Fossil Genes and Microbes Entombed in the Oldest Ice on Earth

SangHoon Lee[1,2], Kay D. Bidle[1], David R. Marchant[3] and Paul G. Falkowski[1,4*]

[1] Environmental Biophysics and Molecular Ecology Program, Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ 08901, USA

[2] Polar Research Institute, Korea Ocean Research & Development Institute, Incheon 406-840, Korea

[3] Department of Earth Sciences, Boston University, Boston, MA 02215, USA

[4] Department of Geological Sciences, Rutgers University, Piscataway, NJ 08901

[*] corresponding author

email: S. Lee (shlee@kopri.re.kr)

K. Bidle (bidle@marine.rutgers.edu)

P. Falkowski (falko@marine.rutgers.edu)

D. Marchant (marchant@bu.edu)

Based on incorporation of radiolabeled substrates and growth in meltwaters, we detected metabolic activity and cell viability of microbes encased in ice samples from the Transantarctic mountains. These samples span the last 8 million years (Ma) and contain the oldest known ice on Earth. Although DNA degraded exponentially with time, with a half-life of ~1.2 Ma, metagenomic analyses of intact sequences from even the oldest ice revealed diverse orthologs to extant metabolic genes. The preservation of genetic material in ancient ice potentially facilitates horizontal gene exchange during periods of rapid climate warming.

The Antarctic continent offers unique environments for understanding the limits to geological and biological preservation of life and genetic material in immobilized dormant states. Analyses of Antarctic microorganisms within the array of subglacial lakes in East Antarctica, including Lake Vostok (*1*), Lake Bonney (*2*), as well as the Taylor Dome region of the Transantarctic Mountains (*3*), have revealed the potential viability of microorganisms preserved in ice over ~300 Ka. In the Dry Valleys of the Transantarctic Mountains, however, patches of much older ice persist yet little is known about the viability of microbes or the state of genetic material in these regions. To that end, we examined ice samples from a debris-covered alpine glacier that heads in Mullins Valley (DLE-98-12) and terminates along a diffuse boundary in Beacon Valley (EME-98-03) (Figure 1A,B).

Geological analyses of this region, including the stratigraphy and age dating of debris overlying the ice (*4-8*) indicate that the minimum age of the ice, as determined by cosmogenic-nuclide exposure-age dating and $^{40}Ar/^{39}Ar$ analysis of *in situ* volcanic ash-fall deposits on the debris capping the ice are ~100 Ka years for DLE-98-12, and approximately 8 million years (Ma) for

EME-98-03 (Figure 1B). The mean annual temperature in this region is -23˚C, and maximum near-surface (2 cm) summertime air temperature is 3˚C. Analyses of the ice crystal structure and its stable-isotope composition ($\delta^{18}O$, δD) indicate that the ice samples have remained frozen, without episodes of melting, since transformation from snow to glacier ice at the head of Mullins Valley (*4, 5, 9*). Hence, microorganisms or DNA encased in the ice at EME-98-03 and DLE-98-12 have been separated and preserved since the late Miocene and late Pleistocene, respectively.

DLE-98-12 and EME-98-03 contained a broad size spectrum of particles and rock debris, ranging from fine silt to coarse sand. This inorganic, solid phase material, which likely originated from rockfall (sandstone and dolorite) onto the ice accumulation zone, was heterogeneously distributed and contributed to variations in chemical properties and particle-associated microzones within and between the samples (*2, 10*) (Supporting Online Material, Table S1). For example, the melt water of EME-98-03 was neutral (pH = 6.9), while that from DLE-98-12 was mildly acidic (pH = 4), due to the chemical reactions of minerals (e.g., pyrite) in the latter. Scanning electron microscopy (SEM) of DLE-98-12 meltwater collected on 0.2 µm polycarbonate filters (Supporting Online Material, Materials and Methods) revealed the presence of distinct coccoid particles, suggestive of intact microbes, interspersed with mineral granules (Fig 1C, arrows). SEM analysis of EME-98-03 revealed a much higher fine-particle load along with abundant sheath-like filaments, which were evenly distributed throughout the sample (Figure 1D). Staining with SYBR Gold, which allows for direct visualization and counting of DNA-containing cells (Figure 1E,F), indicated that bacterial concentrations were 5.07 (±0.98) x 10^5 and 3.28 (±1.56) x 10^4 cells ml^{-1} for DLE-98-12 and EME-98-03, respectively. These bacterial concentrations are comparable to those from polar freshwater (*11, 12*) and sea ices (*13*), but 2 to 3 orders of magnitude higher than in Antarctic snow

(*14*) and subglacial lake ice accretions (*1*).

To more fully understand the genomic composition of the two ice samples, we amplified community DNA and constructed clone libraries with *Bacteria*-specific 16S rDNA primers (Figure 2A). Repeated attempts to amplify community DNA with *Archaea*- and *Eukaryotic*-specific primers were not effective for either sample. Both EME-98-03 and DLE-98-12 were dominated by relatively few bacterial phylotypes, but broader representatives occurred in low frequencies (Figure 2B; Table S2). After removal of potential chimeric sequences (n=36) and de-replication of identical sequences, 30 distinct phylotypes (11 from 96 EME clones, and 19 from 123 DLE clones) were identified; most of these were from known soil microbes. BLAST analysis of the 30 phylotypes (Table S2) revealed that no clones were identical to 16S rDNA sequences in GenBank, and several had relatively reduced phylogenetic similarity (<95%) compared with currently catalogued species. Overall, the EME-98-03 clone library had lower diversity and was dominated by β-proteobacteria (68 of 96 clones), with one phylotype (Group A1, EME076) dominating community representation (62 of 96 clones), followed by two γ-proteobacteria phylotypes (22 clones). The remaining sequences, belonging to α-proteobacteria, *Acidobacteria*, *Firmicutes*, and *Cytophaga-Flavobacterium-Bacteroides* (CFB) divisions, occurred at relatively lower frequencies. Sixty-eight EME-98-03 clones returned BLAST matches with ≥98% identity to *Leptothrix sp.* (EME076; 99% identity), a representative of the sheathed, filamentous bacteria (*15*) (Fig. 1D). More than 80% of the DLE-98-12 clones (96 of 123) represented phylotypes of *Acidobacteria*. A single phylotype of α-proteobacteria represented the second largest division (17 clones) in DLE-98-12, while phylotypes of *Actinobacteria* and CFB were found in lower frequencies. No DLE clones were similar to any known sequences at the species level, while 111 of the DLE clones found matches at

the genus level (≥95% identity).

To examine whether any of the microorganisms in these two ice samples retained metabolic activity, we measured the incorporation of ^{3}H-leucine and ^{3}H-thymidine into protein and nucleic acid pools, respectively, and followed the incorporation and subsequent respiration of ^{14}C-glucose (Materials and Methods). Although radiotracer incorporation was variable due to heterogeneous distribution of inorganic particles and microbes, metabolic activity was readily detected in meltwater from both samples after correcting with formalin-killed controls (Figure 3A,B; Table S3). The radioisotope uptake signal from MilliQ® rinse water was negligible (<0.1%). Overall, DLE-98-12 meltwater displayed higher and earlier ^{3}H-leucine and ^{14}C uptake, as well as ^{14}C respiration compared with the older ice (Figure 3A,B; Table S3). EME-98-03 had higher ^{3}H-thymidine uptake, but ^{3}H-leucine uptake was not detected. In a separate experiment, nutrient augmentation of the meltwaters stimulated earlier and higher radiotracer uptake above that observed in unamended samples (Table S3), indicating that the resuscitated microbes responded to added nutrients. The effect was most evident in the dramatic increase in ^{3}H-leucine uptake by EME-98-03.

Long-term incubation of nutrient-amended meltwater at 4^{o}C in the dark resulted in the recovery of viable cells from both ice samples (Fig 3C,D; Table S3); MilliQ® -inoculated controls showed no detectable growth. For DLE-98-12, two of the nutrient amended samples showed relatively rapid growth over a 70-day period (doubling time of ~7 d), from which we obtained 16S rDNA amplicons (Figure 2F) and isolated viable colonies on plate media (Figure 2E), both of which were identified as *Arthrobacter* sp. via 16S rDNA sequencing (Table S4). Their presence in the initial ice microbial community was verified by applying isolate-specific 16S rDNA primer sets in nested PCR reactions that used the original ice DNA 16S rDNA amplicons as templates (Fig. 3E).

Cultures of EME-98-03 melt water showed extremely slow, low-level growth over the entire 285 d time period (Fig. 3D), with an apparent doubling time of ~200 d. This rate was confirmed by direct staining and enumeration of bacterial cells with SYBR Gold. Unfortunately, repeated attempts to identify the population of viable cells from EME-98-03 via PCR amplification using *Bacteria*-, *Archaea*- and *Eukaryota*-specific primer sets or recovery of cells on solid media failed. The extremely slow and low-level growth, combined with the inability to recover viable cells, suggests that while the microbes encased in ice for 8 Myr are not sterile, their viability is seriously compromised.

The environmental DNA trapped in these ancient ice samples represents a literal gene bank, with deposits made when the ices formed. Our 16S rDNA analyses suggest that deposits were made both from Aeolian fluxes of soil microbes external to the Antarctic continent, as foretold by Darwin in 1839 in the equatorial Atlantic (*16*) and more recently documented by Muñoz et al. for the Antarctic continent (*17*), as well as from endogenous sources, such as glacial till and animal feces (*18*). To examine the preservation and integrity of the genomic material, we compared the average sizes of community DNA. DNA extracted from DLE-98-12 migrated in agarose gels as a tight band at ~20 Kb, while that from EME-98-03 was a broad smear of much lower molecular weight fragments (100 - <1000 bp), indicative of extensive degradation (Figure 4). Using image analysis and densitometry profiles, we calculated the weighted mean size of the extracted genomic DNA to be 18,500 bp for DLE-98-12 and 210 bp for EME-98-03 (*19*). Electrophoretic analyses of DNA extracted from additional ice samples from this region (DLE-98-CS-1, ~300 Ka; MCI-04-13, ~3 ± 1 Ma; and EME-98-8, 5-6 Ma) were performed to further examine the relationship between age and DNA integrity (Figures 1B, 4). That analysis revealed that DNA degradation follows first-

order kinetics with a half-life of ~1.1 Ma. This degradation rate is generally consistent with calculations of DNA survival and depurination rates, where an average bacterial genome (3×10^6 bp) would be broken into ~100 bp fragments in >1.2 Ma at temperatures below -20°C (*20*), although previous studies were not able to detect DNA from 2 to 8 Ma permafrost from Beacon Valley (*21*). The progressive change in DNA size suggests that the buried ice has not undergone melting in the recent geologic past, and that the potential contribution of secondary ice lenses formed by refreezing of modern meltwater to the sampled ices is negligible (*22*).

Given that enzymatic rates of DNA degradation decline by an order of magnitude for every 10°C drop in temperature, enhanced preservation is generally expected for permanently cold environments like Antarctica. What is the primary mechanism causing the degradation of DNA? One possibility is hydrolysis and/or oxidation by hydroxyl ion or proton attack on the sugar backbone or glycosidic bonds (depurination/depyrimidination) (*20*, *23*, *24*). This hypothesis is not supported by the chemical data; the younger, more acidic ice sample follows the same degradation pattern as older, neutral samples (Figure 4). A second possibility is the generation of ionizing radiation from cosmic high-energy particles (e.g., μ-muons and/or protons) or endogenous natural radioactivity in the immediate vicinity of the DNA molecules. Due to the Compton effect (*25*), Antarctica receives among the highest amount of incoming cosmic radiation on the planet (*26*). Given a target size of ~ 10 to 20 $Å^2$ for a covalent bond along the phosphosugar backbone, a minimum of 6 ionizing events is required to reduce the DNA from 20,000 bp to 210 bp over the 8 Myr interval. Assuming a constant flux over this time, and ignoring the upward migration of ice samples toward higher cosmic fluxes nearer the ground surface as overlying ice sublimed (*6*), target theory predicts that a cosmic radiative flux of 2-3 particles cm^{-2} s^{-1} is required to support this

degradation rate of DNA; well within the bounds of the present flux. Given the considerable degradation in the older ice samples and their consistency with the DNA 'kill curve' relationship, a time-dependent increase in the frequency of intermolecular/interstrand crosslinking, both of which might serve to stabilize the DNA over time and reduce fragmentation (*20*), is unlikely. Interestingly, the decay kinetics of DNA size over this time scale suggests that a DNA size model can be used as a geomolecular proxy to independently estimate ice age in this region.

In light of the disparate community DNA sizes and the successful PCR amplification of 1.5 kb 16S rDNA amplicons from both ice types, it is likely that 16S rDNA community reconstruction is more representative for DLE-98-12, as the expected amplicon size is well within the ~18 kb template size. On the other hand, the true community diversity was likely underrepresented in our 16S rDNA reconstructions for EME 98-12; the significant size discrepancy between expected amplicon size and average community DNA size would inevitably be biased towards a small subset of the bacterial population. The correspondence between both endogenous and exogenous sources of the bacterial 16S rDNA, meltwater chemistry, and SEM observations, and the failure to recover either amplicons or viable cells from 0.22 μm filtered MilliQ® waters used to ablate the ice prior to sampling microbes or DNA, strongly argues against artifacts due to contamination. Moreover, our results imply preferential preservation of genomic DNA in a subset of community members (*21*).

In order to further evaluate the molecular composition of microbial genes from the two ice samples, we conducted a metagenomic survey of community DNA by sequencing a subset of randomly picked clones from a bacterial artificial chromosome (BAC) library containing partially BamHI-digested DLE-98-12 community DNA (n=131) and an undigested, end-repaired shotgun library for EME-98-03 community DNA (n=427). Sequencing results reflected the dramatic size

discrepancy in DNA between the different ice samples. A majority of sequence query lengths for EME-98-03 were <300 bp, compared with >600 bp for DLE-98-12 (Figure 5B), with the latter representing only a portion of the insert size due to constraints on sequencing read lengths. Although a significant proportion (45.9% for DLE-98-12 and 39.4% for EME-98-03) of the sequences in both samples had no representative orthologs in public databases, the remaining sequences encompassed a highly diverse set of orthologs to extant functional genes (Figure 5a). These included genes for DNA replication, transcription and translation, energy transduction, carbohydrate and lipid metabolism, cell motility and antibiotic resistance, among others (*27*). An overwhelming majority (>93%) of the gene sequences are derived from the *Bacteria* domain. Indeed, in EME-98-03, orthologs to extant *Actinobacteria* (nonspore-forming, Gram-positive), *Proteobacteria* (gram negative) and *Firmacutes* (spore-forming, Gram-positive) phyla accounted for 52.6%, 27.2% and 8.8% of the BLAST matches, respectively. These data support findings of superior DNA survival by bacteria, most notably the *Actinobacteria* (*21*).

What could cause the relatively high fraction of the community genomes that have no known orthologs? One possibility is these are truly novel genes not represented in the database. However, a second possibility is that small fragment size, due to DNA degradation, introduced an analysis artifact, preventing a statistically significant comparison with homologous orthologs in the database. If the latter, we would expect a correlation between length of query fragment and e-value score. That is not observed in either sample (Figure 5B). Rather, we observed a more random pattern, where a subset of small fragments frequently had high homology with extant genes and others returned ‘no hit’ implying that many of the recovered “no-hit” sequences are truly novel genes.

The community DNA immobilized in Antarctic ice is essentially a “gene popsicle” that potentially can be acquired by extant organisms upon thawing. Given the widespread influence of lateral gene transfer within microbial populations and its putative influence on the tempo of microbial evolution (*28*, *29*), one can envision periods in Earth’s history when very large numbers of ancient genes became environmentally available as ice sheets melt. The vast majority of the ice on the Antarctic continent is <1 Ma; release of the microbes and DNA from such ice into the environment could potentially influence microbial genome content and structure. Although the volume of ice from older formations is relatively small, the DNA in ice ~ 8 Ma contains identifiable genetic information that potentially can be acquired, repaired, incorporated and used by microbes. Our analysis suggests that melting of polar ice in the geological past may have provided a conduit for large scale, phage independent, horizontal gene transfer, potentially scrambling microbial phylogenies and accelerating the tempo of microbial evolution. Indeed, the tempo of evolution following major global glaciations appears to have increased dramatically (*30*), although causal mechanisms have been poorly defined.

Finally, we note that the preservation of microbes and their genes in icy comets may have allowed transfer of genetic material across the solar system. However, given the extremely high cosmic radiation flux in space, our DNA damage curve would most likely be conservative. Hence, in so far as our analysis can be extrapolated to comets, our data suggest it is highly unlikely that life on Earth (or on any of the other three terrestrial planets orbiting our sun) could have been seeded by genetic material external to this solar system.

References and Notes

1. D. M. Karl *et al.*, *Science* **286**, 2144 (1999).

2. J. C. Priscu *et al.*, *Science* **280**, 2095 (1998).
3. B. Christner *et al.*, *Icarus* **144**, 479 (2000).
4. D. R. Marchant *et al.*, *GSA Bulletin* **114**, 718 (2002).
5. D. E. Sugden *et al.*, *Nature* **376**, 412 (1995).
6. J. Schäfer *et al.*, *Earth and Planetary Science Letters* **179**, 91 (2000).
7. E. Rignot, B. Hallet, A. Fountain, *Geophysical Research Letters* **10.1029/2001/GL013494** (2002).
8. J. S. Levy, D. R. Marchant, J. W. I. Head, *Antarctic Science* **in press** (2006).
9. Analysis of synthetic aperture-radar interferometry indicate that ice within the Mullins Valley debris-covered glacier ranges from modern at the valley head to ~300k near the tributary mouth and that buried ice that continues to flow out of Mullins Valley and onto the floor of upper and central Beacon Valley could be as much as 10 Ma (4). Our sample DLE-98-12, which is assigned an age of ~100 ka, lies about half-way down Mullins Valley, while sample EME-98-03 lies in the zone estimated to be up to 10 Ma. These ages were confirmed by cosmogenic dating of surface boulders and $^{40}Ar/^{39}Ar$ dating of surface ash-fall deposits. For cosmogenic dating, a suite of 10 surface cobbles was collected along the length of the Mullins Valley debris-covered glacier and dated using ^{3}He exposure age chronology. The ages range from Holocene near the valley head (~13 ka) to ~700 ka just out onto the floor of upper Beacon Valley; these ages are fully consistent with the ages provided in Rignot *et al* (7). We also dated several ash-fall deposits on top of buried glaciers in Beacon Valley using the $^{40}Ar/^{39}Ar$ techniques (5). Two separate ash-fall deposits collected near EME-98-03 yielded ages of 7.39±0.08 Ma and 8.98±0.25 Ma; another ash-fall deposit, 0.5 km distant, yielded an age of 7.69±0.10 Ma. Three additional ash-fall deposits situated about 3 km up glacier from ice sample EME-98-03 (i.e., ice younger than that at EME-98-03) yield ages of 4.00±0.47 Ma, 3.89±0.50 Ma, and 3.73±0.14 Ma. The 0° C isotherm reaches a maximum depth of ~15 cm in the region; the ice samples we analyzed come from well below this depth (a minimum of 35 cm depth). It should be noted that there are no cryoconite holes in the glacier surface and there are no morphologic or soil data that show that meltwater formed at these sample localities in the past (5).
10. To recover potentially viable microbes and genetic material without damage due to exposure of harsh chemicals or ultraviolet light, we decontaminated the exterior of the ice with a controlled melting process, similar to that used by Karl *et al.* on accreted ice from Lake Vostok (1). Melting procedures and sample processing were performed in bleach- and ethanol-sterilized work areas and used a UV-sterilized laminar flow hood and ‘personal protection’, including lab coats and sterilized gloves. A section of block ice (ca. 1 kg) was thoroughly rinsed with ice-cold ethanol, followed by copious amounts of flowing 0.22 µm-filtered MilliQ® water, effectively melting the exterior ~3 cm of each ice sample. The cleaned ice

was subsequently thawed in a sterile plastic bag at 4°C, and the meltwater was handled aseptically in a sterilized laminar flow hood. To control for false-positive contamination signals, one liter of MilliQ® rinse water was frozen, thawed, and filtered on a polycarbonate filter, and subjected to the same DNA extraction procedure. The material extracted from the contamination control filter was used as a template in PCR with the Bacterial-specific primers, and subsequently examined by agarose gel electrophoresis. Thawed MilliQ® rinse water was also used in radiotracer incubations.

11. D. A. Pearce, *Microb. Ecol.* **46**, 92 (2003).
12. C. Säwström, P. Mumford, W. Marshall, A. Hodson, J. Laybourn-Parry, *Polar Biol.* **25**, 591 (2002).
13. R. Brinkmeyer *et al.*, *Appl. Environ. Microbiol.* **69**, 6610 (2003).
14. E. J. Carpenter, S. Lin, D. G. Capone, *Appl. Environ. Microbiol.* **66**, 4514 (2000).
15. J. G. Holt, N. R. Krieg, P. H. A. Sneath, J. T. Staley, S. T. Williams, *Bergey's manual of determinative bacteriology* (Williams and Wilkins, Baltimore, ed. 9th, 1994), pp. 787.
16. C. Darwin, *The Voyage of the Beagle* (Henry Colburn, reprinted by Penguin Books, Ltd., London, 1839), pp. 425.
17. J. Muñoz, A. M. Felicísimo, F. Cabezas, A. R. Burgaz, I. Martínez, *Science* **304**, 1144 (2004).
18. Analysis of bacterial 16S rDNA clones retrieved from the DLE-98-12 ice sample revealed a 91% identity (1313/1435 bp) of clone group DLE084 to uncultured bacterium clone KD3-117 detected in penguin droppings from Ardley Island, Antarctica (Accession no. AY218602). DLE084 had 94% identity to soil bacterium clone 1417 from soil aggregates (Accession no. AY493956) (Table S2). GenBank accession numbers corresponding to 16S rDNA clone sequences are EF127594 - EF127627.
19. The possibility that these dramatic differences in DNA resulted from shearing can be ruled out, as all ice samples were processed identically, with no harsh physical treatments and minimal vortexing. Furthermore, genomic DNA extracted from a culture of *Escherichia coli* produced similar DNA sizes to those observed for our ‘young’ ice sample.
20. E. Willerslev, A. J. Hansen, H. N. Poinar, *Trends Ecol. Evol.* **19**, 141 (2004).
21. E. Willerslev *et al.*, *Curr. Biol.* **14**, R9 (2004).
22. B. Hagedorn, R. S. Sletten, B. Hallet, E. J. Steig, *Geophys. Res. Abstracts* **8, No. 08476** (2006).
23. M. Eigen, L. deMaeyer, *Proc. R. Soc. London A.* **247**, 505 (1958).
24. T. Lindahl, *Nature* **362**, 709 (1993).
25. G. Lemaitre, M. S. Vallarta, *The Physical Review* **43**, 87 (1933).
26. J. C. Gosse, F. M. Phillips, *Quaternary Science Reviews* **20**, 1475 (2001).
27. The relative frequencies of functional gene groups in DLE-98-12 and EME-98-03 ice samples, shaped by their different environments and ages, would probably change with a more extensive sequencing of the

BAC and end-repaired shotgun libraries. The main point from this analysis is that remnants of diverse, recognizable metabolic genes and numerous ORFs of unknown function have 'survived' in this ancient ice.

28. E. V. Koonin, K. S. Makarova, L. Aravind, *Annu. Rev. Microbiol.* **55**, 709 (2001).
29. H. Ochman, J. G. Lawrence, E. A. Groisman, *Nature* **405**, 299 (2000).
30. A. H. Knoll, *Life on a young planet: the first three billion years of evolution on Earth* (Princeton University Press, Princeton, NJ, 2003), pp.
31. Special thanks to John Heidelberg and Bill Nelson at The Institute for Genomic Research (TIGR) for their assistance in metagenomic sequencing and analysis. We also thank Ron Lauck and Sybil Seitzinger for the chemical analyses, Felisa Wolfe-Simón and Costantino Vetriani for the help with the phylogenetic analyses, Valentin Starovoytov for SEM analysis, and Jörg Schäfer, Gordon Thompson and Michael Bender for helpful discussions on cosmic fluxes in Beacon Valley ice. This study was supported by The Gordon and Betty Moore Foundation (to PGF and KDB). Additional support was provided by the US National Science Foundation (OPP-0338291 to DRM) and the Korea Antarctic Research Program (PP-04104, -05001).

Supporting Online Material

Figure Legends

Figure 1. Geologic setting of buried ice from Mullins and Beacon Valleys in the Transantarctic Mountains and evidence of encased microbes. (A) Photograph of Beacon Valley with view to the northeast towards Taylor Glacier. Note surface-debris cover with polygons that caps buried ice throughout Beacon Valley. (B) Digital Elevation Model (DEM) showing the distribution of ice samples collected at various locations in Mullins (white box) and Beacon valleys. Ice between locations marked 1 and 7 spans an ~8 Myr transect (1: DLE-98-12, ~100 Ka; 2: DLE-98-11, ~200 Ka; 3: DLE-98-CS-1, ~300 Ka; 4: MCI-04-013, ~2 Ma; 5: EME-98-08, ~5-6 Ma; 6: EME-98-01, ~6-7 Ma; 7: EME-98-03, ~8 Ma; ages for 1-3 are based on ^{3}He cosmogenic-nuclide analyses of overlying cobbles and on surface velocity measurements from synthetic-aperture radar interferometry (*7*); ages for 4 and 7 are based on ^{40}Ar/^{39}Ar analyses of in-situ ash fall overlying the ice; and ages for 5 and 6 are based on extrapolation from ice-flow models). Inset panels in (B): left panel shows the location of Beacon Valley in relation to the Antarctic continent; right panel shows an overview of Beacon Valley and nearby Taylor Glacier. (C-D) Scanning electron micrographs and (E-F) SYBR-Gold epifluorescence micrographs of ice samples from DLE-98-12 (C,E) and EME-98-03 (D, F), illustrating the relative abundance of DNA-containing bacteria cells and their morphology compared to glacial till. DLE-98-12 possessed cocci-like cells scattered among mineral debris (arrows), whereby EME-98-03 had the ubiquitous presence of filamentous, sheath-like structures (arrows).

Figure 2. PCR amplification and phylogenetic analysis of encased microbial community DNA in ice samples. (A) Agarose gel images of extracted ice DNA and bacterial 16S rDNA PCR products, including MilliQ® contamination control. Lane 1: environmental DNA from DLE-98-12 ice meltwater; Lane 2: *Bacteria*-specific 16S rDNA PCR amplicons for DLE-98-12; (2a, raw; 2b, gel purified) and *E. coli* (2c, + control); Lanes 3: *Archaea*-specific 16S rDNA PCR amplicons for DLE-98-12 DNA (3a) and *Haloferax volcanii* (3b, + control); Lane 5: *Bacteria*-specific 16S rDNA PCR amplicons for EME-98-03 (5a) and *E. coli* (5b, + control). Lanes 4 and 6 correspond to PCR reagent negative controls. Lanes 7: *Bacteria*-specific 16S rDNA PCR amplicons for DLE-98-12. DNA template serially diluted x1 (7a), x5 (7b), x25 (7c), and x125 (7d). Lanes 8a to d, same as the Lane 7, but for the MilliQ® water contamination control. The lack of amplification in control samples argues against contamination during sample processing. M1 and M2, molecular weight ladders 0.1 and 1 kb, respectively. Ice DNA template was routinely diluted (>10x) to suppress the enzyme

inhibition. (B) Maximum-likelihood phylogenetic tree comparing 16S rDNA phylotypes from encased microbial communities (blue or red) to closely related, reference relatives (black). Numbers in parentheses indicate clone frequency. Phylotypes from the melt water cultures are in the yellow boxes. Phylogenetic trees were generated using the MOLPHY analysis program, followed by refined analysis and optimization using the BASEML feature of PAML. The scale bar represents 0.1 nucleotide substitutions per position.

Figure 3. Assessment of the metabolic activity and viability of microorganisms encased in DLE-98-12 and EME-98-03. (A-B) Incorporation and respiration of ^{3}H-thymidine, ^{3}H-leucine, and ^{14}C-glucose in ice meltwater samples during incubation at 4°C. Data is corrected for formalin-killed controls. Note difference in scale for ^{3}H-leucine incubation in (A) and ^{14}C-respiration in (A-B). (C-D) Time course of viable cell growth, as determined by optical density, in nutrient supplemented ice water cultures. Note the differences in scales, reflecting dramatic growth discrepancies between DLE-98-12 and EME-98-03. Symbols indicate the ice and the nutrient type (see methods). Arrows indicate time points when subsamples were removed from the cultures and used for PCR amplification of *Bacteria*-specific rDNA and plating onto solid media for the isolation of viable cells (E). Amplification and colony formation was only successful for DLE liquid cultures with recovered phylotypes given in Table S4. Despite repeated attempts, EME cultures did not amplify with *Bacteria*-, *Archaea*- or *Eukaryota*-specific 16S rDNA primers, nor did it yield isolated cells. (F) Nested PCR amplification verified the presence of the cultured isolates in the original ancient ice microbial community. A primer set specific to DLE011i and *Arthrobacter roseus* CMS90 (Arose 232F and 1008R; Methods) produced the expected 770bp fragment when the original *Bacterial* 16S rDNA amplicons of the DLE ice DNA was used as a template (Lane 1a) and to the genomic DNA of DLE011i (Lane 1b ,+ control). Similarly, nested PCR with an Arthro primer set (846F and 1286R; Methods) specific to *Arthrobacter* sp. produced the expected ~440 bp fragment when applied to the original *Bacterial* 16S rDNA amplicons (Lane 2a), and DLE011i genomic DNA (Lane 2b, + control). Lanes 3a and b show negative controls (*E. coli* genomic DNA) for the two primer sets.

Figure 4. Time-dependent DNA degradation for buried ice distributed through the Beacon and Mullins Valleys. A DNA 'kill curve' for ice samples collected in Beacon and Mullins valleys, illustrating the influence of geologic time on the weighted mean size (bp) of community DNA. The weighted mean DNA size was calculated by running the total extracted DNA on a 0.8% agarose gel, staining with SYBR Gold and obtaining a densitometry profile of the entire ice DNA using ImageJ

1.34s (National Institute of Health). Representative agarose gels of community DNA frm DLE-98-12 and EME-98-03 are provided above their respective data points, illustrating the dramatic differences in DNA size (DLE-98-12>15 kb and EME-98-03 ~210 bp). Densitometry data was imported to Excel to calculate the weighted mean (distance) of the DNA peak and DNA size (bp) was calculated using the distance-bp calibration for molecular weight standards. Error bars representing standard deviation of DNA size and ice age are provided, and where not visible are within the symbol size. Curve fit is based on a linear regression ($y= 16804e^{-0.0007x}$, $r^2=0.96$).

Figure 5. The genetic content of community ice DNA from DLE-98-12 and EME-98-03. (A) The distribution of gene categories for a random sampling of BAC (DLE-98-12; n=131 sequences) and end-repaired (EME-98-03; n=427 sequences) clone libraries, as determined by BLASTx analysis. A slight majority (~55-60 %) the sequences from both ice sources had BLASTx matches to a wide variety of metabolic genes (key), while others (~40-45%) had no statistical similarity to sequences in GenBank below the 10^{-5} e-value cut off. The overwhelming majority of BLASTx matches derived from the *Bacteria* (93.9% for DLE-98-12 and 98.7% for EME-98-03), supporting superior DNA preservation by members of this domain. The contribution (percent) for the top 5 categories is indicated. The remaining sequences with database matches derived from: Virus (3.7%, 3 sequences) and *Eukaryota* (2.4%, 2 sequences) for DLE-98-12; Archaea (0.4%, 1 sequence) and Eukaryota (0.9%, 2 sequences) for EME-98-03. (B) An analysis of the impact of sequence query size on the statistical similarity to sequences in GenBank (e-score). Given that the ice community DNA possessed small DNA fragments, especially EME-98-03 where the mean weighted DNA size was ~210 bp, query size might introduce an artifact against finding a BLASTx sequence match within the e-value criteria. The Log(e-score) per unit query length as a function of query length was plotted for each recovered ice sequence. Sequences for which no BLASTx hit was recovered are distributed along the x-axis.