. Sci. USA* 104, 14401–14405
23. Eisenhofer, R. *et al.* (2016) Isolating viable ancient bacteria: what you put in is what you get out. *Genome Announc.* 4, e00712-16
24. Willerslev, E. *et al.* (2004) Long-term persistence of bacterial DNA. *Curr. Biol.* 14, R9–R10
25. Austin, J.J. *et al.* (1997) Problems of reproducibility – does geologically ancient DNA survive in amber-preserved insects? *Proc. R. Soc. Lond. B Biol.* 264, 467–474
26. van der Valk, T. *et al.* (2021) Million-year-old DNA sheds light on the genomic history of mammoths. *Nature* 591, 265–269
27. Kjær, K.H. *et al.* (2022) A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. *Nature* 612, 283–291
28. Fernandez-Guerra, A. *et al.* (2023) A 2-million-year-old microbial and viral communities from the Kap København Formation in North Greenland. *bioRxiv*, Published online June 12, 2023. https://doi.org/10.1101/2023.06.10.544454
29. Courtin, J. *et al.* (2022) Pleistocene glacial and interglacial ecosystems inferred from ancient DNA analyses of permafrost sediments from Batagay megaslump, East Siberia. *Environ. DNA* 4, 1265–1283
30. Kennedy, M.J. *et al.* (1994) Preservation records of microorganisms: evidence of the tenacity of life. *Microbiology* 140, 2513–2529
31. Nerber, H.N. and Sorg, J.A. (2024) The small acid-soluble proteins of spore-forming organisms: Similarities and differences in function. *Anaerobe* 87, 102844
32. Liang, R. *et al.* (2019) Aspartic acid racemization constrains long-term viability and longevity of endospores. *FEMS Microbiol. Ecol.* 95, fiz132
33. Vreeland, R.H. *et al.* (2000) Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. *Nature* 407, 897–900
34. Cano, R.J. and Borucki, M.K. (1995) Revival and identification of bacterial spores in 25- to 40-million-year-old Dominican amber. *Science* 268, 1060–1064
35. Mackelprang, R. *et al.* (2017) Microbial survival strategies in ancient permafrost: insights from metagenomics. *ISME J.* 11, 2305–2318
36. Liang, R. *et al.* (2019) Predominance of anaerobic, spore-forming bacteria in metabolically active microbial communities from ancient Siberian permafrost. *Appl. Environ. Microbiol.* 85, e00560-00519
37. Ulrich, N. *et al.* (2018) Experimental studies addressing the longevity of *Bacillus subtilis* spores – the first data from a 500-year experiment. *PLoS One* 13, e0208425
38. Brouchkov, A. *et al.* (2016) Is the ancient permafrost bacteria able to keep DNA stable? *J. Genet.* 95, 1003–1007
39. Ng, F. *et al.* (2005) Fast-growing till over ancient ice in Beacon Valley, Antarctica. *Geology* 33, 121–124
40. Legendre, M. *et al.* (2014) Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc. Natl. Acad. Sci. USA* 111, 4274–4279
41. Alempic, J.-M. *et al.* (2023) An update on eukaryotic viruses revived from ancient permafrost. *Viruses* 15, 564
42. Ng, T.F.F. *et al.* (2014) Preservation of viral genomes in 700-y-old caribou feces from a subarctic ice patch. *P Natl Acad. Sci. USA* 111, 16842–16847
43. Holmes, E.C. (2014) Freezing viruses in time. *Proc. Natl. Acad. Sci. USA* 111, 16643–16644
44. Zhong, Z.-P. *et al.* (2024) Glacier-preserved Tibetan Plateau viral community probably linked to warm–cold climate variations. *Nat. Geosci.* 17, 912–919
45. Raoult, D. *et al.* (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* 306, 1344–1350
46. Smith, O. *et al.* (2019) Ancient RNA from Late Pleistocene permafrost and historical canids shows tissue-specific transcriptome survival. *PLoS Biol.* 17, e3000166
47. Friedländer, M.R. and Gilbert, M.T.P. (2024) How ancient RNA survives and what we can learn from it. *Nat. Rev. Mol. Cell Biol.* 25, 417–418
48. Castello, J.D. *et al.* (1999) Detection of tomato mosaic tobamovirus RNA in ancient glacial ice. *Polar Biol.* 22, 207–212
49. Peyambari, M. *et al.* (2019) A 1,000-year-old RNA virus. *J. Virol.* 93, e01188-18
50. Nishimura, L. *et al.* (2022) Detection of ancient viruses and long-term viral evolution. *Viruses* 14, 1336
51. Scheel, M. *et al.* (2023) Abrupt permafrost thaw triggers activity of copiotrophs and microbiome predators. *FEMS Microbiol. Ecol.* 99, fiad123
52. Zhong, Z.-P. *et al.* (2021) Glacier ice archives nearly 15,000-year-old microbes and phages. *Microbiome* 9, 160
53. Rogers, S.O. *et al.* (2005) Recommendations for elimination of contaminants and authentication of isolates in ancient ice cores. In *Life in Ancient Ice* (Rogers, S.O. and Castello, J.D., eds), pp. 5–21, Princeton University Press
54. Kowalchuk, G.A. *et al.* (2007) Valid recovery of nucleic acid sequence information from high contamination risk samples – ancient DNA and environmental DNA. *Perspect. bioanal.* 2, 357–371
55. Eisenhofer, R. *et al.* (2019) Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends Microbiol.* 27, 105–117
56. Pérez, V. *et al.* (2022) A case study for the recovery of authentic microbial ancient DNA from soil samples. *Microorganisms* 10, 1623
57. Warinner, C. *et al.* (2017) A robust framework for microbial archaeology. *Annu. Rev. Genomics Hum. Genet.* 18, 321–356
58. Key, F.M. *et al.* (2017) Mining metagenomic data sets for ancient DNA: recommended protocols for authentication. *Trends Genet.* 33, 508–520
59. Briggs, A.W. *et al.* (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci.* 104, 14616–14621
60. Dabney, J. *et al.* (2013) Ancient DNA damage. *CSH Perspect. Biol.* 5, a012567
61. Weiß, C.L. *et al.* (2016) Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. *R. Soc. Open Sci.* 3, 160239
62. Mitchell, D. *et al.* (2005) Damage and repair of ancient DNA. *Mutat. Res. Fund.Mol. M.* 571, 265–276
63. Kistler, L. *et al.* (2017) A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic Acids Res.* 45, 6310–6320
64. Cai, P. *et al.* (2006) Interactions of DNA with clay minerals and soil colloidal particles and protection against degradation by DNase. *Environ. Sci. Technol.* 40, 2971–2976
65. Saidi-Mehrabad, A. *et al.* (2020) Optimization of subsampling, decontamination, and DNA extraction of difficult peat and silt permafrost samples. *Sci. Rep.* 10, 14295
66. Garcia-Garcera, M. *et al.* (2011) Fragmentation of contaminant and endogenous DNA in ancient samples determined by shotgun sequencing; prospects for human palaeogenomics. *PLoS One* 6, e24161
67. Weiß, C.L. *et al.* (2020) Mining ancient microbiomes using selective enrichment of damaged DNA molecules. *BMC Genomics* 21, 432

68. Gansauge, M.-T. and Meyer, M. (2014) Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res.* 24, 1543–1549
69. Borry, M. *et al.* (2021) PyDamage: automated ancient damage identification and estimation for contigs in ancient DNA de novo assembly. *PeerJ* 9, e11845
70. Neukamm, J. *et al.* (2021) DamageProfiler: fast damage pattern calculation for ancient DNA. *Bioinformatics* 37, 3652–3653
71. Hübler, R. *et al.* (2019) HOPS: automated detection and authentication of pathogen DNA in archaeological remains. *Genome Biol.* 20, 280
72. Eisenhofer, R. and Weyrich, L.S. (2019) Assessing alignment-based taxonomic classification of ancient microbial DNA. *PeerJ* 7, e6594
73. Cárdenas, Y.O.A. *et al.* (2022) Benchmarking metagenomics classifiers on ancient viral DNA: a simulation study. *PeerJ* 10, e12784
74. Kim, D. *et al.* (2016) Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res.* 26, 1721–1729
75. Pochon, Z. *et al.* (2023) aMeta: an accurate and memory-efficient ancient metagenomic profiling workflow. *Genome Biol.* 24, 242
76. Michelsen, C. *et al.* (2022) metaDMG – a fast and accurate ancient DNA damage toolkit for metagenomic data. *bioRxiv*, Published online December 9, 2022. https://doi.org/10.1101/2022.12.06.519264
77. Peyrégne, S. and Peter, B.M. (2020) AuthentiCT: a model of ancient DNA damage to estimate the proportion of present-day DNA contamination. *Genome Biol.* 21, 246
78. Gansauge, M.-T. and Meyer, M. (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* 8, 737–748
79. Gansauge, M.-T. *et al.* (2020) Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nat. Protoc.* 15, 2279–2300
80. Vollmers, J. *et al.* (2022) How clear is our current view on microbial dark matter?(Re-) assessing public MAG & SAG datasets with MDMcleaner. *Nucleic Acids Res.* 50, e76
81. Zamkovaya, T. *et al.* (2021) A network approach to elucidate and prioritize microbial dark matter in microbial communities. *ISME J.* 15, 228–244
82. Gancz, A.S. *et al.* (2023) Ancient dental calculus reveals oral microbiome shifts associated with lifestyle and disease in Great Britain. *Nat. Microbiol.* 8, 2315–2325
83. Smith, O. *et al.* (2014) A complete ancient RNA genome: identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus. *Sci. Rep.* 4, 4003
84. Der Sarkissian, C. *et al.* (2021) Ancient metagenomic studies: considerations for the wider scientific community. *Msystems* 6, e01315-01321
85. Goncharov, A. *et al.* (2016) Draft genome sequence of *Enterococcus faecium* strain 58m, isolated from intestinal tract content of a woolly mammoth, *Mammuthus primigenius*. *Genome Announc.* 4, e01706-15
86. van Bergeijk, D.A. *et al.* (2024) Taxonomic and metabolic diversity of Actinomycetota isolated from faeces of a 28,000-year-old mammoth. *Environ. Microbiol.* 26, e16589
87. Claverie, J.-M. and Abergel, C. (2016) Giant viruses: the difficult breaking of multiple epistemological barriers. *Stud. Hist. Phil. Biol. Biomed. Sci.* 59, 89–99
88. Levasseur, A. *et al.* (2016) Comparison of a modern and fossil pithovirus reveals its genetic conservation and evolution. *Genome Biol. Evol.* 8, 2333–2339
89. Duchêne, S. and Holmes, E.C. (2018) Estimating evolutionary rates in giant viruses using ancient genomes. *Virus Evol.* 4, vey006
90. de-Dios, T. *et al.* (2023) An adagio for viruses, played out on ancient DNA. *Genome Biol. Evol.* 15, evad047
91. Duchêne, S. *et al.* (2014) Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proc. R. Soc. B Biol. Sci.* 281, 20140732
92. Duchêne, S. *et al.* (2016) Genome-scale rates of evolutionary change in bacteria. *Microb. Genom.* 2, e000094
93. Eaton, K. *et al.* (2023) Plagued by a cryptic clock: insight and issues from the global phylogeny of *Yersinia pestis*. *Commun. Biol.* 6, 23
94. Rozwalak, P. *et al.* (2024) Ultraconserved bacteriophage genome sequence identified in 1300-year-old human palaeofaeces. *Nat. Commun.* 15, 495
95. Vreeland, R. and Rosenzweig, W. (2002) The question of uniqueness of ancient bacteria. *J. Ind. Microbiol. Biotechnol.* 28, 32–41
96. Jørgensen, B.B. and Marshall, I.P. (2016) Slow microbial life in the seabed. *Annu. Rev. Mar. Sci.* 8, 311–332
97. Becraft, E.D. *et al.* (2021) Evolutionary stasis of a deep subsurface microbial lineage. *ISME J.* 15, 2830–2842
98. Weyrich, L.S. *et al.* (2019) Laboratory contamination over time during low-biomass sample analysis. *Mol. Ecol. Resour.* 19, 982–996
99. Glassing, A. *et al.* (2016) Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog.* 8, 24
100. Knights, D. *et al.* (2011) Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* 8, 761–763
101. Knowlton, C. *et al.* (2013) Microbial analyses of ancient ice core sections from Greenland and Antarctica. *Biology* 2, 206–232
102. Liang, R. *et al.* (2021) Genomic reconstruction of fossil and living microorganisms in ancient Siberian permafrost. *Microbiome* 9, 110
103. Rigou, S. *et al.* (2022) Metagenomic survey of the microbiome of ancient Siberian permafrost and modern Kamchatkan cryosols. *microLife* 3, uqac003
104. Pääbo, S. *et al.* (1990) DNA damage promotes jumping between templates during enzymatic amplification. *J. Biol. Chem.* 265, 4718–4721
105. Silverman, J.D. *et al.* (2021) Measuring and mitigating PCR bias in microbiota datasets. *PLoS Comput. Biol.* 17, e1009113
106. Kanagawa, T. (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J. Biosci. Bioeng.* 96, 317–323
107. Ziesemer, K.A. *et al.* (2015) Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. *Sci. Rep.* 5, 16498
108. Morita, R.Y. (1975) Psychrophilic bacteria. *Bacteriol. Rev.* 39, 144–167
109. Lemke, P. *et al.* (2007) Observations: Changes in Snow, Ice and Frozen Ground. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Solomon, S. *et al.*, eds), Cambridge University Press, Cambridge, UK and New York, NY, USA
110. Rowley, T. *et al.* (2015) Periglacial processes and landforms in the critical zone. In *Developments in Earth Surface Processes*, pp. 397–447, Elsevier
111. Lachenbruch, A.H. (1962) *Mechanics of Thermal Contraction Cracks and Ice-wedge Polygons in Permafrost*, Geological Society of America Published online https://doi.org/10.1130/SPE70-p1
112. Seppälä, M. (2011) Synthesis of studies of palsa formation underlining the importance of local environmental and physical characteristics. *Quat. Res.* 75, 366–370
113. Raynaud, D. *et al.* (2020) Antarctic air bubbles and the long-term ice core record of CO2 and other greenhouse gases. In *Past Antarctica*, pp. 27–50, Elsevier
114. Juck, D. *et al.* (2005) Utilization of fluorescent microspheres and a green fluorescent protein-marked strain for assessment of microbiological contamination of permafrost and ground ice core samples from the Canadian High Arctic. *Appl. Environ. Microbiol.* 71, 1035–1041
115. Bang-Andreasen, T. *et al.* (2017) Potential microbial contamination during sampling of permafrost soil assessed by tracers. *Sci. Rep.* 7, 43338
116. Miteva, V. *et al.* (2014) Penetration of fluorescent microspheres into the NEEM (North Eemian) Greenland ice core

to assess the probability of microbial contamination. *Polar Biol.* 37, 47–59
117. Christner, B.C. *et al.* (2005) Glacial ice cores: a model system for developing extraterrestrial decontamination protocols. *Icarus* 174, 572–584
118. Barbato, R.A. *et al.* (2016) Removal of exogenous materials from the outer portion of frozen cores to investigate the ancient biological communities harbored inside. *JoVE*, e54091
119. Bakermans, C. *et al.* (2003) Reproduction and metabolism at −10C of bacteria isolated from Siberian permafrost. *Environ. Microbiol.* 5, 321–326
120. Mykytczuk, N. *et al.* (2013) Bacterial growth at −15C; molecular insights from the permafrost bacterium *Planococcus halocryophilus* Or1. *ISME J.* 7, 1211–1226
121. Rogers, S.O. *et al.* (2004) Comparisons of protocols for decontamination of environmental ice samples for biological and molecular examinations. *Appl. Environ. Microbiol.* 70, 2540–2544
122. Vishnivetskaya, T.A. *et al.* (2014) Commercial DNA extraction kits impact observed microbial community composition in permafrost samples. *FEMS Microbiol. Ecol.* 87, 217–230
123. Feng, S. *et al.* (2023) Comparison of three DNA extraction methods for recovery of microbial DNA from Arctic permafrost. *microPublic. Biol.*, Published online June 21, 2023. https://doi.org/10.17912/micropub.biology.000834
124. Fulton, T.L. (2012) Setting up an ancient DNA laboratory. *Ancient DNA Methods Protoc.* 1–11
125. Jónsson, H. *et al.* (2013) mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29, 1682–1684
126. Al-Asadi, H. *et al.* (2019) Inference and visualization of DNA damage patterns using a grade of membership model. *Bioinformatics* 35, 1292–1298
127. Everett, R. and Cribdon, B. (2023) MetaDamage tool: examining post-mortem damage in sedaDNA on a metagenomic scale. *Front. Ecol. Evol.* 10, 888421
128. Skoglund, P. *et al.* (2014) Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci. USA* 111, 2229–2234