# Awakening ancient polar *Actinobacteria*: diversity, evolution and specialized metabolite potential

Natalie Millán-Aguiñaga[1], Sylvia Soldatou[2], Sarah Brozio[2], John T. Munnoch[2], John Howe[3], Paul A. Hoskisson[2] and Katherine R. Duncan[2,*]

## Abstract

Polar and subpolar ecosystems are highly vulnerable to global climate change with consequences for biodiversity and community composition. Bacteria are directly impacted by future environmental change and it is therefore essential to have a better understanding of microbial communities in fluctuating ecosystems. Exploration of Polar environments, specifically sediments, represents an exciting opportunity to uncover bacterial and chemical diversity and link this to ecosystem and evolutionary parameters. In terms of specialized metabolite production, the bacterial order *Actinomycetales*, within the phylum *Actinobacteria* are unsurpassed, producing 10000 specialized metabolites accounting for over 45% of all bioactive microbial metabolites. A selective isolation approach focused on spore-forming *Actinobacteria* of 12 sediment cores from the Antarctic and sub-Arctic generated a culture collection of 50 strains. This consisted of 39 strains belonging to rare *Actinomycetales* genera including *Microbacterium, Rhodococcus* and *Pseudonocardia*. This study used a combination of nanopore sequencing and molecular networking to explore the community composition, culturable bacterial diversity, evolutionary relatedness and specialized metabolite potential of these strains. Metagenomic analyses using MinION sequencing was able to detect the phylum *Actinobacteria* across polar sediment cores at an average of 13% of the total bacterial reads. The resulting molecular network consisted of 1652 parent ions and the lack of known metabolite identification supports the argument that Polar bacteria are likely to produce previously unreported chemistry.

## INTRODUCTION

The bacterial domain is one of the most genetically diverse, pervasive and ubiquitous domains of life and is essential to all ecosystems. Since the introduction of streptomycin in 1944 for the treatment of tuberculosis, bacteria have been the primary source of bioactive natural products, often termed specialized metabolites. Indeed, between 1981 and 2014, 73% of all approved antibacterial agents were unaltered natural products, or natural product derivatives [1] many of which are produced by bacteria. In particular, bacteria of the order *Actinomycetales* (actinomycetes) are responsible for the production of more than 70% of antibiotics and are unsurpassed for chemical exploitation due to their ability to produce biologically active and chemically diverse metabolites [2–5]. Actinomycetes can be further divided into the genus

*Streptomyces* and all remaining genera are often referred to as 'rare actinomycetes' due to their less-frequent isolation [5–7]. A study of the genomes of 21 marine-derived rare actinomycetes revealed novel and diverse biosynthetic gene clusters (BGCs) coding for specialized metabolites; for example, the genera *Actinomadura* and *Nocardia* were reported to contain 44 and 38 BGCs per strain, respectively [8].

Previous studies have suggested that geographical isolation affects bacterial population structure, which can be investigated by estimating bacterial richness and the environmental factors responsible for species delineation [9]. The marine environment covers 70% of the Earth's surface, yet represents both the most biodiverse and the most understudied ecosystem. Isolating actinomycetes from the marine environment has previously been shown to provide new species

with pharmaceutical promise [10, 11]. For example, a *Saccharomonospora* strain isolated from marine sediment in California yielded the unprecedented alkaloid iodopyridone with cytotoxic activity against HCT-116 human colon cancer cells with an $IC_{50}$ of 3.6 μM [12]. Furthermore, a recently described species of the actinomycete *Pseudonocardia antitumoralis* was isolated from deep-sea sediment in the northern South China Sea and was found to be a natural producer of the synthetic antitumour agent deoxynyboquinone and three new derivatives, pseudonocardians A, B and C [10]. Additionally, phocoenamicins B and C from a *Micromonospora* strain isolated from Canary Island sediment showed antibiotic activity against methicillin-resistant *Staphylococcus aureus* (MRSA) (MIC: 4–64 μg ml⁻¹) and *Mycobacterium tuberculosis* H37Ra (MIC: 16–32 μg ml⁻¹) [13].

Of all marine ecosystems, Polar and sub-Polar ecosystems are highly vulnerable to global climate change with consequences for bacterial biomass, diversity and community composition [14]. Due to the severe climatic conditions and logistical limitations, Polar regions (Arctic and Antarctic marine environments) have remained poorly studied when compared with oceans worldwide [15]. However, research programmes such as the Natural Environment Research Council's Northern Seas Arctic Programme and the German Science Foundation's ANDEEP (Antarctic Benthic Deep-Sea Biodiversity) have furthered the investigation of marine sediment across the Polar regions through seabed photographs and sediment profile images from box, gravity and multi-core sediment samples [16, 17]. These can often represent geologically ancient sediments; for example, an Arctic core sample recovered recently from a water depth of 1126 m was ¹⁴C radiocarbon dated to approximately 26 900 years ago [17] and an Antarctic sediment core sample recovered from 4542 m was dated to 1542 years ago with ²⁰⁶/²⁰⁷Pb isotope ratios [16]. These sediments therefore represent an opportunity to explore ancient sources of novel bacterial diversity and thus chemical potential.

Recent studies have reported *Actinobacteria* from Polar regions through culture-dependent and culture-independent studies [15, 18, 19]. Using 454-sequencing technology it was found that *Actinobacteria* comprised 10 % of the bacterial community composition of Arctic sediment samples [20]. This was a similar finding to a study by Carr *et al.*, who analysed pyrosequencing data from Adélie Basin sediment in Antarctica, with *Actinobacteria* comprising approximately 10 % of the sequence reads [21]. Culture-dependent studies have been successful in isolating a number of Polar *Actinobacteria* strains. For example, Zhang *et al.*, isolated 152 strains which represented 18 *Actinobacteria* genera, including a potentially novel genus, from 21 Arctic sediment samples and González-Rocha *et al.* reported 80 bacteria isolated from Antarctic samples that were distributed across three main phyla, *Actinobacteria*, *Bacteroidetes* and *Proteobacteria* [15, 19]. In some cases, antibacterial and antifungal activity from Arctic actinobacterial strains including *Nocardiopsis* and *Streptomyces* species has also been reported with preliminary metabolomics data; for example, Antarctic *Streptomyces*

strains have been shown to be an exciting source of diverse and potentially novel chemistry [22, 23].

Here we report the comparative analysis of selectively isolated rare actinomycetes from both the Antarctic and sub-Arctic using both culture-dependent and metagenomic investigation. The aims of this study were to analyse the community composition and functional genes associated with depth profiles of sediment cores and assess specialized metabolite production of cultured isolates. This will lead to a greater understanding of the evolutionary and ecological processes that underpin the distribution of these biotechnologically important microorganisms in extreme deep-water Polar environments.

## METHODS

### Sediment core collection and transportation

Sediment cores were obtained from PS Polarstern (Antarctic) and from RRS James Clark Ross (Arctic) from 2002 and 2005 using a multiple (multi), box, piston and gravity corer. The multi-core samples used a hydraulically damped corer enabling undisturbed cores of up to 0.8 m length to be collected with the sediment–water interface preserved. Cores were extruded onboard and sliced, and sediment slices were preserved in sample bags at a constant temperature storage facility at 4 °C. Box core subsamples were collected by pushing a 0.5 m polycarbonate liner into the box corer sample (Antarctic only). Piston and gravity core samples were obtained using a 3 m (gravity) or 12 m (piston) barrel length. Cores were taken using a continuous polycarbonate liner, cut into 1 m length and split longitudinally onshore. All sediment samples were transported in cool storage to the Scottish Association for Marine Science where they were also stored at 4 °C constant temperature.

### Processing sediment cores

Twelve sediment cores were sampled from the sediment core collection at the Scottish Association for Marine Science (SAMS) in Scotland. The latitude, longitude, year of sample collection, Polar region, sediment depth, type of corer used and depth recovered/core length were also recorded (Table S1, available in the online version of this article). The sediment cores were processed by aseptically removing sediment samples from the upper, middle and lower sections of each core into a Falcon tube (50 ml), which was stored at −20 °C until required. The 36 sediment samples were pretreated using the previously described stamping method used by Mincer *et al.* [24] with the following modifications: approximately 1 g of sediment was aseptically placed into a sterile Petri dish, dried over 48 h in a laminar flow hood, ground lightly with a spatula, pressed into a sterile foam (20 mm in diameter), and inoculated onto three different agar media by stamping six or seven times in a circular fashion, giving a serial dilution effect to target the isolation of spore-forming actinomycetes only. The three media were A1, SW and SC, all containing Instant Ocean (18 g l⁻¹), cycloheximide (25 μg ml⁻¹) and nystatin (25 μg ml⁻¹). The media compositions were as follows:

A1 [starch, 10 g l$^{-1}$; yeast extract, 4 g l$^{-1}$; peptone, 2 g l$^{-1}$; agar, 14 g l$^{-1}$), SW (agar, 14 g l$^{-1}$) and SC (starch, 10 g l$^{-1}$; KNO$_3$, 2 g l$^{-1}$; K$_2$HPO$_4$, 2 g l$^{-1}$; casein, 0.3 g l$^{-1}$; MgSO$_4$·7H$_2$O, 0.05 g l$^{-1}$; CaCO$_3$, 0.02 g l$^{-1}$; FeSO$_4$·7H$_2$O, 0.01 g l$^{-1}$; agar, 18 g l$^{-1}$) [24, 25]. All bacterial culture work was carried out under aerobic conditions, at room temperature, as no growth was observed after 5 months at 4 °C. All strains were isolated within 4 weeks of incubation, subcultured until pure, cryopreserved with glycerol (20 %) and stored at −80 °C.

## 16S rRNA gene amplification and phylogenetic analysis

Colony-PCR amplification of the 16S rRNA gene was carried out for most of the strains using the primers: FC27 (5′ AGAGTTTGATCCTGGCTCAG 3′) and RC1492 (5′ TACG GCTACCTTGTTACGACTT 3′) [26]. When 16S rRNA gene amplification did not work using colony-PCR, DNA extractions were performed using the ISOLATE II Genomic DNA kit (Bioline). Thermocycler conditions were 95 °C for 5 min, 30 cycles of 94 °C for 10 min, 55 °C for 30 s and 72 °C for 3 min; followed by 8 min at 72 °C then 1 min at 20 °C. The partial and nearly complete 16S rRNA gene sequences (approximately 1355 bp) were analysed using the Basic Local Alignment Search Tool (BLAST) on the NCBI database [27] and EzTaxon for the closest type strain [28]. Sequences were checked for accuracy, trimmed and a multiple alignment was created using Geneious Pro 9.1.5 [29]. Maximum-likelihood (ML) trees were created using RAxML [30] implemented on the CIPRES Portal v2.2 at the San Diego Supercomputer Centre [31]. Analysis included 1000 bootstrap replicates using the most complex model GTR+GAMMA for bootstrapping and final ML optimization using default parameters settings. The tree was visualized with FigTree v1.3.1 (http://tree.bio. ed.ac.uk/software/figtree). The 16S rRNA gene sequences of the strains isolated in this study were deposited in GenBank under accession numbers MH725277–MH725319.

## Fermentation and metabolite extraction

Fifteen strains were selected based on isolation location and phylogeny for metabolite analysis. These strains were pre-cultured (10 ml, 28 °C, 200 r.p.m. for 10 days) in ISP2 media (yeast extract, 4 g l$^{-1}$; malt extract, 10 g l$^{-1}$; glucose, 4 g l$^{-1}$; Instant Ocean, 18 g l$^{-1}$) [32]. Each fermentation [ISP2 media (50 ml), activated HP-20 resin (Sigma) (2.5 g)] was inoculated with a 5 % (v/v) preculture and fermented for 10 days (28 °C, 200 r.p.m) in duplicate. Each fermentation was centrifuged (4000 r.p.m, 10 min) and the supernatant discarded. The cell/HP-20 resin pellet for each strain was then frozen (−80 °C, overnight) and lyophilized (MicroModulyo, ThermoSavant) until dry. The lyophilized pellet was then extracted with ethyl acetate (20 ml, 100 r.p.m, 25 °C) for 1 h, and then this was repeated a second time. The two extracts were then combined, dried *in vacuo* and weighed.

## Mass spectrometry

High-resolution tandem mass spectrometry (HR-MS/MS) data were generated for the 30 crude bacterial fermentation extracts (15 strains in duplicate), ISP2 medium and solvent controls. The EtOAc extracts were prepared at 1 mg ml$^{-1}$ in MeOH and MS/MS experiments were carried out using a Finnigan LTQ Orbitrap coupled to a Surveyor Plus HPLC pump (Thermo Fisher) in positive ionization mode as previously described [23, 33] with the following modifications: an MS range of $m/z$ 100–15000; an MS2 range of $m/x$ 200–1500, an MSn range of $m/z$ 0–1000 and 30000 resolution.

## Molecular network

LC-MS data were acquired using Xcalibur version 3.1 and converted from .raw to .mzXML file format using the Trans-Proteomic pipeline [34], and clustered to create a molecular network with MS-Cluster using Global Natural Products Social (GNPS) Molecular Networking [35]. The data were filtered by removing all MS/MS peaks within 17 Da of the precursor $m/z$. MS/MS spectra were window filtered by choosing only the top six peaks in the 50 Da window throughout the spectrum. The data were then clustered with MS-Cluster with a parent mass tolerance of 0.02 Da and an MS/MS fragment ion tolerance of 0.2 Da to create consensus spectra. Furthermore, consensus spectra that contained fewer than one spectrum were discarded. A network was then created where edges were filtered to have a cosine score above 0.7 and more than six matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least six matched peaks. To visualize the data, they was imported into Cytoscape suite (version 3.6.1) [36], and nodes and edges were displayed (nodes correspond to a specific consensus spectrum; edges represent significant pairwise alignment between nodes). Cosine similarity scores, ranging from 0 to 1 (1 being identical spectra), were computationally combined as consensus spectra if more than six ion fragmentation spectra matched. A minimum cosine score of 0.7 was selected to subdue the clustering of different compound classes under the same molecular family when visualizing the data as a network of nodes connected by edges [37].

## Sediment core metagenomics using MinION sequencing

DNA was extracted from the upper and lower sections from two sediment cores (one Antarctic and one Arctic) using an ISOLATE II Genomic DNA kit (Bioline). The protocol was modified slightly to include 250 mg of sediment with a pre-lysis incubation of 3 h at 55 °C. Metagenomic DNA quality and quantity was measured using a NanoDrop spectrophotometer and Qubit. The four libraries were prepared using the Rapid Sequencing kit (SQK-RAD003) using at least ~400 ng of template DNA. The flow cell priming kit was used (EXP-FLP001) for sequencing and custom MinKNOW (version 1.13.1) scripts allowed between 7 and 24 h sequencing runs
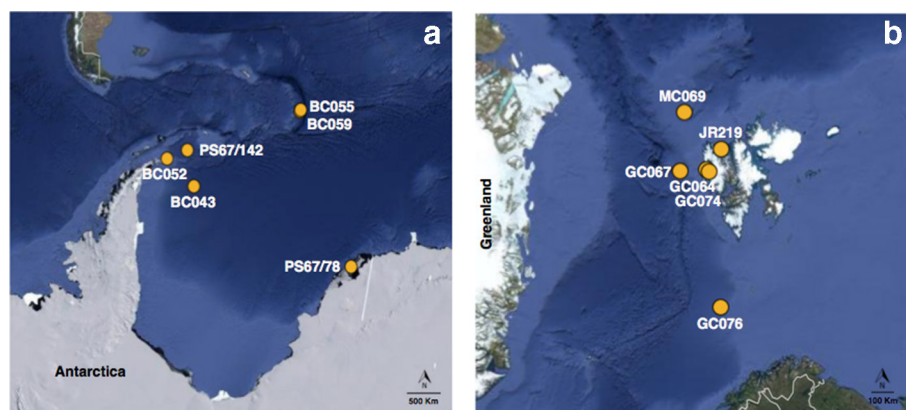
**Fig. 1.** Sediment core locations. (a) Six sediment cores named BC052, BC059, BC043, PS67/142, PS67/78 and BC055 from Antarctica. (b) Six sediment cores named GC067, GC076, JR219, MC069, GC064 and GC074 from the sub-Arctic were investigated in this study.

to be performed with live basecalling. Custom MinKNOW (version 1.13.1) was setup to automatic quality score and transfer the reads into fastq sequences. The generated fastq sequence data were analysed using the metagenomic analysis server MG-RAST, which assigns taxonomy based on predicted proteins and rRNA genes [38]. The MetaPathways pipeline was used for taxonomic profiling [39] using the fastq data as input files. Stacked bar graphs were generated using Microsoft Excel.

## RESULTS

### Culturable bacterial diversity and phylogeny

Six sediment cores from Antarctic and six sediment cores from sub-Arctic locations were sampled to investigate the actinobacterial diversity using a selective (for spore-forming *Actinobacteria*) culture-dependent analysis (Fig. 1, Table S1). Using a room-temperature selective-isolation approach (for spore-forming actinomycetes), a total of 50 strains were isolated based on actinomycete-like morphology. When these were sequenced they belonged to three phyla: *Actinobacteria*, *Proteobacteria* and *Firmicutes* (Fig. 2, Table S2). A total of 27 strains were isolated using the rich nutrient medium A1, comprising 10 genera, three of which were exclusively isolated on this medium. Additionally, minimum medium SW recovered 16 strains affiliated to six genera, with one genus exclusively isolated on this medium (Table S2). No isolates were observed on any of the plates processed at 4 °C.

A total of 78% of the strains isolated belonged to the phylum *Actinobacteria* and all were taxonomically assigned to rare actinomycete genera, as no *Streptomyces* strains were isolated. The majority (34/39) of these strains were isolated from the six Antarctic sediment cores and they represented greater phylogenetic diversity, spanning five genera (*Pseudonocardia*, *Salinibacterium*, *Agrococcus*, *Microbacterium* and *Rhodococcus*). In comparison, only five strains, assigned to three *Actinobacteria* genera (*Microbacterium*, *Dietzia* and *Rhodococcus*) were isolated from the six sub-Arctic sediment cores.

In total, 20% of the strains belonged to the phylum *Proteobacteria* and encompassed the four genera *Pseudomonas*, *Halomonas*, *Sphingorhabdus* and *Methylobacterium*. Within this phylum, only the genus *Sphingorhabdus* was isolated from both Polar regions, while the remaining three were exclusively from Arctic samples. The final 2 % of the strains belonged to the phylum *Firmicutes*, with only one genus (*Bacillus*) isolated from one Arctic core. Isolation of strains was attempted from the upper, middle and lower sections of each sediment core, depending on core length, which varied from 16 to 86.5 cm (Table S1). When comparing these sediment depth profiles, the lower (deepest) section of the Antarctic sediment cores represented the highest cultured bacterial diversity, with all five *Actinobacteria* genera represented. This was in comparison with the lower section of the Arctic sediment cores where only one *Actinobacteria* genus was isolated (Tables S1 and S2).

Interestingly, only one *Actinobacteria* genus, *Rhodococcus*, was isolated from both Antarctic and Arctic sediment cores (Fig. 2). Therefore, in terms of culturable actinobacterial diversity, the majority of strains isolated were location specific. In particular, the sediment cores from the Antarctic sites BC043 and BC052 yielded four genera within the *Actinobacteria* (*Pseudonocardia*, *Salinibacterium*, *Agrococcus* and *Rhodococus*), in comparison to the Arctic core GC067, which yielded two genera (*Microbacterium* and *Rhodococcus*) within this phylum.

A total of 25 *Pseudonocardia* strains were only isolated from three (of the six) Antarctic sediment cores (BC043, BC052 and BC059). Interestingly, these strains grouped into two sister-clades (Fig. 2), one of which comprised 17 strains, 16 from BC043 and one from BC059. The second comprised seven strains, all from BC052. A third clade consisted of one strain from BC043. This clade grouping suggests a pattern between core and phylogeny. However, the characteristics of this are too complicated to suggest a single hypothesis as it could be a result of a number of factors (depth, collection time, location, etc.). The highest number of *Pseudonocardia* strains were
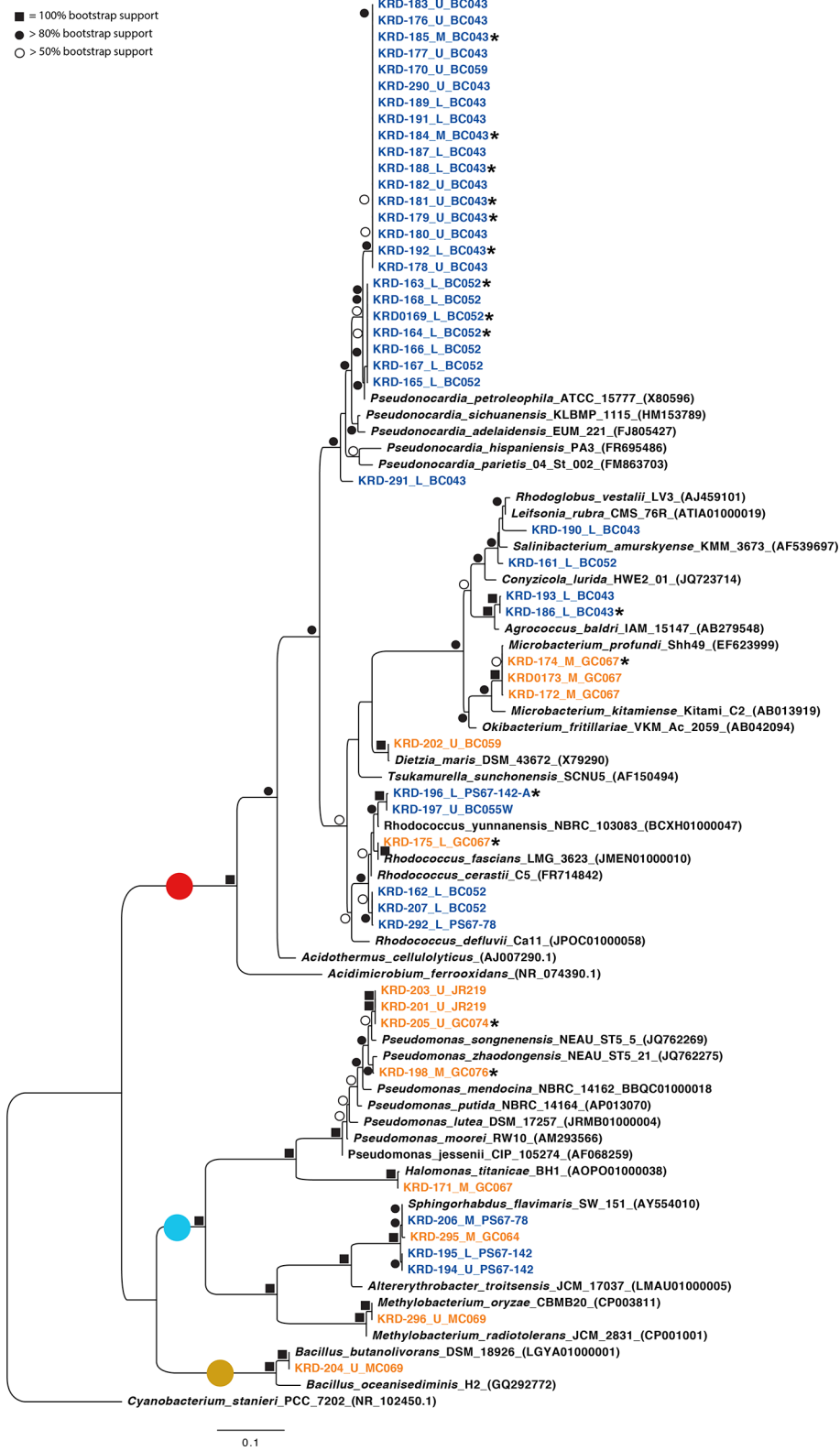
**Fig. 2.** ML tree based on 16S rRNA gene sequences of 50 strains isolated from Antarctic (blue) and Arctic (orange) sediment cores. Colour-coded strain numbers are followed by the section they were isolated: U: upper, M: medium, L: lower; and the number of the sediment core location. The asterisk next to strains indicates that were selected for chemistry analyses. Included are closest type strains and accession numbers. Bootstrap values are indicated for various ranges of support for ML and neighbour-joining analysis. The circles at the beginning of the different clades represent different phyla: red, *Actinobacteria*; blue, *Proteobacteria*; yellow, *Firmicutes*.

isolated using A1 medium, yielding 13 strains, in comparison to SW medium, which yielded 11 isolates (Fig. 2, Table S2). The almost-complete 16S rRNA gene sequences (~1355 bp) from the 25 *Pseudonocardia* strains shared 98.88–99.8% similarity with the closest type strains. Phylogenetically the type strains form three different clades, revealing a well-supported (80% bootstrap support) bifurcation within the *Pseudonocardia* clades (Fig. 2), suggesting that they could represent new species. A more robust tree including sequences from all 59 previously described *Pseudonocardia* species verified the presence of the three Antarctic clades from this study (Fig. S1). The first clade included 14 *Pseudonocardia* strains with 76% bootstrap support, a second clade with six *Pseudonocardia* strains with 85% bootstrap support and a last clade with two *Pseudonocardia* strains with 100% bootstrap support (Fig. S2), suggesting new species within this genus.

When comparing the number of cultured strains from each core, two-thirds of the isolates [29] were obtained from two Antarctic sediment cores BC043 and BC052, with the remaining 20 isolates obtained from the remaining four Antarctic and six sub-Arctic cores (Table S2). It is interesting to note that BC043 and BC052 are the cores closest to the Antarctic peninsula (Fig. 1) and are the deepest cores at greater than 4000 m (Table S1). However, too many variables exist between sediment cores to draw a comparison. These variables include, for example, year of collection, depth, pressure and temperature.

## Metagenomic analysis of bacterial diversity and function

Bacterial diversity was examined using MinION sequencing due to the long sequence reads and potential applications within the field. Initially, DNA extraction was attempted from eight samples from the upper and lower sections of four cores (two Antarctic and two sub-Arctic). However, only DNA from two Polar sediment cores and two different sections of each core (upper and lower) were obtained in higher quality yet low yields (GC067-U and GC067-L: 45.1 ng µl⁻¹; BC043-U: 29.1 ng µl⁻¹; and BC043-L: 129.2 ng µl⁻¹). It is important to note that replicates were not possible for these cores, due to the low yield of extracted DNA, and therefore only data from one sequencing run have been included. Sequence counts generated ranged from 265 to 7408 (Table S3), with eukaryotic sequence reads in higher abundance (>94%) in both sediment cores and across both sections. Bacteria reads varied between 1 and 5 % of the total reads (Fig. 3a). Viral and archaeal reads were found in low abundance (<1%) in some samples and virus reads were only observed in the upper section of the Antarctic sediment core, while archaeal reads were observed in the lower section of both polar sediment cores (Fig. 3a).

Of the 185 total taxonomic microbial reads, the phyla *Proteobacteria*, *Firmicutes*, *Actinobacteria*, *Euryarchaeota* and *Tenericutes* were represented in the sediment cores samples. The most representative group in both sediments cores was the phylum *Proteobacteria*, with 15–39% of the reads, followed by

the phylum *Actinobacteria* with 4–32% of the reads. Within the phylum *Proteobacteria* there were taxonomically diverse classes such as *Gammaproteobacteria* present across all four sediment samples. *Deltaproteobacteria* was only found in the upper section of the Arctic sediment sample, and the family *Enterobactereacea* was only found in the Antarctic sediment core (Fig. 3b). Within the phylum *Actinobacteria*, the genus *Mycobacterium* was present in all the samples with higher abundance in the Arctic than in the Antarctic (Fig. 3b). The phylum *Firmicutes* was found in the lower section of both polar sediment cores and the phylum *Tenericutes* was only found in the lower section of the Antarctic sediment core (Fig. 3b).

Analysis of the 365 functional reads with predicted protein functions annotated to the KO (KEGG Orthology) database showed the most abundant functional categories in both Polar sediment cores were genetic processing and environment information with an average of 34% (Fig. 4). Genes found within the category of genetic processing included reads involved in transcription, translation, replication and repair in contrast to the category of environmental information that includes reads involved in membrane transport, signalling molecules and interaction. The category of organismal systems was found in both polar sediment cores in low abundance (~1% of reads). While both the taxonomic and the functional reads are of interest due to the extreme nature of the cores, it is important to note that here they serve only as a preliminary overview, as the lack of quality DNA prevented replicates and statistical analysis.

## Specialized metabolite potential

Fifteen strains were selected for metabolite analysis based on isolation location and phylogeny (shown by an asterisk in Fig. 2). The selected strains belonged to the genera *Pseudonocardia*, *Microbacterium*, *Agrococcus*, *Rhodococcus and Pseudomonas* and were isolated from both Antarctic and Arctic sediment samples. The resulting fermentation extracts were assessed for bioactivity against both Gram-positive and Gram-negative bacterial pathogens, with no observed antibiotic activity (data not shown). However, analysis of the MS/MS data by bacterial genus resulted in a molecular network containing 1652 parent ions (nodes) (Fig. 5). The node size represents the number of strains producing each parent ion and, as expected, media components have a larger node size as they are omnipresent in bacterial extracts. Due to the imbalance in the number of strains per genus, the number of nodes was averaged to nodes-per-strain for each genus. For example, 620 nodes were attributed to eight *Pseudonocardia* strains and 52 parent ions were produced by one *Microbacterium* strain. Taking into consideration the different number of strains per genus, the number of nodes attributed to one *Pseudonocardia* strain averaged 69, whereas the parent ions produced by the sole *Microbacterium* strain remain unaltered [40]. Interestingly, the genera *Pseudonocardia*, *Rhodococcus* and *Pseudomonas* were represented by a similar number of nodes (69, 76 and 77, respectively) whereas the genera *Microbacterium* and *Agrococcus* were shown to produce a
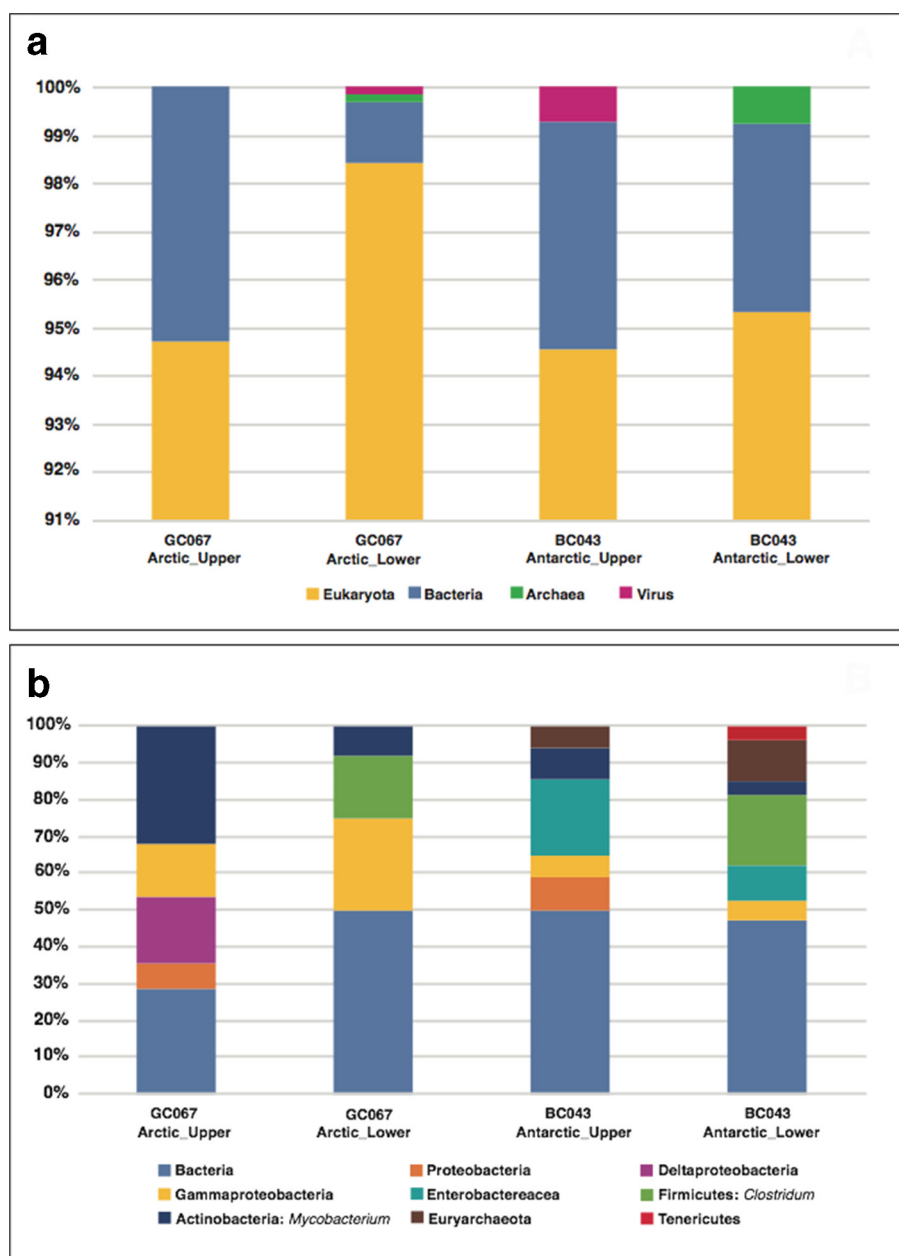
**Fig. 3.** Taxonomic profiling of two polar sediment cores and two different sections using MinION sequencing data. (a) Distribution of taxa using MG-RAST analysis. (b) Taxonomic analysis using the MethaPathways algorithm.

smaller number of parent ions (52 and 59, respectively). No parent ion within the molecular network matched any known compounds present in the GNPS libraries, indicating the great potential of these Polar bacteria to produce novel chemistry. However, manual data analysis using the Natural Product Atlas, a newly launched open access database designed to cover all microbially derived natural products (http://www.npatlas.org), revealed the presence of several candidate matches. This included a parent ion (*m/z* 691.387 [M+Na]$^+$) produced by one Antarctic *Agrococcus* strain (KRD 186) which was the sodium adduct of *m/z* 669.40521 ([M+H]$^+$). Based on its High-Resolution ElectroSpray Ionisation Mass

Spectrometry (HRESIMS) (*m/z* 668.3983), the molecular formula of the non-charged ion was identified as C$_{32}$H$_{60}$O$_{14}$. A search of the accurate mass together with the molecular formula generated one Natural Product Atlas hit which corresponded to the known glycoglycerolipid GGL3 previously reported from a *Microbacterium* species associated with the marine sponge *Halichondria panacea* [41] (Fig. 5). The molecular network further supported the presence of the known glycoglycerolipid, as the node corresponding to GGL3 was structurally related (cosine score >0.9) to a parent ion (*m/z* 677.371 [M+Na]$^+$), which was produced by more than one strain. Specifically, this parent ion is present in
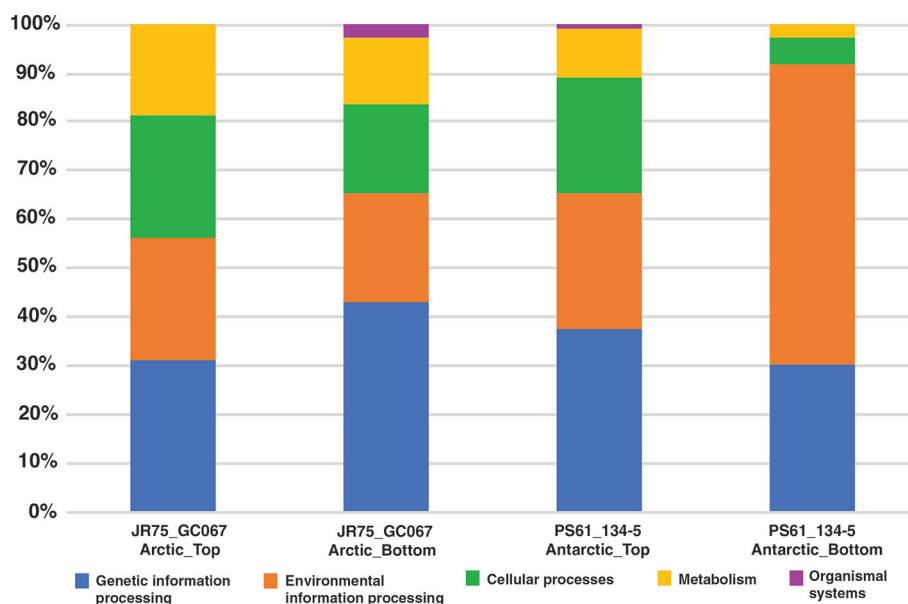
**Fig. 4.** Distribution of functional categories using the reads with predicted protein functions annotated to the KO (KEGG Orthology) database, using MG-RAST results.

extracts derived from the Antarctic *Agrococcus* strain (KRD 186) which produces GGL3 and the Arctic *Microbacterium* strain (KRD 174). The parent ion *m/z* 677.371 [M+Na]⁺ is the sodium adduct of *m/z* 655.38922 ([M+H]⁺). Based on its HRESIMS, the molecular formula of the non-charged ion was identified as $C_{31}H_{58}O_{14}$, which agrees with a mass difference of 14 calculated between the two parent ions. Dereplication of the accurate mass in combination with the generated molecular formula yielded a known haemolysin [1] (Fig. 5) originally isolated from the green alga *Ulva pertusa* [42]. The fragmentation pattern agreed with the dereplication results, which further confirmed the presence of this haemolysin [1]. The third node (*m/z* 755.404) of the molecular family shown in Fig. 5 did not correspond to any previously reported compound, which could suggest that *Agrococcus* strain KRD 186 produces a potentially new metabolite that is structurally related to this glycolipid family.

To further investigate the potentially undiscovered compounds produced by these polar bacterial strains, the Antarctic sediment core BC043 was selected as it was dated to be almost 2000 years old. In particular, six *Pseudonocardia* strains (two from each core section) were chosen to assess the influence of depth (time scale) on the specialized metabolites observed as they were isolated across all three sections (U, M, L) of the sediment core (Fig. 6). The relatively small number of strains per core section used in this analysis was due to the limited number of *Pseudonocardia* strains isolated from the middle core section. Specifically, strains KRD 179 and KRD 181 were isolated from the upper section of the sediment core whereas strains KRD 184 and KRD 185, and KRD 188 and KRD 192 were isolated from the middle and lower core sections, respectively. The number of parent ions

produced (i.e. non-media components) represented 58.2% of the nodes in the molecular network and was found to vary across the three sediment core sections, as *Pseudonocardia* strains isolated from the upper and lower section of the sediment core respectively produce 18.6% and 17.8% of the total number of nodes while the two strains derived from the middle section are accountable for only 9.8% of the total number of nodes present in the molecular network. As shown in Fig. 6, only a small number of nodes (orange) are shared between more than two strains. Interestingly, 11.9% of the total parent ions are shared between strains isolated from the upper, middle and lower section of sediment core BC043, with only 3.3% of the produced parent ions (i.e. non-media components) shared amongst strains isolated from all three sections of the core, demonstrating the chemical diversity and potential across core depth (and evolutionary time).

## DISCUSSION

The isolation of *Actinobacteria* from understudied marine sediment provides an opportunity to find novel species that can lead to new chemistry [43, 44]. As aerobic conditions were used for culture, it suggests that the spores isolated may not be metabolically active in the anaerobic sediment cores. In this study, rare actinomycete strains were cultured, including 25 Antarctic *Pseudonocardia* strains. The genus *Pseudonocardia* consists of 59 recognized species according to the LSPN (List of Prokaryotic names with Standing in Nomenclature, (http://www.bacterio.net/) and they have previously been isolated from environments such as deep-sea sediment [10], leafcutter ants [45], the Atacama desert [46] and in association with plants [47]. One previous study
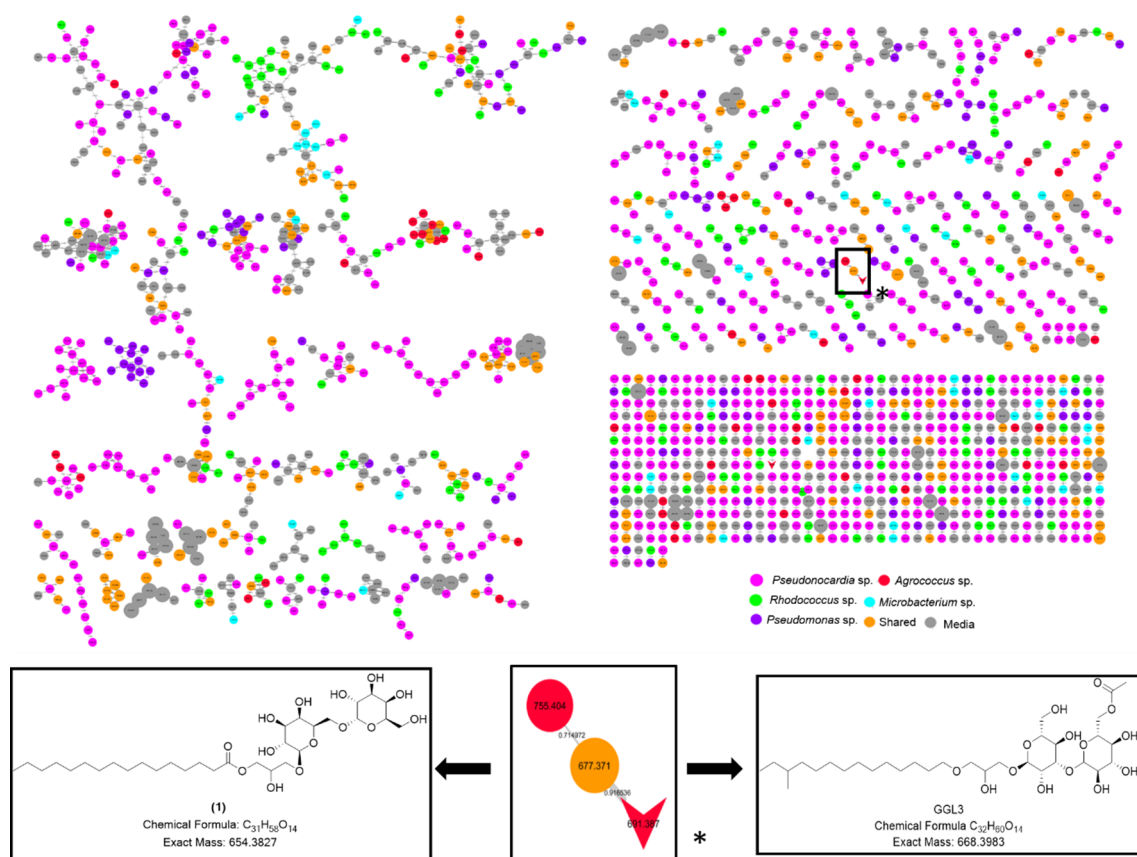
**Fig. 5.** Molecular network of 1652 parent ions produced by 15 bacterial strains (9 *Pseudonocardia* species, 1 *Microbacterium* species, 2 *Rhodococcus* species, 1 *Agrococcus* species and 2 *Pseudomonas* species). Nodes are colour-coded based on bacterial genus (pink: *Pseudonocardia*, blue: *Microbacterium*, red: *Agrococcus*, green: *Rhodococcus* and purple: *Pseudomonas*). Grey nodes represent media components whereas orange nodes represent parent ions that are produced by more than one different genus. Node size reflects the number of strains that produced each parent ion. Nodes highlighted in boxes represent parent ions that are identified as previously discovered metabolites. The reverse triangle shaped node reflects the sodium adduct of GGL3.

has reported the isolation of one *Pseudonocardia* strain from the McMurdo Dry Valleys region of Antarctica and this was described as a new species, *Pseudonocardia antarctica* [48]. The type strain of *Pseudonocardia sediminis* isolated from a marine South China Sea sediment sample [49] was the closest relative of KRD-291 within this study. The 24 remaining *Pseudonocardia* strains isolated were related to the type strains of *Pseudonocardia petrolephila* and *Pseudonocardia serianimatus,* the latter of which was isolated from the leaves of *Artemisia annua* L. [50]. The highly conserved nature of the 16S rRNA gene has resulted in its application for species delineation. The sequence similarity cut-off value for this has increased from <97% in 1994 [51] to 98.63% in 2006 [40]. More recent studies have proposed that this value is not applicable to all genera (e.g. *Mycobacterium*, *Salinispora*) as it has been suggested that any nucleotide change in the 16S rRNA gene (>98.65 %) may be meaningful from a taxonomic perspective [52–55]. Although further analysis will need to be conducted, comparing the phylogenetic distance and 16S rRNA gene sequence similarity values of these strains with all reported strains within this genus (including *P. antarctica*)

suggests that the *Pseudonocardia* strains from this study may represent novel species.

The Arctic sediment core GC067 was collected in a water depth of 1226 m with 3.80 m recovered and was radiocarbon dated by Howe *et al* [56] between 8287 (upper) and 26000 (lower) years old. This age indicates that the majority of the core was deposited during the pre-Last Glacial Maximum (LGM) (mid-Weichselian) and LGM (Late Weichselian) period [56]. Culture-dependent studies from this ancient core revealed three genera of the phylum *Actinobacteria* (*Halomonas*, *Microbacterium* and *Rhodococcus*) (Fig. 2). The bacteria isolated from this core have essentially been preserved (in a slow or dormant state). Culture-independent studies of the same core showed one of the highest reads for eukaryotes (94–98% of the total reads) corroborating the description of this core by Howe *et al.* [56] with a high abundance of dinoflagellate cysts that is consistent with possible high seasonal productivity [56, 57]. Antarctic sediment core BC043 was collected in a water depth of 4060 m with 50 cm of sediment recovered from the Weddell Abyssal Plain area.
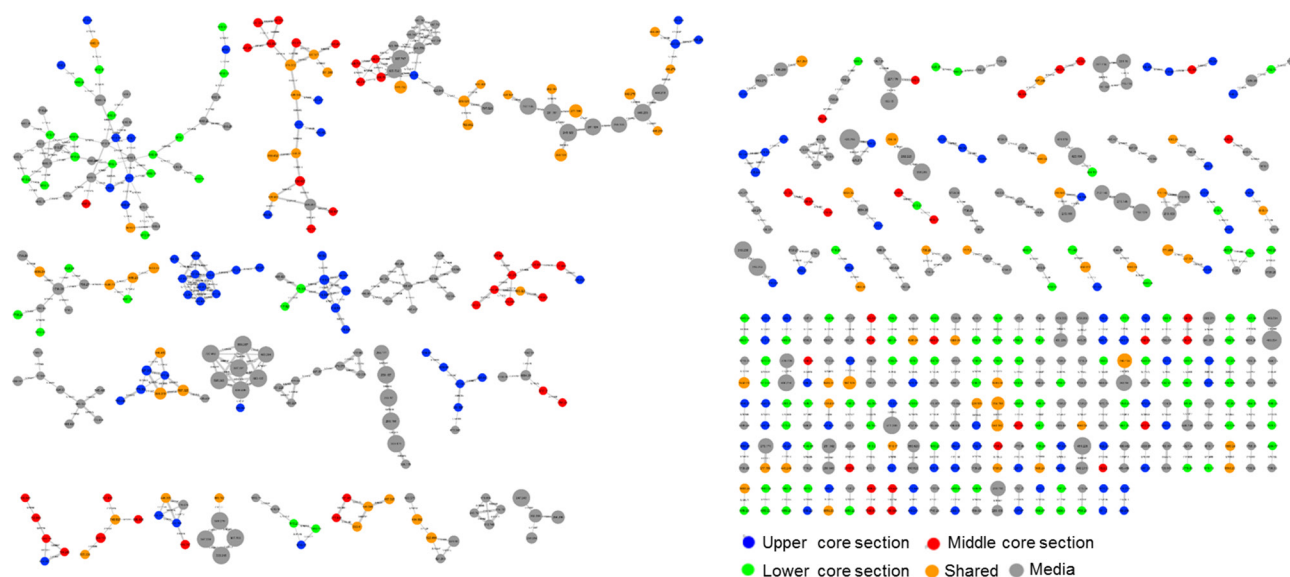
**Fig. 6.** Molecular network of 629 parent ions produced by six *Pseudonocardia* strains isolated from sediment core BC043 from the upper (KRD 179, KRD 181), middle (KRD 184, KRD 185) and lower (KRD 188, KRD 192) sections of the core. Nodes are colour-coded based on core section (blue: upper, red: middle, green: lower). Grey nodes represent media components whereas orange nodes represent parent ions that are produced by strains from more than one sediment core section. Node size reflects the number of strains that produced each parention.

The sediment was described by Howe *et al.* [16] as greenish grey silt and clay with pure siliceous biogenic material in the form of frustules, centric diatoms and radiolaria with a diatomaceous-rich phytodetritus 'mat' on the surface of the box core [16]. The high abundance of eukaryote (diatoms) reads in the sediment core (~95% of the reads) confirms the description mentioned above; knowing that diatoms are the only group of eukaryotic microalgae with a diplontic life history dominated by a duplicated genome [58]. Howe *et al.* [16] estimated the age of this sediment core based on the isotope $^{210}$Pb to 1542 years old [16]. Culture-dependent studies of this sediment core resulted in the isolation of three *Actinobacteria* genera (*Pseudonocardia, Agrococcus, Salinibacterium*). Both the sediment core of the Arctic and Antarctic have been dated to be more than 1500 years old with a depth of more than 1000 m, thus representing a remarkable opportunity to culture bacterial life. In particular, the *Pseudonocardia* strains that have been isolated from the Antarctic sediment have revealed potentially novel species and will greatly aid our understanding of this relatively understudied genus.

Molecular networking has been proven as a valuable approach for time-effective chemical dereplication and the discovery of new specialized metabolites when dealing with complex mixtures such as crude extracts of bacterial cultures [35, 59, 60]. In this study, a representative sample of 15 strains were selected based on isolation location (Polar region, depth, core section) as well as phylogeny for chemical investigation. In particular, these strains belonged to the genera *Pseudonocardia, Microbacterium, Agrococcus, Rhodococcus* and *Pseudomonas*, which were isolated from depths ranging from 379 to 4539 m. Using the GNPS platform, a

molecular network of 1652 parent ions was built based on bacterial genus. Data analysis of the molecular network and the acquired MS/MS spectra revealed the presence of two previously reported metabolites, GGL3 and compound 1 (Fig. 5), originally isolated from a *Microbacterium* species living symbiotically with the marine sponge *Halichondria panacea* collected off the Adriatic coast [41] and the green alga *Ulva pertusa* [42], respectively. In our dataset, GGL3 was produced by an Antarctic *Agrococcus* strain (KRD 186) isolated from sediment core BC043 which was collected at 4060 m below sea level whereas compound 1 (Fig. 5) was identified at a depth of 1226 m. Moreover, chemical diversity and potential across core depth (and evolutionary time) was demonstrated by analysis of *Pseudonocardia* strains isolated across all three sections (U, M, L) of sediment core BC043 (Fig. 6). Despite the small sample of strains which was due to limited *Pseudonocardia* strains isolated from the middle section of the core, the molecular network obtained suggests that depth may influence chemical diversity. A recent report covering the literature since 2001 demonstrated that only 29 new compounds have been discovered from Antarctic and (sub-) Arctic bacteria, of which 13 were isolated from marine actinomycetes [61]. The lack of known metabolite identifications within this study supports the theory that Polar bacteria have considerable chemical potential.

## References

1. Newman DJ, Cragg GM. Natural products as sources of new drugs from 1981 to 2014. *J Nat Prod* 2016;79:629–661.

2. Bérdy J. Thoughts and facts about antibiotics: where we are now and where we are heading. *J Antibiot* 2012;65:385–395.

3. Qin Z, Munnoch JT, Devine R, Holmes NA, Seipke RF *et al*. Formicamycins, antibacterial polyketides produced by *Streptomyces formicae* isolated from African *Tetraponera* plant-ants. *Chem Sci* 2017;8:3218–3227.

4. Seipke RF. Strain-level diversity of secondary metabolism in Streptomyces albus. *PLoS One* 2015;10:e0116457–14.

5. Bérdy J. Bioactive microbial metabolites. *J Antibiot* 2005;58:1–26.

6. Baltz RH. Marcel Faber roundtable: is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *J Ind Microbiol Biotechnol* 2006;33:507–513.

7. Lazzarini A, Cavaletti L, Toppo G, Marinelli F. Rare genera of actinomycetes as potential producers of new antibiotics. *Antonie van Leeuwenhoek* 2000;78:399–405.

8. Schorn MA, Alanjary MM, Aguinaldo K, Korobeynikov A, Podell S *et al*. Sequencing rare marine actinomycete genomes reveals high density of unique natural product biosynthetic gene clusters. *Microbiology* 2016;162:2075–2086.

9. Jensen PR, Mafnas C. Biogeography of the marine actinomycete Salinispora. *Environ Microbiol* 2006;8:1881–1888.

10. Tian XP, Long L-J, Li S-M, Zhang J, Xu Y *et al*. Pseudonocardia antitumoralis sp. nov., a deoxynyboquinone-producing actinomycete isolated from a deep-sea sediment. *Int J Syst Evol Microbiol* 2013;63:893–899.

11. Gerwick WH, Moore BS. Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chem Biol* 2012;19:85–98.

12. Maloney KN, Macmillan JB, Kauffman CA, Jensen PR, Dipasquale AG *et al*. Lodopyridone, a structurally unprecedented alkaloid from a marine actinomycete. *Org Lett* 2009;11:5422–5424.

13. Pérez-Bonilla M, Oves-Costales D, de la Cruz M, Kokkini M, Martín J *et al*. Phocoenamicins B and C, new antibacterial Spirotetronates isolated from a marine *Micromonospora* sp. *Mar Drugs* 2018;16:95–13.

14. Makhalanyane TP, Van Goethem MW, Cowan DA. Microbial diversity and functional capacity in polar soils. *Curr Opin Biotechnol* 2016;38:159–166.

15. Zhang G, Cao T, Ying J, Yang Y, Ma L. Diversity and novelty of actinobacteria in Arctic marine sediments. *Antonie van Leeuwenhoek* 2014;105:743–754.

16. Howe JA, Shimmield TM, Diaz R. Deep-Water sedimentary environments of the northwestern weddel sea and south sandwich islands, antarctica. *Deep Sea Research Part II* 2004;51:1489–1514.

17. Howe JA, Shimmield TM, Harland R. Late quaternary contourites and glaciomarine sedimentation in the Fram Strait. *Sedimentology* 2008;55:179–200.

18. Jorgensen SL, Hannisdal B. Correlating microbial community profiles with geochemical data in highly stratified sediments from the arctic mid-ocean ridge. *Proc Natl Acad Sci USA* 2012.

19. González-Rocha G, Muñoz-Cartes G, Canales-Aguirre CB, Lima CA, Domínguez-Yévenes M *et al*. Diversity structure of culturable bacteria isolated from the fildes peninsula (King George Island, Antarctica): a phylogenetic analysis perspective. *PLoS One* 2017;12:e0179390–18.

20. Bienhold C, Boetius A, Ramette A. The energy-diversity relationship of complex bacterial communities in Arctic deep-sea sediments. *ISME J* 2012;6:724–732.

21. Carr SA, Orcutt BN, Mandernack KW, Spear JR. Abundant atribacteria in deep marine sediment from the adélie basin, antarctica. *Front Microbiol* 2015;6:403–412.

22. Yuan M, Yu Y, Li H-R, Dong N, Zhang X-H. Phylogenetic diversity and biological activity of actinobacteria isolated from the chukchi shelf marine sediments in the arctic ocean. *Mar Drugs* 2014;12:1281–1297.

23. Purves K, Macintyre L, Brennan D, Hreggviðsson G, Kuttner E *et al*. Using molecular networking for microbial secondary metabolite Bioprospecting. *Metabolites* 2016;6:2–18.

24. Mincer TJ, Jensen PR, Kauffman CA, Fenical W. Widespread and persistent populations of a major new marine actinomycete taxon in ocean sediments. *Appl Environ Microbiol* 2002;68:5005–5011.

25. Mohseni M, Norouzi H, Hamedi J, Roohi A. Screening of antibacterial producing actinomycetes from sediments of the Caspian sea. *Int J Mol Cell Med* 2013;2:64–71.

26. Eden PA, Schmidt TM, Blakemore RP, Pace NR. Phylogenetic analysis of Aquaspirillum magnetotacticum using polymerase chain reaction-amplified 16S rRNA-specific DNA. *Int J Syst Bacteriol* 1991;41:324–325.

27. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.

28. Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y *et al*. Introducing EzBioCloud: a taxonomically United database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* 2017;67:1613–1617.

29. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M *et al*. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012;28:1647–1649.

30. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 2006;22:2688–2690.

31. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES science gateway for inference of large phylogenetic trees. gateway computing environments workshop (GCE). *IEEE* 2010:1–8.

32. Shirling EB, Gottlieb D. Methods for characterization of Streptomyces species. *Int J Syst Bacteriol* 1966;16:313–340.

33. Macintyre L, Zhang T, Viegelmann C, Martinez IJ, Cheng C *et al*. Metabolomic tools for secondary metabolite discovery from marine microbial symbionts. *Mar Drugs* 2014;12:3416–3448.

34. Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H. A guided tour of the Trans-Proteomic Pipeline. In: Martens L, Hermjakob H (editors). *Proteomics*, 10; 2010. pp. 1150–1159.

35. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N *et al*. Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nature* 2016;34:828–837.

36. Smoot M, Ono K, Ideker T, Maere S. PiNGO: a Cytoscape plugin to find candidate genes in biological networks. *Bioinformatics* 2011;27:1030–1031.

37. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N *et al*. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2007;2:2366–2382.

38. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM *et al*. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 2008;9:386–388.

39. Hanson NW, Konwar KM, Hawley AK, Altman T, Karp PD *et al*. Metabolic pathways for the whole community. *BMC Genomics* 2014;15:619–14.

40. Stackebrandt E. Taxonomic parameters revisited: tarnished gold standards. *Microbiology. Today* 2006;33:152–155.

41. Wicke C, Hüners M, Wray V, Nimtz M, Bilitewski U *et al.* Production and structure elucidation of glycoglycerolipids from a marine sponge-associated *Microbacterium* species. *J Nat Prod* 2000;63:621–626.

42. Fusetani N, Hashimoto Y. Structures of two water soluble hemolysins isolated from the green alga Ulva pertusa. *Agricultural and Biological Chemistry* 1975;39:2021–2025.

43. Fenical W, Jensen PR. Developing a new resource for drug discovery: marine actinomycete bacteria. *Nat Chem Biol* 2006;2:666–673.

44. Jensen PR, Moore BS, Fenical W. The marine actinomycete genus Salinispora: a model organism for secondary metabolite discovery. *Nat Prod Rep* 2015;32:738–751.

45. Holmes NA, Innocent TM, Heine D, Bassam MA, Worsley SF *et al.* Genome analysis of two Pseudonocardia phylotypes associated with Acromyrmex leafcutter ants reveals their biosynthetic potential. *Front Microbiol* 2016;7:4307–4316.

46. Trujillo ME, Idris H, Riesco R, Nouioui I, Igual JM *et al.* Pseudonocardia nigra sp. nov., isolated from Atacama desert rock. *Int J Syst Evol Microbiol* 2017;67:2980–2985.

47. Chen HH, Qin S, Li J, Zhang YQ, Xu LH *et al.* Pseudonocardia endophytica sp. nov., isolated from the pharmaceutical plant Lobelia clavata. *Int J Syst Evol Microbiol* 2009;59:559–563.

48. Prabahar V, Dube S, Reddy GSN, Shivaji S. Pseudonocardia antarctica sp. nov. an actinomycetes from McMurdo dry Valleys, Antarctica. *Syst Appl Microbiol* 2004;27:66–71.

49. Zhang DF, Jiang Z, Li L, Liu BB, Zhang XM *et al.* Pseudonocardia sediminis sp. nov., isolated from marine sediment. *Int J Syst Evol Microbiol* 2014;64:745–750.

50. Zhao GZ, Zhu WY, Li J, Xie Q, Xu LH *et al.* Pseudonocardia serianimatus sp. nov., a novel actinomycete isolated from the surface-sterilized leaves of Artemisia annua L. *Antonie van Leeuwenhoek* 2011;100:521–528.

51. Stackebrandt E, Goebel BM. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Evol Microbiol* 1994;44:846–849.

52. Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* 2014;64:346–351.

53. Rossi-Tamisier M, Benamar S, Raoult D, Fournier PE. Cautionary tale of using 16S rRNA gene sequence similarity values in identification of human-associated bacterial species. *Int J Syst Evol Microbiol* 2015;65:1929–1934.

54. Millán-Aguiñaga N, Chavarria KL, Ugalde JA, Letzel A-C, Rouse GW *et al.* Phylogenomic insight into Salinispora (bacteria, actinobacteria) species designations. *Sci Rep* 2017;7:3564.

55. Beye M, Fahsi N, Raoult D, Fournier PE. Careful use of 16S rRNA gene sequence similarity values for the identification of *Mycobacterium* species. *New Microbes New Infect* 2018;22:24–29.

56. Howe JA, Wilson CR, Shimmield TM, Diaz RJ, Carpenter LW. Recent deep-water sedimentation, trace metal and radioisotope geochemistry across the southern ocean and Northern Weddell Sea, Antarctica. *Deep Sea Research Part II* 2007;54:1652–1681.

57. Hou Y, Lin S. Distinct Gene Number-Genome Size Relationships for Eukaryotes and Non-Eukaryotes: Gene Content Estimation for Dinoflagellate Genomes. In: Redfield RJ (editor). *PLoS ONE*, 4; 2009. pp. e6978–8.

58. Koester JA, Swalwell JE, von Dassow P, Armbrust EV. Genome size differentiates co-occurring populations of the planktonic diatom Ditylum brightwellii (Bacillariophyta). *BMC Evol Biol* 2010;10:1.

59. Yang JY, Sanchez LM, Rath CM, Liu X, Boudreau PD *et al.* Molecular networking as a dereplication strategy. *J Nat Prod* 2013;76:1686–1699.

60. Quinn RA, Nothias L-F, Vining O, Meehan M, Esquenazi E *et al.* Molecular networking as a drug discovery, drug metabolism, and precision medicine strategy. *Trends Pharmacol Sci* 2017;38:143–154.

61. Tian Y, Li YL, Zhao FC. Secondary metabolites from polar organisms. *Mar Drugs* 2017;15:28–30.