



Microbial community analysis with a specific statistical approach after a record breaking snowfall in Southern Italy

Pamela Monaco, Fabio Divino, Gino Naclerio and Antonio Bucci* 

Abstract

Purpose: Snow and ice ecosystems present unexpectedly high microbial abundance and diversity. Although arctic and alpine snow environments have been intensively investigated from a microbiological point of view, few studies have been conducted in the Apennines. Accordingly, the main purpose of this research was to analyze the microbial communities of the snow collected in two different locations of Capracotta municipality (Southern Italy) after a snowfall record occurred on March 2015 (256 cm of snow in less than 24 h).

Methods: Bacterial communities were analyzed by the Next-Generation Sequencing techniques. Furthermore, a specific statistical approach for taxonomic hierarchy data was introduced, both for the assessment of diversity within microbial communities and the comparison between different microbiotas. In general, diversity and similarity indices are more informative when computed at the lowest level of the taxonomic hierarchy, the species level. This is not the case with microbial data, for which the species level is not necessarily the most informative. Indeed, the possibility to detect a large number of unclassified records at every level of the hierarchy (even at the top) is very realistic due to both the partial knowledge about the cultivable fraction of microbial communities and limitations to taxonomic assignment connected to the quality and completeness of the 16S rRNA gene reference databases. Thus, a global approach considering information from the whole taxonomic hierarchy was adopted in order to obtain a more consistent assessment of the biodiversity.

Result: The main phyla retrieved in the investigated snow samples were *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Firmicutes*. Interestingly, DNA from bacteria adapted to thrive at low temperatures, but also from microorganisms normally associated with other habitats, whose presence in the snow could be justified by wind-transport, was found. Biomolecular investigations and statistical data analysis showed relevant differences in terms of biodiversity, composition, and distribution of bacterial species between the studied snow samples.

Conclusion: The relevance of this research lies in the expansion of knowledge about microorganisms associated with cold environments in contexts poorly investigated such as the Italian Apennines, and in the development of a global statistical approach for the assessment of biological diversity and similarity of microbial communities as an additional tool to be usefully combined with the barcoding methods.

Keywords: Snow, Microbial communities, Next-Generation Sequencing, Record snowfall, Diversity indices, Similarity indices

* Correspondence: antonio.bucci@unimol.it

Department of Biosciences and Territory, University of Molise, Contrada Fonte Lappone, 86090 Pesche, IS, Italy



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Background

An understanding of the temporal and spatial structures, functions, interactions, and population dynamics of microbial communities is critical for many aspects of life, including scientific discovery, biotechnological development, sustainable agriculture, energy security, environmental protection, and human health (Bucci et al. 2017). Accordingly, several methods (cultivation-dependent and molecular approaches) have been employed to reveal microbial community composition and responses to environmental changes in various environments and in different contexts (Malik et al. 2008; Ligi et al. 2014; Vanwonterghem et al. 2014; Bucci et al. 2014, 2015; Warden et al. 2016; Crescenzo et al. 2017; Monaco et al. 2020).

The cryosphere refers to the portion of the Earth where the water is in solid form as snow or ice, including mountain glaciers, ice sheets, ice shelves, sea, lake or river ice, snow cover, permafrost, and seasonal frozen ground (Xiao et al. 2015). Snow and ice environments cover up to 21% of the Earth's surface (Maccario et al. 2015); in winter, up to 12% of the planet's land can be covered by snow (Marshall 2011) while glacial ice extends over approximately 10% of surface, storing 75% of the world's fresh water (Maccario et al. 2015).

For a long period, frozen environments have been considered to be limiting for the development of life due to their extremely harsh climatic conditions such as low temperatures, low atmospheric humidity, low liquid water availability, and high levels of radiation (Cowan and Tow 2004; Lopatina et al. 2016), and they have received much less attention compared to hot habitats. Nevertheless, over the past 20 years, microorganisms inhabiting the cryosphere have been increasingly studied especially for the potential discovery of enzymes with biotechnological interest and to expand knowledge on the ecology of "extreme" environments (Margesin and Miteva 2011; Arrigo 2014). Snow and ice have unexpectedly high microbial abundance and diversity. Arctic and alpine snow have been intensively investigated (Bachy et al. 2011; Varin et al. 2012; Hell et al. 2013; Maccario et al. 2014; Lopatina et al. 2016) and bacteria belonging to several phylogenetic groups such as *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, *Firmicutes*, *Bacteroidetes*, and *Actinobacteria* have been detected (Segawa et al. 2005; Amato et al. 2007; Møller et al. 2013; Maccario et al. 2014; Cameron et al. 2015). On the other hand, studies were also carried out on the surface snow in Antarctica (Carpenter et al. 2000; Brinkmeyer et al. 2003; Christner et al. 2003; Fujii et al. 2010; Lopatina et al. 2013) and revealed the presence of representatives of *Proteobacteria*, *Bacteroidetes*, *Cyanobacteria*, and *Verrucomicrobia* (Brinkmeyer et al. 2003; Lopatina et al. 2013, 2016).

Unlike *Bacteria*, *Archaea* were only rarely observed (Maccario et al. 2015). In Arctic spring snow samples,

sequences associated with *Thaumarchaeota* and *Euryarchaeota* were detected with a relative abundance below 1% over soil, sea ice, and ice sheets (Møller et al. 2013; Maccario et al. 2014; Cameron et al. 2015). In Antarctic sea ice, *Archaea* were estimated at up to 6.6% of the microbial community (Cowie et al. 2011).

The main purpose of this research was to analyze and compare the microbial communities of the snow collected in two different locations of Capracotta, a municipality in the Molise region (Southern Italy; Additional file 1), after a snowfall record that occurred on March 5–6, 2015 (256 cm of snow during an 18-hour period) (Teague and Gallicchio 2017). This village, approximately 220 km east of Rome, is known for its dramatic weather, but this event has been debatably record-breaking (Teague and Gallicchio 2017). Bacterial communities were analyzed by the Next-Generation Sequencing techniques (NGS), a molecular tool widely used to investigate many microbial ecosystems (Yang et al. 2016; Bucci et al. 2017), which returns a huge amount of data to be further processed and analyzed. In addition, the effectiveness of a global approach to assess diversities and similarities was evaluated in order to summarize the innumerable information derived from NGS, both for evaluation of the biodiversity within microbial communities and for comparison between different microbiotas.

Results

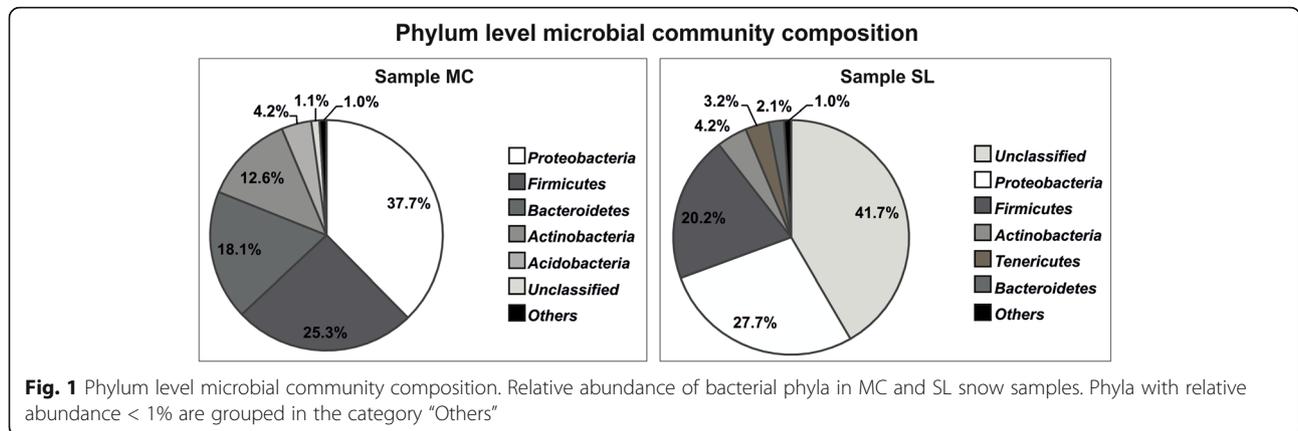
Biomolecular investigations

MiSeq runs produced a total of 297,564 raw reads for Monte Civetta (MC) sample and a total of 81,403 raw reads for Santa Lucia (SL) sample, including V3 and V4 regions of the 16S rRNA gene. The total number of reads that passed quality filtering was 284,844 for MC sample and 75,980 for SL sample. Analyses focused on the *Bacteria* domain.

A different percentage of unclassified microorganisms for each taxonomic level was retrieved in the two investigated snow microbial communities. Unclassified sequences varied between 1% and 42% in MC and SL samples, respectively, already at phylum level, reaching values of ca. 45% (MC) and 73% (SL) at the taxonomic level of species.

Classified OTUs belonged to 23 phyla, 51 classes, 103 orders, 238 families, 656 genera, and 1725 species in MC sample, while in SL sample, there were 21 phyla, 41 classes, 83 orders, 174 families, 418 genera, and 744 species.

As shown in Fig. 1, the five top dominant phyla in MC sample were *Proteobacteria* (37.7%), *Firmicutes* (25.3%), *Bacteroidetes* (18.1%), *Actinobacteria* (12.6%), and *Acidobacteria* (4.2%) whereas SL sample was mainly composed by *Proteobacteria* (27.7%), *Firmicutes* (20.2%), *Actinobacteria* (4.2%), *Tenericutes* (3.2%), and *Bacteroidetes* (2.1%).



The relative abundance of *Acidobacteria* in SL sample was 0.2%. The phylum *Tenericutes* had a much higher relative abundance in SL sample than in MC (0.1%).

At class taxonomic level (Fig. 2a), most of classified reads in MC sample belonged to *Alphaproteobacteria* (26.0%), followed by *Sphingobacteriia* (17.2%), *Bacilli* (16.7%), *Actinobacteria* (12.1%), and *Clostridia* (8.4%). On the other hand, SL sample was characterized by *Betaproteobacteria* (19.5%) and *Bacilli* (19.1%), which together comprised about 39% of total reads, and *Actinobacteria*, *Gammaproteobacteria*, and *Alphaproteobacteria* with a relative abundance of ca. 4% each. *Mollicutes* ranked sixth among classified classes in SL sample with a percentage of 3.2%, while they were poorly represented in MC sample (0.1%).

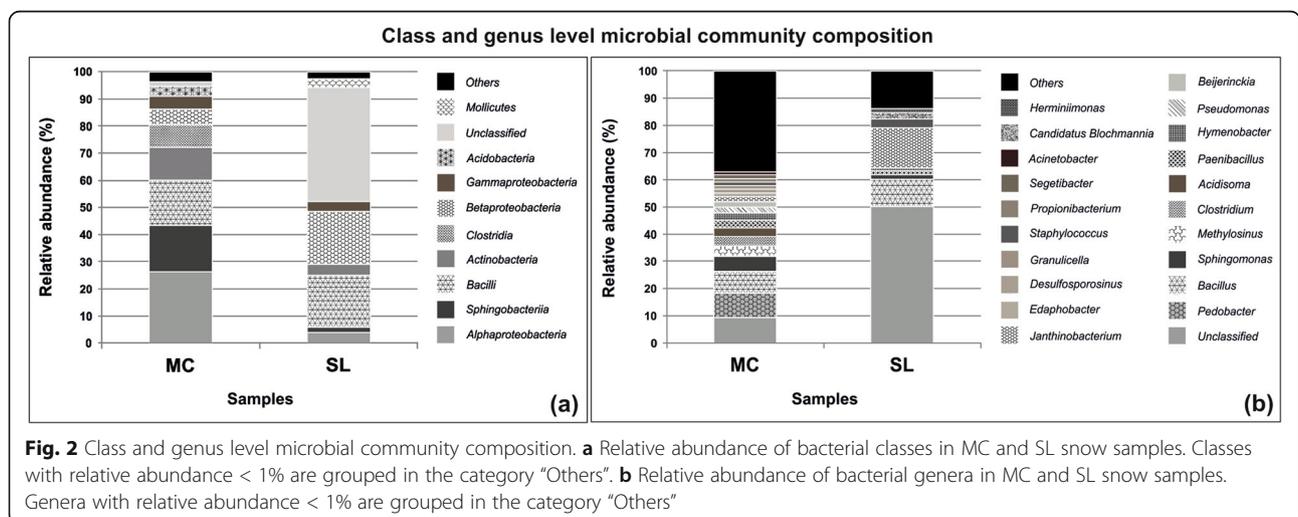
At the taxonomic level of genus (Fig. 2b), *Pedobacter*, *Bacillus*, *Sphingomonas*, and *Methylosinus* represented the four dominant taxa in MC sample, with a relative abundance of 9.0%, 7.9%, 5.5%, and 3.9%, respectively. *Clostridium*, *Acidisoma*, *Paenibacillus*, and *Hymenobacter* were fairly represented.

In SL sample, *Janthinobacterium* and *Bacillus* predominated with percentages of 14.8% and 10.2%, respectively. Also *Staphylococcus*, *Candidatus Blochmannia*, *Paenibacillus*, *Herminiimonas*, *Sphingomonas*, and *Hymenobacter* were found, but with lower relative abundance values (ranging from ca. 1% to ca. 3%).

Furthermore, with a relative abundance of about 3%, *Acidisoma tundrae* and *Pedobacter kwangyangensis* were the main known species found in MC sample, followed by *Beijerinckia mobilis* (1.5%), *Sphingomonas oligophenolica* (1.2%), *Pedobacter cryoconitis* (1.2%), *Sphingomonas wittichii* (1.2%), and *Edaphobacter modestus* (1.1%).

In SL sample, the most representative species was *Janthinobacterium agaricidamnosum* with a percentage of 6.0%, followed by *Bacillus badius*, *Candidatus Blochmannia rufipes*, and *Bacillus smithii* (relative abundance > 1%).

However, a high number of classified OTUs (ca. 43% of total reads in MC sample and ca. 16% in SL sample) identified bacterial species that, taken individually, had



a relative abundance less than 1%, suggesting an extreme fragmentation of microbial communities, probably composed by numerous distinct bacterial populations.

Microbial biodiversity assessment

In order to have a coherent assessment of the microbial diversity and similarity, and overcome the potential non-consistency when indices are referred to different taxonomic levels, it was considered the global approach presented in the “A global approach to the analysis of taxonomic data” section. The assessment was obtained following a compositional data approach (Aitchison 1986) with the counts transformed in proportions: for each taxonomic level j , each category count was divided by the total number of classified individuals.

Two selected indices were considered at each taxonomic level j . In particular, to evaluate the diversity within each sample, it was computed the Shannon entropy (Shannon and Weaver 1963) in the relative form (h) introduced by Pielou (Pielou 1969):

$$h_j = \frac{-1}{\log(k_j)} \sum_{i=1}^{k_j} f_{ij} \log(f_{ij}),$$

where k_j denotes the number of recognized taxa and f_{ij} the sampling proportion of microbial individuals classified in the category i of the taxonomic level j .

In addition, to assess the similarity between the two samples, it was computed the percentage model affinity (pma) index as introduced by Novak and Bode (Novak and Bode 1992) and investigated in terms of statistical properties by Ärje et al. (Ärje et al. 2016):

$$pma_j = 1 - \frac{1}{2} \sum_{i=1}^{k_j} |p_{ij} - q_{ij}|,$$

where p_{ij} and q_{ij} denote the sampling proportions of microbial individuals classified in the category i of the taxonomic level j , referred to the two samples, respectively.

Diversity

In Additional file 2, the weights w_j , normalized to unity (first and third columns), and the values of the relative entropy h_j (second and fourth columns) are reported for each sample, while in Fig. 3, the respective plots of the relative Shannon entropy are shown.

The global entropy (h_g) was computed from the data in Additional file 2 for both the samples:

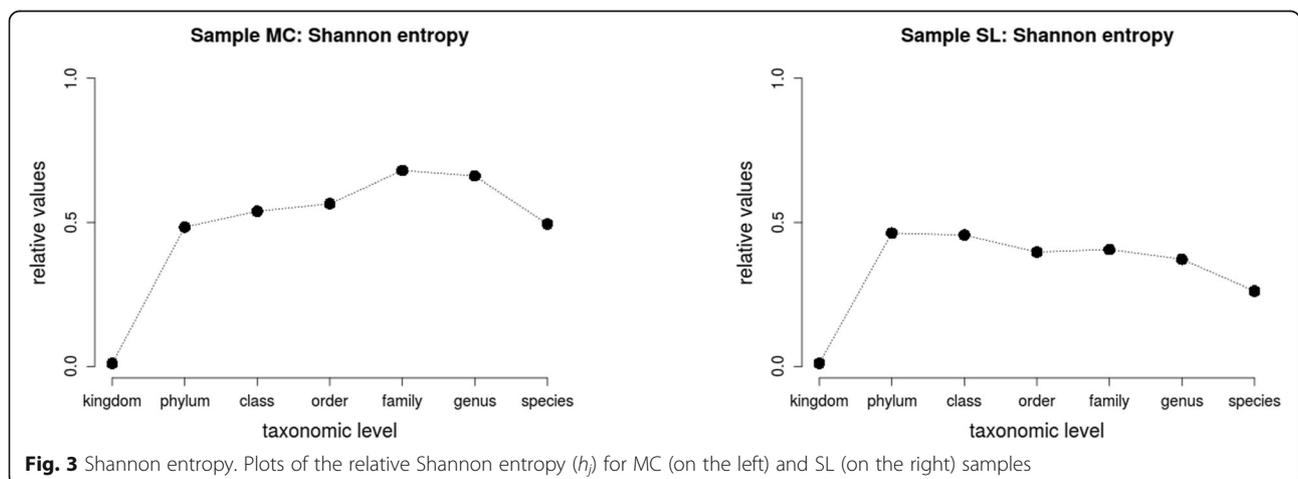
$$h_g = \sum_{j=1}^J w_j h_j.$$

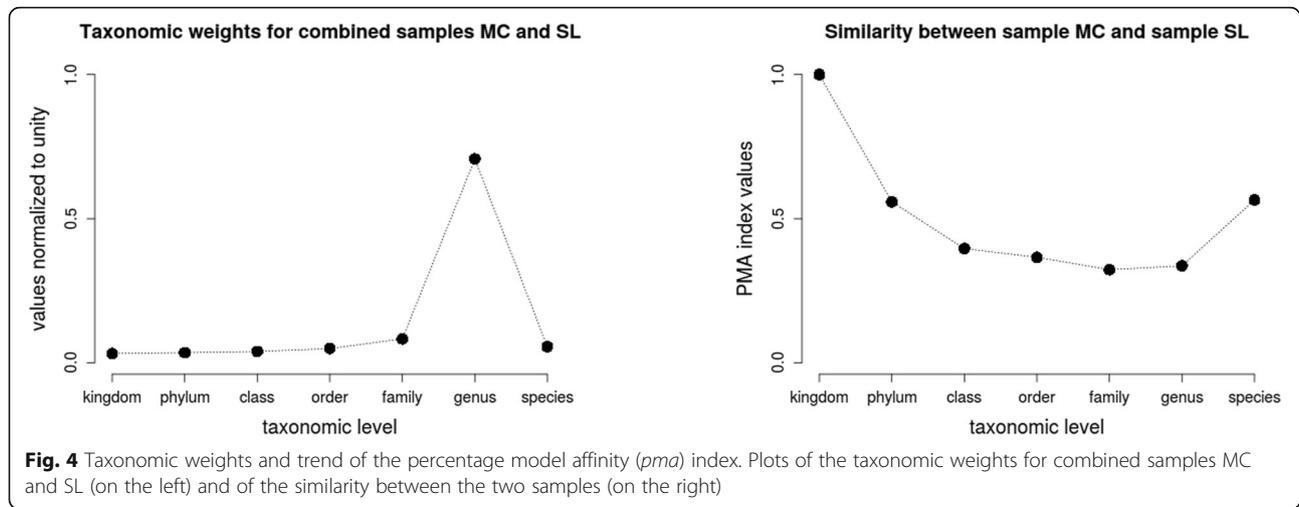
The obtained values, $h_g = 0.520$ and $h_g = 0.365$, showed that MC sample presented a higher level of biodiversity than SL sample. Furthermore, from data in Additional file 2, it is interesting to notice that in the sample MC, the species level presented the highest weight, with value equal to 0.301, while in the sample SL, the most informative level was the genus level with the respective weight equal to 0.707.

Similarity

With respect to the assessment of the microbial similarity between the two samples, the construction of the respective weights is different because it is necessary to combine the scaled abundance and the scaled richness of both the samples. Therefore, in this case, the mean and standard deviation were computed with respect to the values v_j and κ_j . In Additional files 3 and 4, all the quantities necessary to the construction of the weights were reported: classified records n_j , unclassified records u_j , recognized taxa k_j , scaled abundance v_j , scaled richness κ_j , mean μ_j , standard deviation σ_j , and weight w_j .

In Additional file 5, the weights w_j and the values of the percentage model affinity (pma_j) index were reported,





with the respective plots shown in Fig. 4. Then, the global similarity index (pma_g) is given as follows:

$$pma_g = \sum_{j=1}^J w_j pma_j,$$

with value $pma_g = 0.472$ that denotes a relevant degree of microbial dissimilarity between the two samples of data.

Discussion

This study represents one of the few works concerning the analysis of the snow microbiota in the Apennines, and its relevance is also related to the extraordinary snowfall event that took place in the municipality of Capracotta in 2015. The exceptionally adverse weather conditions caused serious problems with movements of people within the study area and did not allow to develop the experimental plan in an optimally manner, mainly with reference to the small sampling size. Nonetheless, the research constitutes a preliminary step for further and more accurate investigations of an environmental context poorly characterized till now.

Snow microbial communities were analyzed and compared by using NGS and a specific statistical approach for data interpretation, here described for the very first time. The assessment of biological diversity and similarity is relevant in many environmental studies. Indeed, biodiversity trends in time and space may often suggest environmental dynamics or changes in the status of ecosystems (Spellerberg 2005). In particular, in microbial studies, biodiversity assessment is important to describe communities and formulate hypotheses about the potential relations with the environments where the communities are observed. When information is obtained from genetic sequencing technologies, microbial data usually consist of taxonomic counts observed from one or more

samples of the community of interest, and referred to every level of the taxonomic hierarchy, from kingdom to species. In general, when shifting from kingdom to species along the taxonomic hierarchy, the data are typically characterized by slightly decreasing levels of abundance (n_j) and significantly increasing levels of richness (k_j). Therefore, diversity and similarity indices are generally more informative when computed at the lowest level of the hierarchy, the species level for instance.

Unfortunately, this is not the case with microbial data, for which the possibility to detect a large number of unclassified records (u_j) at every level of the hierarchy (even at the top) is realistic, due to the partial knowledge about the cultivable fraction of microbial communities and to the limitations of NGS technologies, caused chiefly by shorter read length and impacting on precision of species identification (Bukin et al. 2019). Therefore, in microbial studies, the species level is not necessarily the most informative. In these situations, to obtain a more consistent assessment of the diversity and similarity coherently with the taxonomy, a global approach is required: it is necessary to consider information from the whole taxonomic hierarchy and not from only one specific level (the species level for instance).

NGS results were in agreement with data available in scientific literature concerning snow microbial communities composition at phylum level, with *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Firmicutes* representing the main groups found in these ecosystems (Liu et al. 2006; Zhang et al. 2010; Michaud et al. 2014; Mortazavi et al. 2015). These phyla include different bacterial genera and species with an extreme metabolic diversity and the ability to adapt to snow and ice environments, such as *Polaribacter*, *Psychroflexus*, and *Pedobacter*. Some have the ability to form endospores, which confers an important selective advantage in terms of adaptation to harsh environmental conditions.

Nevertheless, biomolecular investigations and subsequent statistical data analysis showed relevant differences in terms of biodiversity, composition, and distribution of bacterial species between the studied snow samples. Indeed, the value of the global similarity index (pma_g) between MC and SL samples was 0.472. The standard threshold 0.7 discussed in Ärje et al. (Ärje et al. 2016) is reputed as the critical level under which two communities are not considered similar in composition. Therefore, the observed value $pma_g = 0.472$ denotes a clear level of dissimilarity.

After family ($pma = 0.323$), genus was the taxonomic level in which the main differences between MC and SL snow samples were concentrated ($pma = 0.336$). These differences must be sought in the percentage at which the different genera occur.

Furthermore, in accord with the global approach based on the weighted averaging, the data analysis showed that species and genus were the most informative taxonomic levels (in MC and SL samples, respectively).

In general, NGS analysis showed the presence of genera comprising bacterial species adapted to thrive at low temperatures and typical of snow/ice ecosystems, thus, probably constituting a resident microbiota (*Sphingomonas*, *Methylobacterium*, *Acidisoma*, *Janthinobacterium*, *Paenibacillus*, *Hymenobacter*), and genera including microorganisms whose presence could be justified by wind-transport (transient species) such as *Beijerinckia*, *Deinococcus*, *Geodermatophilus*, *Maricaulis*, *Marinibacillus*, *Marinitoga*, *Marinobacter*, *Marinobacterium*, *Marinococcus*, and *Marinomonas*.

Sphingomonas species can tolerate intense radiation, drying, and low concentrations of nutrients (Liu et al. 2006). Several studies highlighted the presence of bacteria belonging to this genus in different cold environments such as the Arctic snow, the Tibetan snow and glaciers, and the Antarctic snow and soils (Christner et al. 2002; Segawa et al. 2005; Liu et al. 2006; Miteva 2008; An et al. 2010; Xiang et al. 2010; Lopatina et al. 2013; Michaud et al. 2014; Mortazavi et al. 2015). Metatrophic bacteria belonging to *Methylobacterium* genus are often found in cold environments in a metabolically active form (Lopatina et al. 2013) as well as species belonging to the genus *Acidisoma* (*Acidisoma tundrae* and *Acidisoma sibiricum*) which are psychrotolerant and moderately acidophilic bacteria capable of growth at 2–30 °C and pH 3.0–7.6 (Belova et al. 2009).

With reference to other genera commonly found in cold environments and in the snow (Rainey et al. 2005; Zhang et al. 2010; Chuvochina et al. 2011; Ivy et al. 2012; Lee et al. 2014; Mortazavi et al. 2015) and retrieved in samples collected in Capracotta, *Paenibacillus* includes widely distributed psychrotolerant spore-forming bacteria, previously isolated from Arctic snow

samples (Ivy et al. 2012; Mortazavi et al. 2015) whereas *Hymenobacter's* ability to withstand high doses of ionizing radiation represents a selective advantage increasing cell survival chances, not only during wind-mediated transport but also during snow exposition to intense UV radiation (Rainey et al. 2005; Zhang et al. 2010; Chuvochina et al. 2011; Lee et al. 2014).

Concerning transient microorganisms present in snow environments and probably transported by winds, one of the main genera retrieved in MC sample was *Beijerinckia* (with *Beijerinckia mobilis* in third place among classified species). It includes mainly soil bacteria widely distributed in the acid tropical soils of equatorial Africa, Southeast Asia, and South America, but it has been found only sporadically in temperate and subtropical areas. *Beijerinckia* spp. are able to grow in a temperature range between 10 and 35 °C, with optimal values between 20 and 30 °C. However, *Beijerinckia* cells resist freezing (Becking 1959, 1961, 1981). Therefore, considering the geographical distribution of this genus, its presence in MC and SL samples could find an explanation in a long distance transport and deposition during snowfalls.

In addition, also the genera *Deinococcus* and *Geodermatophilus* have been found, consistently with results reported in several scientific works (Carpenter et al. 2000; Chuvochina et al. 2011; Michaud et al. 2014; Mortazavi et al. 2015). *Geodermatophilus obscurus*, the type species of the genus *Geodermatophilus*, grows in the temperature range 18–37 °C, with an optimum of 24–28 °C (Ivanova et al. 2010). These temperatures were much higher than the average temperatures recorded in Capracotta especially in wintertime. Therefore, here too, it is possible to assume that these microorganisms, which do not represent cold adapted species, could be part of a transient microbiota in the snow.

It is generally recognized that microorganisms normally associated with other habitats (e.g., marine species, mesophiles, and even thermophiles) have often been found in snow and ice ecosystems (Liu et al. 2006; Lopatina et al. 2013). This statement is further supported by the presence of bacteria typically associated with marine habitats such as *Maricaulis*, *Marinibacillus*, *Marinitoga*, *Marinobacter*, *Marinobacterium*, *Marinococcus*, and *Marinomonas*. Their recovery could be related to the well-known orographic lift: the cold air coming from North-East, while crossing the Adriatic Sea, charges with humidity and microorganisms before reaching the internal mountainous areas of Molise Apennines giving rise to heavy snowfall (Stocchi and Davolio 2017), and influences the microbiota composition. These findings are in line with the results of Harding et al. (2011) which applied molecular, microscopic, and culture techniques to characterize the microbial communities in snow and air at remote sites in the Canadian High Arctic.

Microorganisms retrieved in cold environments such as Antarctica, the Tibetan Plateau, and alpine regions of Japan, Europe, and North America were found together with microbes coming from diverse biomes in the coastal Arctic contributing to local inocula in the snow. These revealed the importance of aerial transport as a major transport route enabling microbes to colonize the different habitats.

Conclusions

In conclusion, the results obtained have shown that snow microbial communities retrieved in SL and MC samples relevantly differ from each other despite they are represented by bacterial phyla normally associated to cold environments. The reasons could be found in the different location, altitude, and also touristic usability of the sampling sites. In fact, unlike Santa Lucia sampling site, an unfrequented area, Monte Civetta was a place where thousands of people streamed in winter season to reach the alpine ski facilities of Monte Capraro, hosting some refreshment points, and it is known that microbial communities are generally the first responders to environmental chemical parameters/environmental perturbation (Bergk Pinto et al. 2019). An extreme fragmentation was suggested by the presence of several bacterial populations, each representing a small fraction of the whole microbial community. In addition, the presence of DNA from various microorganisms typically associated to other habitats demonstrates the strong potential of the wind-mediated transport in shaping the microbiota composition, enriching the normal resident bacterial communities with transient microorganisms. Although much of the biology and ecology of snow is still unknown, it is common knowledge that seasonal snowpack can influence the local climate, underlying soil, and adjacent ecosystems. For example, by regulating freeze-thaw events, the extent and duration of snow cover can affect soil microbial community composition, microbial-mediated soil nitrogen (N) cycling, and greenhouse gas exchange with the atmosphere. Accordingly, our exploratory approach and results can be used as a starting point to develop further investigations on the Apennines useful to address several research questions on the microbial ecology of these peculiar environments.

Methods

Study area

Capracotta is a small mountain village of about 900 inhabitants in the Molise region (Southern Italy; Additional file 1). It is the second highest municipality in the Apennines at 1421 m above sea level (a.s.l.).

In winter, snowfalls are frequent and abundant with temperatures that can drop down to several degrees below 0 °C whereas summers are mild.

Snow samples for molecular analyses were collected at Monte Civetta (MC), an area located at 1650 m

a.s.l. close to the alpine ski facilities of Monte Capraro, and at Santa Lucia (SL; 1550 m a.s.l.), located at the base of Monte Campo, the highest mountain in the territory of Capracotta.

Sampling

Samples from each study site were obtained after the snowfall record event (March 7, 2015) by mixing snow collected in 6 points at a distance of about 1 m from each other, within an area of 15 m² (5 m × 3 m). Surface snow layer was removed using a sterile spoon to eliminate deposited coarser particles such as dust and plant material (Amato et al. 2007). Collection was performed with a sterile plastic shovel in 6 sterile polyethylene containers (2 l) from each site (Harding et al. 2011). Sterilized suite and gloves were worn during sample collection to minimize contamination (Lopatina et al. 2013). Successively, the 6 subsamples from Monte Civetta and Santa Lucia were transported to the laboratory, the snow was melted over a period of ca. 12 h and occasionally mixed in order to maintain a homogeneous water temperature of ≤ 4 °C (Harding et al. 2011; Lopatina et al. 2013). After melting, 1 l of water from each subsample was measured in a sterile graduated cylinder and added to a single sterile polyethylene container (10 l): a single 6 l water sample was generated from each of the sampling sites (Monte Civetta and Santa Lucia). The two samples were filtered by using a membrane filtration system.

Biomolecular investigations

DNA extraction

For each sample, 6 l of melted snow were filtered through mixed esters of cellulose membrane filters (S-PakTM Membrane Filters, 47-mm diameter, 0.22-µm pore size, Millipore Corporation, Billerica, MA, USA). Filters were stored at - 80 °C until nucleic acid extraction. Total genomic DNA was extracted using the PowerWater[®]DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) and used for the Next-Generation Sequencing analysis.

Dual index 16S rRNA gene amplicon library preparation and bioinformatics analysis

NGS sequencing protocol was performed at BMR Genomics srl (Padova, Italy). For each snow sample, the V3–V4 regions of the 16S rRNA gene were amplified using the primers 331F and 797R (Nadkarni et al. 2002).

The primers were modified with a forward overhang (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGA CAG-[locus-specific sequence]-3') and a reverse overhang (5'-GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAG-[locus-specific sequence]-3'), which were necessary for dual index library preparation.

The library was run on the Illumina MiSeq (San Diego, California, USA) using the 2×300 bp paired-end approach.

The classification step used ClassifyReads, a high-performance implementation of the Ribosomal Database Project (RDP) Classifier described in Wang et al. (Wang et al. 2007). This process involved matching short subsequences of the reads (called words) to a set of 16S rRNA gene reference sequences (the taxonomic database used was an Illumina-curated version of the May 2013 release of the Greengenes Consortium Database). The accumulated word matches for each read were used to assign reads to a particular taxonomic classification. Forward and reverse strands were aligned independently in paired-end runs. Read stitching was not performed but classification required both reads from each cluster to classify to the same taxonomy to not be excluded. Analyses focused on the *Bacteria* domain. Therefore, sequences referred to viruses and microorganisms belonging to other domains, such as *Archaea*, have been excluded from the subsequent investigation. 16S rRNA gene sequences generated in this study are deposited in the NCBI Sequence Read Archive under the accession number PRJNA563617.

A global approach to the analysis of taxonomic data

Introducing some mathematical notation, the taxonomic levels are here denoted by the index j , with $j = 1, \dots, J$, corresponding to the standard hierarchy: kingdom, phylum, class, order, family, genus, and species (hence $J = 7$). Then, for each level j , n_j and u_j indicate the numbers of classified and unclassified records, respectively, while k_j denotes the number of recognized taxa. For data collected from m distinct samples of the community of interest, n_j and u_j are obviously the sum over the m samples of the respective quantities, while k_j is the total number of the recognized taxa in all the m samples, for each taxonomic level $j = 1, \dots, J$. Therefore, from the calculation point of view, distinct samples are considered as pooled.

In this section, a specific procedure based on the averaging method is presented. In particular, considering a generic biodiversity indicator of interest I , with the respective values I_j computed from the data at each level j of the taxonomic hierarchy, the procedure consists in the averaging those values I_j with specific weights w_j determined in terms of amount of information present in the respective taxonomic level j .

Indeed, at each level j of the hierarchy, two relevant quantities are always available: the abundance, in terms of classified records n_j , and the richness, in terms of recognized taxa k_j . These parameters represent the degree of statistical consistence and the degree of microbial variety of the data related to each taxonomic level. Therefore,

it would be natural to define each weight by combining these two types of information, such as through the arithmetic mean of the respective scaled values (denoted here by v_j and κ_j , respectively). The scaling is necessary because abundance and richness have different magnitudes in general.

Furthermore, the more similar the scaled values, the more coherent the respective weight. Therefore, in addition to the average of v_j and κ_j , each weight should also account for their precision, such as the inverse of the standard deviation for instance. Hence, in order to account for both average and precision, for each level j , the respective weight w_j was defined in terms of the ratio between the arithmetic mean and the standard deviation of the scaled values v_j and κ_j .

To show the construction of our proposal, an example with a generic indicator I is presented in the following scheme.

For each taxonomic level j of the hierarchy:

- Compute the scaled values v_j and κ_j as

$$v_j = \frac{n_j}{n_M} \text{ and } \kappa_j = \frac{k_j}{k_M},$$

where $n_M = \max\{n_j; j = 1, \dots, J\}$ and $k_M = \max\{k_j; j = 1, \dots, J\}$ are the respective maxima across the hierarchy;

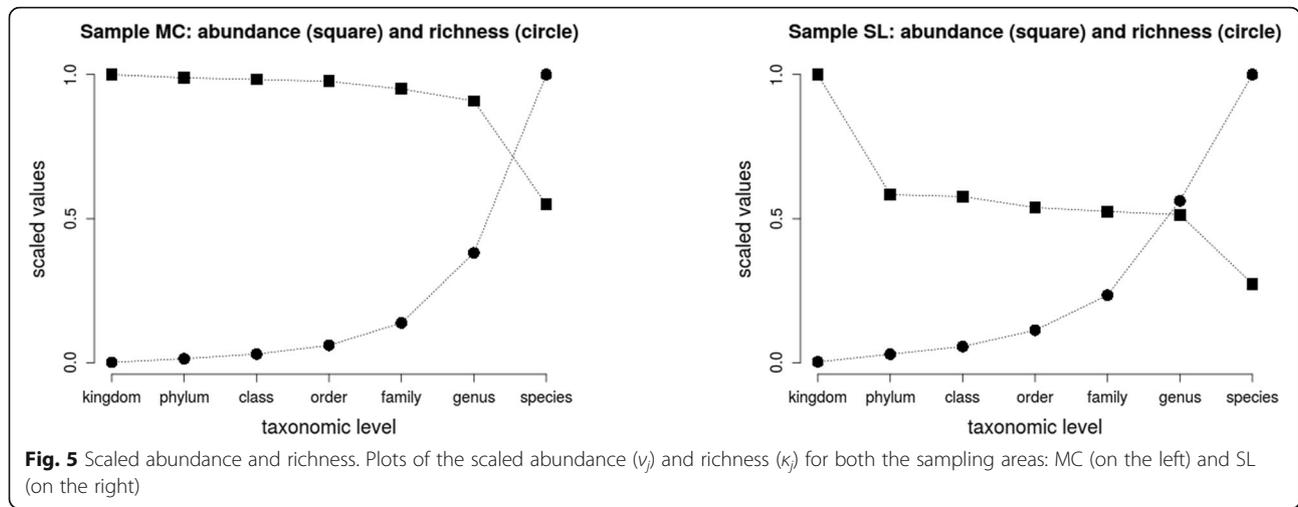
- Compute the mean μ_j and the standard deviation σ_j ;
- Define each weight as the ratio $w_j = \frac{\mu_j}{\sigma_j}$;
- Normalize the weights to unity;
- Compute the global value I_g as the weighted average

$$I_g = \sum_{j=1}^J w_j I_j.$$

The system of weights obtained through this procedure defines each weight as the relative precision (the inverse of the coefficient of variation) of the information included in each taxonomic level j .

This approach can be easily generalized with biodiversity indices, which consider more than one sample of data, such as the similarity indices. Indeed, in this case, at every level j of the hierarchy, it is simply necessary to extend the computation of the mean and standard deviation to the values obtained from all the samples.

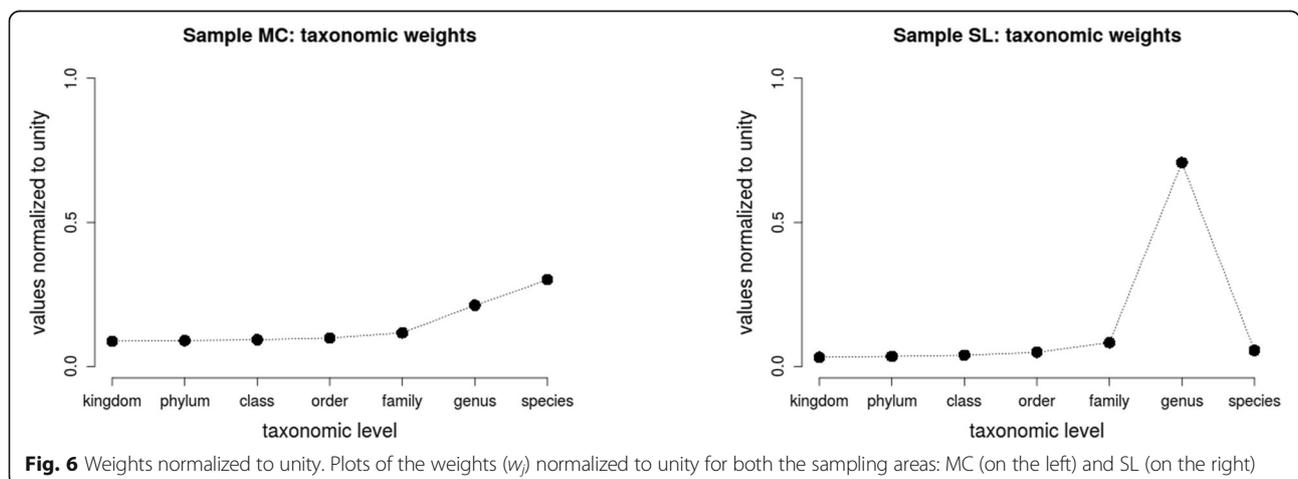
In Fig. 5, the plots of the scaled abundance and the scaled richness are reported for each taxonomic level. These scaled values are necessary to calculate the taxonomic weights as arithmetic means for each taxonomic level and for each sample. In Fig. 6, the system of taxonomic weights with values normalized to unity is shown



for both MC and SL. Each weight represents the importance of the corresponding taxonomic level in terms of relative abundance and richness.

Currently, the assessment of diversity and similarity in biological communities with DNA data is often based on the use of clustering methods which process directly the DNA barcodes (Hebert et al. 2003; Jin et al. 2013). After the clusters are obtained, diversity and similarity indices are computed with respect to such clustering structure of the data. This approach is useful to avoid the problem of unclassified DNA sequences, since the clustering of DNA barcodes is independent from any taxonomic classification. Although the DNA barcoding approach is very useful to identify new species groups and assign unknown individuals to clusters (Savolainen et al. 2005), it strongly depends on the choice of the distance or metrics used in the clustering algorithm, with potentially relevant effects on the numerical values of the diversity

and similarity indices. The approach introduced in this work produces a global assessment of diversity and similarity that considers the information of the taxonomic structure of the data that is how data are distributed along the different levels of the taxonomic hierarchy, combined with the respective relevance of each taxonomic level, in terms of abundance and richness. This aspect is missing in the barcoding approach, as the clustering computation is based only on the DNA sequences, without the taxonomic labels. Therefore, the method proposed in this article allows the calculation of any diversity or similarity indicator preserving the taxonomic structure of the data and represents an additional tool that can be usefully combined with other existing approaches in order to obtain a more complete evaluation of diversity and similarity in biological communities, paving the way to another perspective to look at biodiversity for 16S rDNA NGS data.



Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1186/s13213-020-01604-6>.

Additional file 1: Figure S1. Study area. Localization of Capracotta municipality. (TIF 217 kb)

Additional file 2: Table S1. Taxonomic weights normalized to unity (w_j) and relative Shannon entropy (h_j) for both samples MC and SL, across the taxonomic hierarchy. (DOCX 12.7 kb)

Additional file 3: Table S2. Construction of the taxonomic weights (w_j) for sample MC, across the taxonomic hierarchy. (DOCX 13.0 kb)

Additional file 4: Table S3. Construction of the taxonomic weights (w_j) for sample SL, across the taxonomic hierarchy. (DOCX 12.7 kb)

Additional file 5: Table S4. Taxonomic weights (w_j) for joined samples MC and SL, and values of the percentage model affinity (p_{ma}) index, across the taxonomic hierarchy. (DOCX 12.3 kb)

Acknowledgements

Not applicable.

Authors' contributions

PM, FD, GN, and AB contributed intellectual input to this study. PM collected the samples and performed the experiments and data analysis. FD, GN, and AB provided assistance during the experiments and contributed to the data analysis. PM, FD, and AB wrote the paper. The authors read and approved the final manuscript.

Funding

Fabio Divino is partially supported by the PRIN2015 project "Environmental processes and human activities: capturing their interactions via statistical methods (EPHASTAT)" funded by MIUR (Italian Ministry of Education, University and Scientific Research) [MIUR PRIN 20154X8K23 SH3 - CUP: B86J16002110001].

Availability of data and materials

16S rRNA gene sequences generated during the current study are deposited/available in the NCBI Sequence Read Archive under the accession number PRJNA563617 [<https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA563617>].

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 8 September 2020 Accepted: 9 October 2020

Published online: 09 November 2020

References

- Aitchison J (1986) The statistical analysis of compositional data. Chapman & Hall, London
- Amato P, Hennebelle R, Magand O, Sancelme M, Delort AM, Barbante C et al (2007) Bacterial characterization of the snow cover at Spitzberg, Svalbard. *FEMS Microbiol Ecol* 59:255–264
- An LZ, Chen Y, Xiang SR, Shang TC, Tian LD (2010) Differences in community composition of bacteria in four glaciers in western China. *Biogeosciences* 7: 1937–1952
- Årje J, Choi K, Divino F, Meissner K, Kärkkäinen S (2016) Understanding the statistical properties of the percent model affinity index can improve biomonitoring related decision making. *Stoch Env Res Risk A* 30:1981–2008
- Arrigo KR (2014) Sea ice ecosystems. *Annu Rev Mar Sci* 6:439–467
- Bachy C, López-García P, Vereshchaka A, Moreira D (2011) Diversity and vertical distribution of microbial eukaryotes in the snow, sea ice and seawater near the North Pole at the end of the polar night. *Front Microbiol* 2:106
- Becking JH (1959) Nitrogen-fixing bacteria of the genus *Beijerinckia* in South African soils. *Plant Soil* 11:193–206
- Becking JH (1961) Studies on nitrogen-fixing bacteria of the genus *Beijerinckia*. I. Geographical and ecological distribution in soils. *Plant Soil* 14:49–81
- Becking JH (1981) The family *Azotobacteraceae*. In: Starr MP, Stolp H, Trüper HG, Balows A, Schlegel HG (eds) *The prokaryotes: a handbook on habitats, isolation, and identification of bacteria*. Springer-Verlag, Berlin, pp 795–817
- Belova SE, Pankratov TA, Detkova EN, Kaparullina EN, Dedysh SN (2009) *Acidisoma tundrae* gen. nov., sp. nov. and *Acidisoma sibiricum* sp. nov., two acidophilic, psychrotolerant members of the *Alphaproteobacteria* from acidic northern wetlands. *Int J Syst Evol Micr* 59:2283–2290
- Bergk Pinto B, Maccario L, Dommergue A, Vogel TM, Larose C (2019) Do organic substrates drive microbial community interactions in arctic snow? *Front Microbiol* 10:2492
- Brinkmeyer R, Knittel K, Jürgens J, Weyland H, Amann R, Helmke E (2003) Diversity and structure of bacterial communities in Arctic versus Antarctic pack ice. *Appl Environ Microb* 69:6610–6619
- Bucci A, Allocca V, Naclerio G, Capobianco G, Divino F, Fiorillo F et al (2015) Winter survival of microbial contaminants in soil: an *in situ* verification. *J Environ Sci* 27:131–138
- Bucci A, Petrella E, Celico F, Naclerio G (2017) Use of molecular approaches in hydrogeological studies: the case of carbonate aquifers in Southern Italy. *Hydrogeol J* 25:1017–1031
- Bucci A, Petrella E, Naclerio G, Gambatese S, Celico F (2014) Bacterial migration through low-permeability fault zones in compartmentalised aquifer systems: a case study in Southern Italy. *Int J Speleol* 43:273–281
- Bukin YS, Galachyants YP, Morozov I, Bukin SV, Zakharenko AS, Zemskaia TI (2019) The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Sci Data* 6:190007
- Cameron KA, Hagedorn B, Dierer M, Christner BC, Choquette K, Sletten R et al (2015) Diversity and potential sources of microbiota associated with snow on western portions of the Greenland Ice Sheet. *Environ Microbiol* 17:594–609
- Carpenter EJ, Lin S, Capone DG (2000) Bacterial activity in South Pole snow. *Appl Environ Microb* 66:4514–4517
- Christner BC, Kvitko BH, Reeve JN (2003) Molecular identification of bacteria and eukarya inhabiting an Antarctic cryoconite hole. *Extremophiles* 7:177–183
- Christner BC, Mosley-Thompson E, Thompson LG, Zagorodnov V, Sandman K, Reeve JN (2002) Isolation and identification of bacteria from ancient and modern ice core archive. In: Casassa G, Sepulveda FV, Sinclair R (eds) *Patagonia Ice Fields, a Unique Natural Laboratory for Environmental and Climate Change Studies*. Kluwer, New York, pp 9–16
- Chuvochina MS, Marie D, Chevaillier S, Petit JR, Normand P, Alekhina IA et al (2011) Community variability of bacteria in Alpine snow (Mont Blanc) containing Saharan dust deposition and their snow colonization potential. *Microbes Environ* 26:237–247
- Cowan DA, Tow LA (2004) Endangered Antarctic environments. *Annu Rev Microbiol* 58:649–690
- Cowie ROM, Maas EW, Ryan KG (2011) Archaeal diversity revealed in Antarctic sea ice. *Antarct Sci* 23:531–536
- Crescenzo R, Mazzoli A, Cancelliere R, Bucci A, Naclerio G, Baccigalupi L et al (2017) Beneficial effects of carotenoid-producing cells of *Bacillus indicus* HU16 in a rat model of diet-induced metabolic syndrome. *Benef Microbes* 8:823–831
- Fujii M, Takano Y, Kojima H, Hoshino T, Tanaka R, Fukui M (2010) Microbial community structure, pigment composition, and nitrogen source of red snow in Antarctica. *Microb Ecol* 59:466–475
- Harding T, Jungblut AD, Lovejoy C, Vincent WF (2011) Microbes in High Arctic snow and implications for the cold biosphere. *Appl Environ Microb* 77:3234–3243
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proc R Soc Lond B* 270:313–321
- Hell K, Edwards A, Zarsky J, Podmirseg SM, Girdwood S, Pachebat JA et al (2013) The dynamic bacterial communities of a melting High Arctic glacier snowpack. *ISME J* 7:1814–1826
- Ivanova N, Sikorski J, Jando M, Munk C, Lapidus A, Del Rio TG et al (2010) Complete genome sequence of *Geodermatophilus obscurus* type strain (G-20^T). *Stand Genomic Sci* 2:158–167
- Ivy RA, Raniere ML, Martin NH, Bakker HC, Xavier BM, Wiedmann M et al (2012) Identification and characterization of psychrotolerant sporeformers associated with fluid milk production and processing. *Appl Environ Microb* 78:1853–1864
- Jin Q, Han H, Hu X, Li X, Zhu C, Ho SYW et al (2013) Quantifying species diversity with a DNA barcoding-based method: Tibetan moth species (Noctuidae) on the Qinghai-Tibetan plateau. *Plos One* 8:e64428

- Lee JJ, Srinivasan S, Lim S, Joe M, Lee SH, Kwon SA et al (2014) *Hymenobacter swuensis* sp. nov., a gamma-radiation-resistant bacteria isolated from mountain soil. *Curr Microbiol* 68:305–310
- Ligi T, Oopkaup K, Truu M, Preem JK, Nölvak H, Mitsch WJ et al (2014) Characterization of bacterial communities in soil and sediment of a created riverine wetland complex using high-throughput 16S rRNA amplicon sequencing. *Ecol Eng* 72:56–66
- Liu Y, Yao T, Kang S, Jiao N, Zeng Y, Shi Y et al (2006) Seasonal variation of snow microbial community structure in the East Rongbuk glacier, Mt. Everest. *Chinese Sci Bull* 51:1476–1486
- Lopatina A, Krylenkov V, Severinov K (2013) Activity and bacterial diversity of snow around Russian Antarctic stations. *Res Microbiol* 164:949–958
- Lopatina A, Medvedeva S, Shmakov S, Logacheva MD, Krylenkov V, Severinov K (2016) Metagenomic analysis of bacterial communities of Antarctic surface snow. *Front Microbiol* 7:398
- Maccario L, Sanguino L, Vogel TM, Larose C (2015) Snow and ice ecosystems: not so extreme. *Res Microbiol* 166:782–795
- Maccario L, Vogel TM, Larose C (2014) Potential drivers of microbial community structure and function in Arctic spring snow. *Front Microbiol* 5:413
- Malik S, Beer M, Megharaj M, Naidu R (2008) The use of molecular techniques to characterize the microbial communities in contaminated soil and water. *Environ Int* 34:265–276
- Margesin R, Miteva V (2011) Diversity and ecology of psychrophilic microorganisms. *Res Microbiol* 162:346–361
- Marshall SJ (2011) *The cryosphere*. Princeton University Press, Princeton
- Michaud L, Lo Giudice A, Mysara M, Monsieus P, Raffa C, Leys N et al (2014) Snow surface microbiome on the High Antarctic Plateau (DOME C). *Plos One* 9:e104505
- Miteva V (2008) Bacteria in snow and glacier ice. In: Margesin R, Schinner F, Marx JC, Gerday C (eds) *Psychrophiles: From Biodiversity to Biotechnology*. Springer, Berlin, pp 31–50
- Møller AK, Søborg DA, Al-Soud WA, Sørensen SJ, Kroer N (2013) Bacterial community structure in High-Arctic snow and freshwater as revealed by pyrosequencing of 16S rRNA genes and cultivation. *Polar Res* 32:17390
- Monaco P, Toumi M, Sferra G, Tóth E, Naclerio G, Bucci A (2020) The bacterial communities of *Tuber aestivum*: preliminary investigations in Molise region, Southern Italy. *Ann Microbiol* 70:37
- Mortazavi R, Attiya S, Ariya PA (2015) Arctic microbial and next-generation sequencing approach for bacteria in snow and frost flowers: selected identification, abundance and freezing nucleation. *Atmos Chem Phys* 15: 6183–6204
- Nadkarni MA, Martin FE, Jacques NA, Hunter N (2002) Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology* 148:257–266
- Novak MA, Bode RW (1992) Percent model affinity: a new measure of macroinvertebrate community composition. *J N Am Benthol Soc* 11:80–85
- Pielou EC (1969) *An introduction to mathematical ecology*. Wiley, New York
- Rainey FA, Ray K, Ferreira M, Gatz BZ, Nobre MF, Bagaley D et al (2005) Extensive diversity of ionizing-radiation-resistant bacteria recovered from Sonoran Desert soil and description of nine new species of the genus *Deinococcus* obtained from a single soil sample. *Appl Environ Microb* 71: 5225–5235
- Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Phil Trans R Soc B* 360:1805–1811
- Segawa T, Miyamoto K, Ushida K, Agata K, Okada N, Kohshima S (2005) Seasonal change in bacterial flora and biomass in mountain snow from the Tateyama Mountains, Japan, analyzed by 16S rRNA gene sequencing and real-time PCR. *Appl Environ Microb* 71:123–130
- Shannon C, Weaver W (1963) *The mathematical theory of communication*. Illinois University Press, Urbana
- Spellerberg IF (2005) *Monitoring ecological change*. Cambridge University Press, Cambridge
- Stocchi P, Davolio S (2017) Intense air-sea exchanges and heavy orographic precipitation over Italy: the role of Adriatic sea surface temperature uncertainty. *Atmos Res* 196:62–82
- Teague KA, Gallicchio N (2017) *The evolution of meteorology: a look into the past, present, and future of weather forecasting*. Wiley, Oxford
- Vanwonterghem I, Jensen PD, Ho DP, Batstone DJ, Tyson GW (2014) Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. *Curr Opin Biotech* 27:55–64
- Varin T, Lovejoy C, Jungblut AD, Vincent WF, Corbeil J (2012) Metagenomic analysis of stress genes in microbial mat communities from Antarctica and the High Arctic. *Appl Environ Microb* 78:549–559
- Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microb* 73:5261–5267
- Warden JG, Casaburi G, Omelon CR, Bennett PC, Breecker DO, Foster JS (2016) Characterization of microbial mat microbiomes in the modern thrombolite ecosystem of Lake Clifton, Western Australia using shotgun metagenomics. *Front Microbiol* 7:1064
- Xiang SR, Chen Y, Shang TC, Jing ZF, Wu G (2010) Change of microbial communities in glaciers along a transition of air masses in western China. *J Geophys Res* 115:G04014
- Xiao CD, Wang SJ, Qin DH (2015) A preliminary study of cryosphere service function and value evaluation. *Adv Clim Change Res* 6:181–187
- Yang L, Yang HL, Tu ZC, Wang XL (2016) High-throughput sequencing of microbial community diversity and dynamics during douchi fermentation. *Plos One* 11:e0168166
- Zhang S, Yang G, Wang Y, Hou S (2010) Abundance and community of snow bacteria from three glaciers in the Tibetan Plateau. *J Environ Sci* 22:1418–1424

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

