Antarctic Science 29(1), 3–15 (2017) © Antarctic Science Ltd 2016 doi:10.1017/S0954102016000389

Synthesis
A primer to metabarcoding surveys of Antarctic terrestrial biodiversity

PAUL CZECHOWSKI[1,2], LAURENCE J. CLARKE[1,3,4], ALAN COOPER[1] and MARK I. STEVENS[5,6]

[1]*Australian Centre for Ancient DNA, University of Adelaide, Adelaide, SA 5000, Australia*
[2]*Antarctic Biological Research Initiative, 31 Jobson Road, Bolivar, SA 5110, Australia*
[3]*Australian Antarctic Division, Channel Highway, Kingston, TAS 7050, Australia*
[4]*Antarctic Climate & Ecosystems Cooperative Research Centre, University of Tasmania, Private Bag 80, Hobart, TAS 7001, Australia*
[5]*South Australian Museum, GPO Box 234, Adelaide, SA 5000, Australia*
[6]*School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, SA 5000, Australia*
paul.czechowski@abiori.org

Abstract: Ice-free regions of Antarctica are concentrated along the coastal margins but are scarce throughout the continental interior. Environmental changes, including the introduction of non-indigenous species, increasingly threaten these unique habitats. At the same time, the unique biotic communities subsisting in isolation across the continent are difficult to survey due to logistical constraints, sampling challenges and problems related to the identification of small and cryptic taxa. Baseline biodiversity data from remote Antarctic habitats are still missing for many parts of the continent but are critical to the detection of community changes over time, including newly introduced species. Here we review the potential of standardized (non-specialist) sampling in the field (e.g. from soil, vegetation or water) combined with high-throughput sequencing (HTS) of bulk DNA as a possible solution to overcome some of these problems. In particular, HTS metabarcoding approaches benefit from being able to process many samples in parallel, while workflow and data structure can stay highly uniform. Such approaches have quickly gained recognition and we show that HTS metabarcoding surveys are likely to play an important role in continent-wide biomonitoring of all Antarctic terrestrial habitats.

Received 28 January 2016, accepted 8 July 2016, first published online 13 September 2016

Key words: biomonitoring, conservation, cryptic taxa, high-throughput sequencing, invasive taxa

Introduction

Although only 0.3% of Continental Antarctica is ice-free, Antarctica is home to many organisms including bacteria, unicellular eukaryotes, fungi, lichen, cryptogamic plants and invertebrates that are scattered across the continent and subsist in isolated, remote, island-like habitats (Convey *et al.* 2014), for example, in soils, lakes and cryoconite holes. Availability of biodiversity information from these Antarctic areas is required for three major reasons. First, such data facilitate the investigation of glacial constraints and effects on current biodiversity (Convey *et al.* 2009). Second, it allows investigation of the effects of environmental change on Antarctic ecosystems (Nielsen & Wall 2013), and finally, conservation management becomes possible, also in light of increasing threats from non-indigenous invasive species (Chown *et al.* 2012a, 2012b, 2015b). However, knowledge of terrestrial Antarctic biodiversity is still limited because the vast majority of Antarctica's ice-free areas remain un- or under-studied (McGaughran *et al.* 2011, Convey *et al.* 2014).

Biodiversity research of ice-free habitats in Antarctica is complicated. First, logistic difficulties exacerbated by the harsh environmental conditions may limit biological research to the proximity of research stations, when more extensive field work is required (Convey 2010). Second, traditional biodiversity assessments of many multicellular eukaryotes include manual sorting and morphological identification, which are time consuming and require specific taxonomic expertise, especially for the cryptic and inconspicuous terrestrial life of Antarctica (e.g. Velasco-Castrillón *et al.* 2014a). Molecular methods are better suited to the study of Antarctic biota (Rogers 2007), but may lack resolution when sequence information is not considered (e.g. the analysis of terminal restriction fragment length polymorphisms or similar techniques; Magalhaes *et al.* 2012, Makhalanyane *et al.* 2013, Dreesens *et al.* 2014) or may be labour intensive (e.g. Sanger sequencing; Lawley *et al.* 2004, Fell *et al.* 2006, Velasco-Castrillón & Stevens 2014, Velasco-Castrillón *et al.* 2014b, 2014c).

Readily applied in many other parts of the world (reviewed in Bik *et al.* 2012a, Bohmann *et al.* 2014), metabarcoding approaches (*sensu* Taberlet *et al.* 2012a) present an opportunity to rapidly generate baseline biodiversity information for a variety of terrestrial

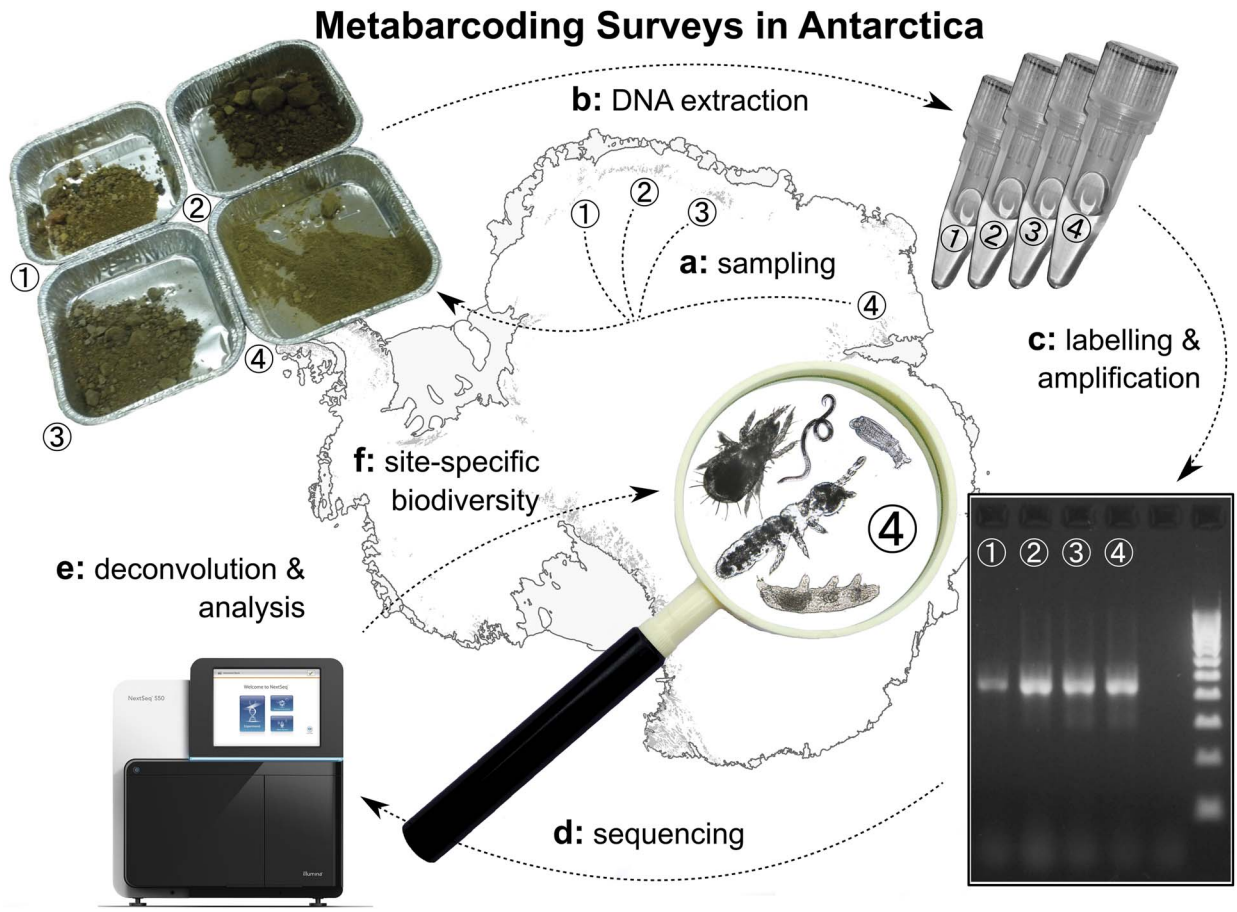

Fig. 1. Workflow for metabarcoding analyses, which can be applied to soil, snow, ice, cryconite holes, lake sediments or nearshore marine environments. **a.** Samples are collected. **b.** The genetic material is extracted in bulk from individual samples. **c.** DNA contained in extracts is amplified with genetic markers and sequencing adapters, multiplex identifier (MID) tags are added. **d.** The library is processed on a high-throughput sequencing device. **e.** After data deconvolution according to sample, reference information assigns individual sequences or sequence clusters with taxonomic information. **f.** Distributional information becomes available. Picture of sequencing device provided courtesy of Illumina (San Diego, CA, USA). Base layers courtesy of the Scientific Committee on Antarctic Research Antarctic Digital Database.

Antarctic habitats (Fig. 1; Chown *et al.* 2015b). Metabarcoding approaches use the genetic material from bulk environmental samples such as soil, permafrost, water, ice, snow or other substrates (Bohmann *et al.* 2014). Then DNA from multiple organisms contained in such samples are identified for taxonomic analyses either with traditional Sanger sequencing or more recently using high-throughput sequencing (HTS; Chown *et al.* 2015b, Cowan *et al.* 2015, Czechowski *et al.* 2016). In a global context, HTS-supported metabarcoding approaches have been applied to monitoring invasive species and surveying biodiversity over large spatial scales (Drummond *et al.* 2015 and reviewed in Bohmann *et al.* 2014). In Antarctica, metabarcoding studies, initially based on Sanger sequencing, have enabled the identification of cryptic organisms and communities such as fungi, yeast and invertebrates (Lawley *et al.* 2004, Fell *et al.* 2006). These techniques have also been applied to viruses (López-Bueno *et al.* 2009), bacteria in hypolithic communities, soil and air (Makhalanyane *et al.* 2013, Bottos *et al.* 2014a, 2014b), as well as fungal and unicellular eukaryotes of soils (Dreesens *et al.* 2014, Niederberger *et al.* 2015). Additionally, the methodological pitfalls of these techniques when applied in Antarctica have become better understood (Lee *et al.* 2012b, Czechowski *et al.* 2016), including amplification and sequencing biases, coupled with sparse reference data. Collectively, HTS metabarcoding, despite not being without flaws, provides a promising method to rapidly gather biodiversity information from Antarctic habitats, with the ability to generate large amounts of biodiversity data from a wide range of taxa with simple sample collection, uniform laboratory workflows and comparable data structures.

https://doi.org/10.1017/S0954102016000389 Published online by Cambridge University Press

Here, we provide a technical introduction to HTS metabarcoding with an Antarctic focus and highlight the potential of such approaches for Antarctic biodiversity research beyond their current applications. This synthesis serves as a starting point for the development of Antarctic HTS metabarcoding surveys. We hope to encourage fellow researchers to participate in the joint effort of understanding Antarctica's biodiversity on a continental scale (Kennicutt *et al.* 2014).

Technical considerations

Metabarcoding projects are influenced by biases inherent to several methodological aspects. These include: i) extraction of representative DNA from a mixed template sample and the intra- and extracellular DNA contained in such a sample, ii) platform-specific sequencing technologies including inherent sequence error patterns, iii) the appropriate choice of markers, iv) methods for generation and v) amplification of sequencing libraries. Finally, informed approaches to HTS data processing and analysis are necessary to achieve research goals.

Sample selection

As shown in a variety of global studies (reviewed in Bohmann *et al.* 2014), it is possible to extract DNA suitable for metabarcoding analyses from a variety of substrates, which offers a unique opportunity to study different environments in Antarctica (Fig. 1a). DNA can be extracted from organisms contained in surface soil (Czechowski *et al.* 2016), permafrost (Bellemain *et al.* 2013), snow (Dalén *et al.* 2007), ice (Willerslev *et al.* 2004), freshwater benthos of lakes (Hajibabaei *et al.* 2012) or nearshore marine sediments (Powell *et al.* 2003). Furthermore, extracts of pre-sorted samples can be analysed (Drummond *et al.* 2015), such as from museum collections. When limited starting material is available, preservatives such as ethanol can be used as a DNA source (Shokralla *et al.* 2010). DNA can also be extracted from faeces (Jarman *et al.* 2013), for example the seal and penguin colonies in coastal regions of Antarctica and sub-Antarctic islands. The variety of potential sample types, coupled with cost-effective sequence data generation, could address problems related to surveying large spatial or temporal scales.

Extraction of environmental samples

Failure to extract representative DNA from a sample (Fig. 1b), so-called 'extraction bias' (Pedersen *et al.* 2014), is a major concern for metabarcoding approaches. Such biases occur when extraction methods inconsistently lyse cells of different organisms, and are further biased by the presence of dead organisms' DNA in substrates (Pedersen *et al.* 2014). Some authors explicitly distinguish an intracellular DNA component from an extracellular DNA component in bulk extracts, and present methods to quantify both fractions in a given sample (Ascher *et al.* 2009). Yet, applying such approaches across large numbers of samples may be cost-prohibitive. Alleviating DNA extraction biases can be achieved through combining different extraction methods, and include blending samples prior to extraction and/or using a large amount of starting material (Delmont *et al.* 2011, 2013, Taberlet *et al.* 2012b). Yet, different extraction methods or batch-wise application of one extraction method may introduce variable levels of non-template contamination (Salter *et al.* 2014). Therefore, randomized drawing of sample batches is recommended (Salter *et al.* 2014). Extraction biases and contamination can be discovered by inclusion of negative and positive controls. Negative controls facilitate the detection of contamination. Positive controls of known taxonomic composition are helpful in detecting compositional deviations between the sequence data and sample source (Salter *et al.* 2014, Czechowski *et al.* 2016). Consequently, both positive and negative controls help to optimize the DNA extraction process and are helpful in streamlining processing parameters in steps following extraction.

High-throughput sequencing platforms

The recent advance of HTS-supported metabarcoding, metagenetics and metagenomics (Bik *et al.* 2012a, Taberlet *et al.* 2012a, Bohmann *et al.* 2014) can be considered a consequence of continuing development of sequencing platforms by companies such as 454 (Roche, Basel, official platform support was discontinued in 2016), Illumina (San Diego, CA, USA), IonTorrent (Thermo Fisher Scientific, Waltham, MA, USA) and others since 2005 (Glenn 2011, van Dijk *et al.* 2014). These devices generate substantially larger amounts of sequencing data than chain-termination sequencing (Bohmann *et al.* 2014), but in comparison produce shorter reads (i.e. ~100–800 base pairs, depending on the technology). Using these platforms in conjunction with metabarcoding (and metagenomic) approaches removes the need to process mixed DNA templates through clone libraries and hence substantially reduces the time to data generation. The platform of choice to conduct metabarcoding biodiversity surveys currently appears to be one of the Illumina platforms, due to the large number of sequences generated which reduces the cost per base and the comparatively low error rate of this sequencing platform (Bokulich *et al.* 2013, Bragg *et al.* 2013). The discontinued 454 platform, although often comparatively expensive to use, will continue for some time to offer the longest read lengths of all platforms

https://doi.org/10.1017/S0954102016000389 Published online by Cambridge University Press

suitable for amplicon sequencing (Van Dijk *et al.* 2014). Comprehensive reviews of HTS platforms are provided in Glenn (2011) and van Dijk *et al.* (2014). Currently, the most common and cost-effective approach to generate metabarcoding information with HTS is parallel sequencing of PCR-amplified bulk DNA extracts, known as 'amplicon sequencing' (Taberlet *et al.* 2012a, Bohmann *et al.* 2014). Important methodological aspects of amplicon sequencing are described below. We also describe how pitfalls of amplicon sequencing can be alleviated and present alternative methods for library generation and sequencing.

Marker choice

Markers for PCR amplification of mixed DNA templates extracted from environmental samples (e.g. soil (Fig. 1c), permafrost, water, ice, snow, etc.) should i) ideally amplify all taxa with similar efficiency despite potential mismatches between primers and the variety of template molecules (Clarke *et al.* 2014a), ii) amplify target regions short enough to allow amplification of degraded DNA, particularly if targeting extracellular DNA (Riaz *et al.* 2011, Coissac *et al.* 2012, Taberlet *et al.* 2012a), iii) exhibit the least possible amount of degenerate bases to allow the application of high annealing temperatures, while decreasing the risk of chimeric amplification (Lenz & Becker 2008, Ahn *et al.* 2012) and iv) target a gene region for which ample reference data are available to allow taxonomic identification of phylotypes (see below).

Finding a primer pair that possesses these desirable, possibly incompatible, qualities is challenging. Two genes that have been widely applied in single-gene and metabarcoding analyses of metazoans, for example, are the nuclear 18S ribosomal DNA (18S rDNA) and mitochondrial cytochrome c oxidase subunit I (COI) genes (Wu *et al.* 2011, Zhan *et al.* 2014). These markers are favoured due to their long history of application, resulting in comparatively abundant reference data in sequence repositories such as GenBank, BOLD and SILVA (Pruesse *et al.* 2007, Ratnasingham & Hebert 2007, Benson *et al.* 2011) (Fig. 1e & f). However, 18S rDNA data may underestimate biodiversity due to low taxonomic resolution, and many COI markers show inherent taxonomic bias due to insufficiently conserved primer binding sites across broad taxonomic groups (Tang *et al.* 2012, Deagle *et al.* 2014). Similar advantages and disadvantages are found analogously in other marker regions applied in metabarcoding studies, for example when targeting fungi using the ITS region or photosynthetic cryptogams via the matK and chloroplast genes (CBOL Plant Working Group 2009, Orgiazzi *et al.* 2013, Drummond *et al.* 2015).

Library generation

Preparing DNA for HTS requires the addition of platform-specific sequencing adapters, and often (particularly for metabarcoding) sample-specific sequence tags (or 'multiplex identifier' (MID) tags) are also required to enable deconvolution of sequence data (Fig. 1c & e). Initially, DNA pools were furnished with MID tags during PCR (Saiki *et al.* 1988) via extended primer sequences or ligation of unmodified primers preceding sequence adapter ligation (Binladen *et al.* 2007, Meyer *et al.* 2008). More recently, library generation via long primer sequences carrying both sequence adaptor and MID tags (fusion primers) has become common (Bik *et al.* 2012a). The application of fusion primers is practical in that it only requires a single PCR, but may be costly for large numbers of samples and difficult for primer lengths above ~50 base pairs due to poor PCR performance. In those cases, more labour intensive ligation protocols may be a better choice (Stiller *et al.* 2009, Kircher *et al.* 2012, O'Neill *et al.* 2013). Also of concern is the informed choice of MID tags. Owing to possible flaws in the underlying algorithms, these tags may not meet the intended expectations of robustness towards sequencing errors (Faircloth & Glenn 2012). Only MID tags that have been explicitly tested for correct Hamming distances (Hamming 1950) are recommended, and this will later enable correct deconvolution and error correction (Faircloth & Glenn 2012).

Amplification

Concordance between the taxonomic composition of a mixed DNA template retrieved from environmental bulk samples and the amplified library requires careful calibration of PCR conditions, for example, length optimization of denaturation, annealing and extension steps as well as the correct temperatures for the primer annealing phase (Fig. 1c). Possible pitfalls include i) introduction of substitutions and insertion/deletions through polymerase activity (Cline *et al.* 1996), ii) formation of chimeric molecules in late amplification stages (Kanagawa 2003), iii) amplification bias when using degenerate primers in combination with high annealing temperatures (Cline *et al.* 1996, Kanagawa 2003), and iv) failure to detect rare variants when little replication is applied (Ficetola *et al.* 2015). Such pitfalls collectively threaten the credibility of the resulting sequence data (Czechowski *et al.* 2016). They may i) alter the similarity of phylotypes to reference sequences, ii) result in artificial phylotypes that match several reference sequences, iii) artificially enrich phylotypes whose library molecules matched the PCR primers well or iv) result in false-negative concealment of phylotypes. Retrieval of higher quality data can be achieved by i) application of proofreading polymerases (Taberlet *et al.* 2012a),

https://doi.org/10.1017/S0954102016000389 Published online by Cambridge University Press

Table I. Selection of analysis software for metabarcoding data of environmental DNA. Possible tasks related to handling of metabarcoding data provided in columns. Multi-purpose tools in top section are suitable for various sequence analysis tasks. Software environments in the middle section are specialized for metabarcoding analysis. Programs and packages in the bottom section are focussed on ecological analysis and/or visualizing results. X = functionality of a given tool for a given task is provided, (X) = functionality of a given tool for a given task is provided to some extent, - = functionality is not provided.

Name	Interface	Marker development	Quality check	Read trimming	Read cleaning	Sequence annotation	Phylotype analysis	Metadata analysis	Selected reference	Web link
AdapterRemoval	Unix shell	-	-	X	-	-	-	-	Lindgreen 2012	https://github.com/slindgreen/AdapterRemoval
Trimmomatic	Unix shell	-	-	X	-	-	-	-	Lohse *et al.* 2012	http://www.usadellab.org/cms/?page=trimmomatic
FastQC	Unix shell, graphical	-	X	-	-	-	-	-	-	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Galaxy	Graphical interface	-	X	X	X	(X)	(X)	-	Giardine *et al.* 2005	https://usegalaxy.org/
UPARSE	Unix shell	X	X	X	X	(X)	-	-	Edgar 2013	http://www.drive5.com/
OBItools,	Unix shell	X	X	X	X	X	(X)	(X)	Boyer *et al.* 2016	http://www.grenoble.prabi.fr/trac/OBITools
QIIME	Unix shell	-	X	X	(X)	X	X	(X)	Caporaso *et al.* 2010	http://qiime.org/
MOTHUR	Unix shell	-	X	X	(X)	X	X	(X)	Schloss *et al.* 2009	http://www.mothur.org/
MG-RAST	Graphical interface	-	X	(X)	(X)	X	(X)	(X)	Wilke *et al.* 2016	http://metagenomics.anl.gov/
MEGAN	Graphical interface	-	-	-	-	X	X	-	Huson & Weber 2013	http://ab.inf.uni-tuebingen.de/software/megan5/
PHYLOSEQ	R	-	-	-	-	-	X	(X)	McMurdie & Holmes 2013	https://joey711.github.io/phyloseq/
VEGAN	R	-	-	-	-	-	(X)	X	Oksanen *et al.* 2015	http://cran.r-project.org/web/packages/vegan/index.html
EXPLICET	Graphical interface	-	-	-	-	-	X	-	Robertson *et al.* 2013	http://www.explicet.org/

ii) using few and long PCR cycles (Kanagawa 2003, Lenz & Becker 2008, Ahn *et al.* 2012), iii) careful testing of annealing temperature (Sipos *et al.* 2007) and iv) processing three or more PCR replicates (Gilbert *et al.* 2010). Analogous to the extraction step, positive and negative controls are important to track contamination during the amplification procedure (Czechowski *et al.* 2016). At the same time, positive controls may be a source of cross-contamination, for example, through unintended PCR-product carry-over (Kwok 1990). Using suitable non-Antarctic control DNA, which can be distinguished from sample DNA in later analysis steps, could reduce the impact of PCR-product carry-over.

Sequence analysis

Most importantly, it needs to be noted that processing HTS data (Fig. 1e & f) is not straightforward and requires a high level of bioinformatics expertise, project-specific software selection and software fine-tuning at every step. To perform a metabarcoding analysis with any given raw dataset, first an analysis workflow needs to be conceptualized. Then, a variety of software algorithms need to be selected with regard to the analysis steps and study goals, keeping in mind available computing hardware, methods of library design, employed sequencing technology, data volume and analysis pitfalls (Coissac *et al.* 2012, Lee *et al.* 2012b). Subsequently, testing programs individually and in order of application using small datasets is advisable. Here, it may be necessary to generate custom (or at least modify existing) scripts, through which data input and output of algorithms is handled and connected.

It is possible that the resulting metabarcoding analysis workflow is initiated by marker selection (Riaz *et al.* 2011), and once sequence data has been generated, several raw data processing steps will follow before the statistical analysis can be attempted (Bik *et al.* 2012a, Bohmann *et al.* 2014). Raw data preparation typically includes quality filtering, removal of sequence adapters, data deconvolution and chimera removal. The clean raw data are then typically clustered, assigned with taxonomy, and subsequently, the resulting data are checked for their suitability for the intended statistical analysis.

Although raw data preparation can be achieved with a variety of programs (see Table I for examples), software environments dedicated to metabarcoding analysis such as QIIME, MOTHUR and MG-RAST (Schloss *et al.* 2009, Caporaso *et al.* 2010, Wilke *et al.* 2016) offer functionality incorporating whole analysis workflows starting from raw data cleaning, phylotype clustering and basic statistical analyses. These metabarcoding/metagenomic software environments themselves usually take advantage of multiple algorithms dedicated to particular sub-routines of analysis workflows.

https://doi.org/10.1017/S0954102016000389 Published online by Cambridge University Press

For example, chimera detection may be achieved with UPARSE (Edgar 2013) in QIIME. Taxonomic assignments may be retrieved with BLAST (Altschul *et al.* 1990) or other algorithms such as the RDP classifier, or UCLUST (Edgar 2010, Lan *et al.* 2012). In general, sub-algorithms employed by software suites need to be carefully considered before attempting data preparation and analysis steps. An overview of raw data preparation software is provided by Zhou & Rokas (2014).

If dedicated metabarcoding/metagenomic analysis environments do not offer desired functionalities for analysis and visualization, some analyses can be achieved through other available software and possibly linked in via 'glue code' written in programming languages such as R (R Development Team 2016), BASH or PYTHON (Van Rossum & Drake 1995). EXPLICET (Robertson *et al.* 2013), for example, offers basic visualization and statistical analysis functionally coupled with a graphical user interface, suitable for novice users. More powerful, but command-driven, the R environment offers several packages for the statistical analysis and visualization of metabarcoding data with packages such as PHYLOSEQ or VEGAN (McMurdie & Holmes 2013, Oksanen *et al.* 2015, R Development Team 2016) and many others. An in-depth review of metabarcoding and metagenomic sequence analysis software is provided by Lindgreen *et al.* (2016).

Biodiversity surveys often seek to quantify α-diversity (species richness) and β-diversity (change in community composition; Whittaker 1960). Three pitfalls should be carefully considered when designing metabarcoding analysis workflows aiming at α- and β-diversity comparisons between a given set of samples, and while relating species occurrences to, for example, their environment. All analyses require a) sufficient sequencing depth, b) α- and β-diversity comparisons require appropriate abundance correction of libraries with different read depths and c) distance-based ordination techniques can only be applied appropriately when using the correct distance measure. Regarding a) above, it is important to realize that sequencing depth of HTS libraries is crucial for the reliable estimation of biodiversity measures in the resulting data (Smith & Peay 2014); while in order to retrieve reliable biodiversity estimates from Antarctic habitats, a high sequencing effort may be necessary to retrieve credible statistical results (Czechowski *et al.* 2016). Regarding b) above, the analysis of differentially abundant phylotypes in metabarcoding data, by comparisons of proportions or rarefied counts, although applied widely, is inappropriate and may yield misleading results (McMurdie & Holmes 2014). Instead of applying such rarefaction methods, other algorithms should be employed to enable comparison between libraries with coverage differences. For instance, R packages DESeq and edgeR offer alternative ways to correct phylotype abundance (Robinson *et al.* 2009, Anders & Huber 2010). Regarding c) above, scarcity of biological data is known to impair ecological statistical analysis in Antarctica due to the low spatial overlap of individual phylotypes (Magalhaes *et al.* 2012, Czechowski *et al.* 2016). Consequently, metabarcoding data from many Antarctic habitats is likely to be difficult to analyse with commonly used distance-based ordination methods including multidimensional scaling (Wish & Carroll 1982), constrained analysis of principal components (CAP) and redundancy analysis (RDA) (Legendre & Andersson 1999). Hence, when employing distance-based ordination approaches, sample comparison may only be possible with metrics established to be suitable, for example the Hellinger distance (Gagné & Proulx 2009). Alternatively, several different distance metrics should be compared (Blanchet *et al.* 2014). Model-based ordination methods may circumvent drawbacks of distance-based ordinations, and should be used where possible (Ellis *et al.* 2012, Wang *et al.* 2012, Hui *et al.* 2015).

Recent improvements of high-throughput sequencing metabarcoding

Retrieving biodiversity information from hundreds of samples over large spatial or temporal scales requires cost-efficient processing. Tagging individual samples with fusion primers for amplicon sequencing is simple, but increases the cost of HTS metabarcoding studies for large-scale approaches. Presumably for this reason, numbers of parallel processed samples in several recent global and Antarctic metabarcoding studies range from seven to twelve samples (Bik *et al.* 2012b, Roesch *et al.* 2012, Dreesens *et al.* 2014, Niederberger *et al.* 2015). Reducing primer-associated costs is possible through modular combination of multiple sequence tags per sample, thus reducing the amount of unique oligonucleotides required for a project. Examples of such modular workflows include using two PCRs to double-tag amplicons for HTS (Bybee *et al.* 2011, de Cárcer *et al.* 2011). Similarly, double-tagging can generate amplicons with minimal work, handling and cost in a single PCR (Clarke *et al.* 2014b).

The PCR biases during library preparation can be alleviated through the application of hybridization approaches. In hybridization approaches, libraries are generated by annealing target DNA to biotinylated oligonucleotide probes (Gnirke *et al.* 2009, Faircloth *et al.* 2012, Lemmon *et al.* 2012). In comparison to PCR, hybridization approaches enable retrieval of multiple conserved regions per reaction, perform well in detecting rare DNA and reduce compositional biases in the resulting data without the need for extensive replication (Taberlet *et al.* 2012a). For example, Denonfoux *et al.* (2013) sequenced bacterial DNA derived from environmental samples after enrichment

https://doi.org/10.1017/S0954102016000389 Published online by Cambridge University Press

with a hybridization approach, demonstrating the benefits outlined here for mixed template DNA sources.

The lengths of genomic regions that can be targeted with single read lengths of a given HTS platform are variable (see section 'High-throughput sequencing platforms'), but usually shorter than the 600–1000 base pairs that can be achieved from a single read using Sanger sequencing technology. Therefore, recent research has investigated the options of adopting shorter fragments of regions that have been used widely in Sanger sequencing, for example, the beginning of the COI gene region or the 18S gene (Machida & Knowlton 2012, Leray *et al.* 2013). Other studies have identified new marker regions with short read lengths suitable for HTS technologies, which retain adequate information allowing comparisons with data from the traditional markers (CBOL Plant Working Group 2009, Epp *et al.* 2012). Furthermore, identification of custom marker regions is now possible with bioinformatics tools such as ecoPrimers incorporated into OBI tools (Riaz *et al.* 2011, Boyer *et al.* 2016) (see Table I). ecoPrimers employs user-curated reference data retrieved from repositories such as GenBank (Benson *et al.* 2011) to identify conserved regions suitable for project-specific primer design for mixed template amplification.

Other methods of streamlining metabarcoding approaches with regard to data yields and cost efficiency have become available. A combination of shotgun sequencing methods and amplicon sequencing, for instance, allow retrieval of full length COI sequences using HTS technology (Liu *et al.* 2013). Furthermore, the omission of library quantification, and instead pooling libraries by volume (coupled with shearing and re-assembly; Feng *et al.* 2015), can reduce time and effort during library construction. Finally, with decreasing sequencing costs, metagenomic studies targeting the entirety of DNA molecules in a sample, including functional genes, without selective amplification or enrichment (Fierer *et al.* 2012) may become viable for large sample numbers.

The potential of metabarcoding and metagenomics for Antarctic biology

Elucidating community structures

Community-level interaction is an important feature of Antarctic ecosystems. Such interactions were believed to be minimal, perhaps owing to the fact that they are hard to measure (Hogg *et al.* 2006). However, biotic community-level interactions are increasingly implicated in facilitating survival in harsh environments, and may be observable through stratified occurrence of different organisms or the exchange of nutrients between strata within communities (Nakai *et al.* 2012, Pointing & Belnap 2012). Community-level organization has been discovered among Antarctic soil crusts, lithobiontic communities, eukaryotes in moss pillars and cyanobacterial mats (Jungblut *et al.* 2012, Nakai *et al.* 2012, Makhalanyane *et al.* 2013, Colesie *et al.* 2014). Evidence for biotic interactions has also been reported among soil arthropods of sub-Antarctic islands (Caruso *et al.* 2013).

Studies describing the community-level organization of Antarctic terrestrial ecosystems will benefit from metabarcoding and metagenomic approaches. Possible studies could include further analyses of prokaryotic and eukaryotic diversity in substrates such as snow, soil crusts and hypolithons, photobiotic and mycobiotic diversity and biogeography of lichen, or the association between fungi and eukaryotes in moss communities, that are still often studied using Sanger sequencing (Carpenter *et al.* 2000, Fernández-Mendoza *et al.* 2011, Khan *et al.* 2011, Jungblut *et al.* 2012, Gokul *et al.* 2013, Altermann *et al.* 2014). The HTS-supported analysis of such communities is becoming more common for eukaryotes and bacteria, e.g. in cyanobacterial mats and hypolithic communities (Lee *et al.* 2012a, Dreesens *et al.* 2014, Niederberger *et al.* 2015). However, similar approaches could be applied to environments such as air (Bottos *et al.* 2014b) or nearshore sediments (Powell *et al.* 2003). Functional aspects of soil microbial communities have been investigated using metagenomic HTS approaches, providing an in-depth picture of ecosystem services (Fierer *et al.* 2012). Similarly, HTS-based metagenomic studies advanced the description of morphologically conserved, rare or small cryptic communities in Antarctica, including their provision of ecosystem services (Goordial *et al.* 2016).

Supporting conservation of Antarctica

The biodiversity and distribution of the terrestrial Antarctic biota is more heterogeneous than anywhere else in the world (Ettema & Wardle 2002, Convey *et al.* 2014, Chown *et al.* 2015a). Large distances between habitats, unique geological and glacial histories, different soil compositions and extreme fluctuations of abiotic conditions amplify this heterogeneity (Bockheim 1997, Marchant & Head 2007, Bintanja *et al.* 2014). Consequently, Antarctic biota exhibit a high degree of endemism and costly adaptation mechanisms to withstand harsh environmental conditions (Convey 1997, Convey & Stevens 2007). Human-mediated environmental changes are anticipated to have profound effects on the spatial extent and structure of Antarctic terrestrial ecosystems (Chown *et al.* 2012a, 2012b). Despite a high degree of isolation between continental habitats (Convey *et al.* 2014), the distribution patterns of Antarctic species may shift southwards and increasingly overlap, possibly eroding the extensive endemism among many Antarctic species (Nielsen & Wall 2013), particularly when considering human-mediated dispersal. Additionally, non-indigenous species may outcompete local endemics in an increasingly accommodating environment, particularly in

https://doi.org/10.1017/S0954102016000389 Published online by Cambridge University Press

the sub-Antarctic (Frenot *et al.* 2005, Hughes & Convey 2010, Hughes *et al.* 2010).

Current Antarctic biology is primarily influenced by the desire to conserve the unique and still largely uncharacterized biodiversity of the continent and surrounding islands. Elucidating distribution patterns of terrestrial communities and identifying biotic elements most vulnerable to climate change have been deemed some of the most important goals of Antarctic biological conservation (Kennicutt *et al.* 2014). Definition and extension of protected areas in Continental Antarctica, particularly in remote locations, is urgently required (Terauds *et al.* 2012, Shaw *et al.* 2014), coupled with increased monitoring of these areas for the introduction of taxa from the sub- and Maritime Antarctic (Chown *et al.* 2012a, Shaw *et al.* 2014, McGeoch *et al.* 2015). Efforts to capture heterogeneous patterns in terrestrial biodiversity, and to assess the future impact of alien species, require densely spaced biological and environmental survey data (Shaw *et al.* 2014, McGeoch *et al.* 2015). The HTS-supported molecular methods are particularly powerful in resolving Antarctic endemics from non-indigenous species that are not easily detected or are difficult to identify (Hughes & Convey 2012, Chown *et al.* 2015b); such methods could inform, for example, on the number of eukaryotic alien and invasive species per biogeographical region in standardized frameworks (McGeoch *et al.* 2015).

Continent-wide survey data and time series monitoring

The HTS-based metabarcoding approaches are regarded as more efficient compared with morphological methods for assessing the ecological integrity and health of diverse marine and terrestrial environments, by providing a uniform, swift and economical means of species identification (Aylagas *et al.* 2014, Drummond *et al.* 2015). Potential applications to Antarctic environments are now being realized, where HTS-based metabarcoding studies similarly offer a simple, cost-efficient workflow and rich sequence information that can be easily combined or re-analysed in more detailed integrative studies (Gutt *et al.* 2012, Chown *et al.* 2015b).

In order to use the full potential of HTS for Antarctic biodiversity research and ecology, we suggest i) designing studies with close consideration of research goals defined by the international community (Kennicutt *et al.* 2015), ii) designing studies with larger numbers of samples, for example, across variable spatial and temporal scales, similar to approaches used by Dornelas *et al.* (2014) and Howard-Williams *et al.* (2006) or contributing towards such efforts, iii) using a variety of DNA sources for analysis, including historical material from museum collections or historical Antarctic voyages (Headland 2009), iv) providing well-documented analysis code with all published HTS data, v) further developing laboratory and analysis protocols for metabarcoding and metagenomic approaches suitable to investigate Antarctic habitats, and finally, vi) generating reference DNA sequences for Antarctic species identification using α taxonomic (including morphological) approaches (Turrill 1938).

Summary and conclusions

Metabarcoding analysis of mixed template and environmental DNA is a valuable option to describe the composition and distribution of the cryptic and heterogeneously distributed terrestrial biota of Antarctica. Metabarcoding and metagenomic approaches have proven helpful in describing bacterial and hypolithic communities in ice-free regions of Antarctica and could similarly be applied to many other taxa on the continent, including communities inhabiting snow and ice, as well as lake and marine sediments. In comparison to traditional molecular methods, HTS-based approaches yield large amounts of detailed data with relatively simple and time-efficient laboratory workflows, coupled with straightforward fieldwork. Multiple laboratory developments have recently improved the cost efficiency of PCR-based library generation allowing parallel processing of large sample numbers. Drawbacks of amplicon library generation can be alleviated by alternative library preparation methods. By providing a consistent and efficient means of species identification, as well as insights into the functional diversity of such habitats, HTS-based metabarcoding and metagenomic studies will be a useful tool for assessing the ecological integrity and health of Antarctic habitats. When applied to large sample numbers, across large spatial scales and multiple biota, HTS-based metabarcoding and metagenomic approaches will improve our understanding of Antarctic terrestrial biodiversity on a continental scale.

Acknowledgements

The authors declare no competing interests. The Australian Antarctic Division provided funding under science project 2355 to M.S. The Australian Research Council supported this work through funds from linkage grant LP0991985 to A.C. and M.S. The University of Adelaide supported this project through an International Post-Graduate Research Scholarship to P.C. We thank two anonymous reviewers for helpful comments that improved the manuscript.

Author contribution

P.C. prepared, edited and revised the manuscript. M.S., L.C. and A.C. edited and revised the manuscript.

References

Ahn, J.H., Kim, B.Y., Song, J. & Weon, H.Y. 2012. Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. *Journal of Microbiology*, **50**, 1071–1074.

Altermann, S., Leavitt, S.D., Goward, T., Nelsen, M.P. & Lumbsch, H.T. 2014. How do you solve a problem like Letharia? A new look at cryptic species in lichen-forming fungi using Bayesian clustering and SNPs from multilocus sequence data. *PLoS ONE*, **9**, 10.1371/journal.pone.0097556.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Anders, S. & Huber, W. 2010. Differential expression analysis for sequence count data. *Genome Biology*, **11**, 10.1186/gb-2010-11-10-r106.

Ascher, J., Ceccherini, M.T., Pantani, O.L., Agnelli, A., Borgogni, F., Guerri, G., Nannipieri, P. & Pietramellara, G. 2009. Sequential extraction and genetic fingerprinting of a forest soil metagenome. *Applied Soil Ecology*, **42**, 10.1016/j.apsoil.2009.03.005.

Aylagas, E., Borja, Á. & Rodríguez-Ezpeleta, N. 2014. Environmental status assessment using DNA metabarcoding: towards a genetics based marine biotic index (gAMBI). *PLoS ONE*, **9**, 10.1371/journal.pone.0090529.

Bellemain, E., Davey, M.L., Kauserud, H., Epp, L.S., Boessenkool, S., Coissac, E., Geml, J., Edwards, M., Willerslev, E., Gussarova, G., Taberlet, P. & Brochmann, C. 2013. Fungal palaeodiversity revealed using high-throughput metabarcoding of ancient DNA from arctic permafrost. *Environmental Microbiology*, **15**, 1176–1189.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Sayers, E.W. 2011. GenBank. *Nucleic Acids Research*, **39**, 10.1093/nar/gkq1079.

Bik, H.M., Porazinska, D.L., Creer, S., Caporaso, J.G., Knight, R. & Thomas, W.K. 2012a. Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology & Evolution*, **27**, 233–243.

Bik, H.M., Sung, W., De Ley, P., Baldwin, J.G., Sharma, J., Rocha-Olivares, A. & Thomas, W.K. 2012b. Metagenetic community analysis of microbial eukaryotes illuminates biogeographic patterns in deep-sea and shallow water sediments. *Molecular Ecology*, **21**, 1048–1059.

Binladen, J., Gilbert, M.T.P., Bollback, J.P., Panitz, F., Bendixen, C., Nielsen, R. & Willerslev, E. 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE*, **2**, 10.1371/journal.pone.0000197.

Bintanja, R., Severijns, C., Haarsma, R. & Hazeleger, W. 2014. The future of Antarctica's surface winds simulated by a high-resolution global climate model: 2. Drivers of 21st century changes. *Journal of Geophysical Research - Atmospheres*, **119**, 7160–7178.

Blanchet, F.G., Legendre, P., Bergeron, J.A.C. & He, F.L. 2014. Consensus RDA across dissimilarity coefficients for canonical ordination of community composition data. *Ecological Monographs*, **84**, 10.1890/13-0648.1.

Bockheim, J.G. 1997. Properties and classification of cold desert soils from Antarctica. *Soil Science Society of America Journal*, **61**, 224–231.

Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Yu, D.W. & de Bruyn, M. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, **29**, 358–367.

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., Mills, D.A. & Caporaso, J.G. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, **10**, 10.1038/NMETH.2276.

Bottos, E.M., Scarrow, J.W., Archer, S.D.J., McDonald, I.R. & Cary, S.C. 2014a. Bacterial community structures of Antarctic soils. *In* Cowan, D.A., *ed. Antarctic terrestrial microbiology*. Berlin: Springer, 9–33.

Bottos, E.M., Woo, A.C., Zawar-Reza, P., Pointing, S.B. & Cary, S.C. 2014b. Airborne bacterial populations above desert soils of the McMurdo Dry Valleys, Antarctica. *Microbial Ecology*, **67**, 120–128.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P. & Coissac, E. 2016. OBITools: a UNIX-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, **16**, 176–182.

Bragg, L.M., Stone, G., Butler, M.K., Hugenholtz, P. & Tyson, G.W. 2013. Shining a light on dark sequencing: characterising errors in ion torrent PGM data. *PLoS Computational Biology*, **9**, 10.1371/journal.pcbi.1003031.

Bybee, S.M., Bracken-Grissom, H., Haynes, B.D., Hermansen, R.A., Byers, R.L., Clement, M.J., Udall, J.A., Wilcox, E.R. & Crandall, K.A. 2011. Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biology and Evolution*, **3**, 1312–1323.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., *et al.* 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Carpenter, E.J., Lin, S.J. & Capone, D.G. 2000. Bacterial activity in South Pole snow. *Applied and Environmental Microbiology*, **66**, 10.1128/AEM.66.10.4514-4517.2000.

Caruso, T., Trokhymets, V., Bargagli, R. & Convey, P. 2013. Biotic interactions as a structuring force in soil communities: evidence from the micro-arthropods of an Antarctic moss model system. *Oecologia*, **172**, 10.1007/s00442-012-2503-9.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 12 794–12 797.

Chown, S.L., Clarke, A., Fraser, C.I., Cary, S.C., Moon, K.L. & McGeoch, M.A. 2015a. The changing form of Antarctic biodiversity. *Nature*, **522**, 10.1038/nature14505.

Chown, S.L., Hodgins, K.A., Griffin, P.C., Oakeshott, J.G., Byrne, M. & Hoffmann, A.A. 2015b. Biological invasions, climate change and genomics. *Evolutionary Applications*, **8**, 10.1111/eva.12234.

Chown, S., Huiskes, A.H.L., Gremmen, N.J.M., Lee, J.E., Terauds, A., Crosbie, K., Frenot, Y., Hughes, K.A., Imura, S., Kiefer, K., Lebouvier, M., Raymond, B., Tsujimoto, M., Ware, C., van de Vijver, B. & Bergstrom, D.M. 2012a. Continent-wide risk assessment for the establishment of nonindigenous species in Antarctica. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 10.1073/pnas.1119787109.

Chown, S.L., Lee, J.E., Hughes, K.A., Barnes, J., Barrett, P.J., Bergstrom, D.M., Convey, P., Cowan, D.A., Crosbie, K., Dyer, G., Frenot, Y., Grant, S.M., Herr, D., Kennicutt, M.C., Lamers, M., Murray, A., Possingham, H.P., Reid, K., Riddle, M.J., Ryan, P.G., Sanson, L., Shaw, J.D., Sparrow, M.D., Summerhayes, C., Terauds, A. & Wall, D.H. 2012b. Challenges to the future conservation of the Antarctic. *Science*, **337**, 158–159.

Clarke, L.J., Soubrier, J., Weyrich, L.S. & Cooper, A. 2014a. Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, **14**, 1160–1170.

Clarke, L.J., Czechowski, P., Soubrier, J., Stevens, M.I. & Cooper, A. 2014b. Modular tagging of amplicons using a single PCR for high-throughput sequencing. *Molecular Ecology Resources*, **14**, 117–121.

Cline, J., Braman, J.C. & Hogrefe, H.H. 1996. PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research*, **24**, 3546–3551.

Coissac, E., Riaz, T. & Puillandre, N. 2012. Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology*, **21**, 1834–1847.

Colesie, C., Gommeaux, M., Green, T.G.A. & Bueddel, B. 2014. Biological soil crusts in Continental Antarctica: Garwood Valley, southern Victoria Land, and Diamond Hill, Darwin Mountains region. *Antarctic Science*, **26**, 115–123.

Convey, P. 1997. How are the life history strategies of Antarctic terrestrial invertebrates influenced by extreme environmental conditions? *Journal of Thermal Biology*, **22**, 10.1016/S0306-4565(97)00062-4.

Convey, P. 2010. Terrestrial biodiversity in Antarctica – Recent advances and future challenges. *Polar Science*, **4**, 135–147.

Convey, P. & Stevens, M.I. 2007. Antarctic biodiversity. *Science*, **317**, 1877–1878.

Convey, P., Stevens, M.I., Hodgson, D.A., Smellie, J.L., Hillenbrand, C.D., Barnes, D.K.A., Clarke, A., Pugh, P.J.A., Linse, K. & Cary, S.C. 2009. Exploring biological constraints on the glacial history of Antarctica. *Quaternary Science Reviews*, **28**, 3035–3048.

Convey, P., Chown, S.L., Clarke, A., Barnes, D.K.A., Bokhorst, S., Cummings, V., Ducklow, H.W., Frati, F., Green, T.G.A., Gordon, S., Griffiths, H.J., Howard-Williams, C., Huiskes, A.H.L., Laybourn-Parry, J., Lyons, W.B., McMinn, A., Morley, S.A., Peck, L.S., Quesada, A., Robinson, S.A., Schiaparelli, S. & Wall, D.H. 2014. The spatial structure of Antarctic biodiversity. *Ecological Monographs*, **84**, 203–244.

Cowan, D.A., Ramond, J.B., Makhalanyane, T. & de Maayer, P. 2015. Metagenomics of extreme environments. *Current Opinion in Microbiology*, **25**, 10.1016/j.mib.2015.05.005.

Czechowski, P., Clarke, L.J., Breen, J., Cooper, A. & Stevens, M.I. 2016. Antarctic eukaryotic soil diversity of the Prince Charles Mountains revealed by high-throughput sequencing. *Soil Biology & Biochemistry*, **95**, 10.1016/j.soilbio.2015.12.013.

Dalén, L., Götherström, A., Meijer, T. & Shapiro, B. 2007. Recovery of DNA from footprints in the snow. *Canadian Field-Naturalist*, **121**, 321–324.

De Cárcer, D.A., Denman, S.E., McSweeney, C. & Morrison, M. 2011. Strategy for modular tagged high-throughput amplicon sequencing. *Applied and Environmental Microbiology*, **77**, 6310–6312.

Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F. & Taberlet, P. 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, **10**, 10.1098/rsbl.2014.0562.

Delmont, T.O., Simonet, P. & Vogel, T.M. 2013. Mastering methodological pitfalls for surviving the metagenomic jungle. *BioEssays*, **35**, 744–754.

Delmont, T.O., Robe, P., Cecillon, S., Clark, I.M., Constancias, F., Simonet, P., Hirsch, P.R. & Vogel, T.M. 2011. Accessing the soil metagenome for studies of microbial diversity. *Applied and Environmental Microbiology*, **77**, 10.1128/AEM.01526-10.

Denonfoux, J., Parisot, N., Dugat-Bony, E., Biderre-Petit, C., Boucher, D., Morgavi, D.P., Le Paslier, D., Peyretaillade, E. & Peyret, P. 2013. Gene capture coupled to high-throughput sequencing as a strategy for targeted metagenome exploration. *DNA Research*, **20**, 185–196.

Dornelas, M., Gotelli, N.J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C. & Magurran, A.E. 2014. Assemblage time series reveal biodiversity change but not systematic loss. *Science*, **344**, 10.1126/science.1248484.

Dreesens, L.L., Lee, C.K. & Cary, S.C. 2014. The distribution and identity of edaphic fungi in the McMurdo Dry Valleys. *Biology*, **3**, 10.3390/biology3030466.

Drummond, A.J., Newcomb, R.D., Buckley, T.R., Xie, D., Dopheide, A., Potter, B.C.M., Heled, J., Ross, H.A., Tooman, L., Grosser, S., Park, D., Demetras, N.J., Stevens, M.I., Russell, J.C., Anderson, S.H., Carter, A. & Nelson, N. 2015. Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience*, **4**, 10.1186/s13742-015-0086-1.

Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.

Edgar, R.C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, **10**, 996–998.

Ellis, N., Smith, S.J. & Pitcher, C.R. 2012. Gradient forests: calculating importance gradients on physical predictors. *Ecology*, **93**, 156–168.

Epp, L.S., Boessenkool, S., Bellemain, E.P., Haile, J., Esposito, A., Riaz, T., Erseus, C., Gusarov, V.I., Edwards, M.E., Johnsen, A., Stenoien, H.K., Hassel, K., Kauserud, H., Yoccoz, N.G., Brathen, K., Willerslev, E., Taberlet, P., Coissac, E. & Brochmann, C. 2012. New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Molecular Ecology*, **21**, 1821–1833.

Ettema, C.H. & Wardle, D.A. 2002. Spatial soil ecology. *Trends in Ecology & Evolution*, **17**, 177–183.

Faircloth, B.C. & Glenn, T.C. 2012. Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. *PLoS ONE*, **7**, 10.1371/journal.pone.0042543.

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T. & Glenn, T.C. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, **61**, 10.1093/sysbio/sys004.

Fell, J.W., Scorzetti, G., Connell, L. & Craig, S. 2006. Biodiversity of micro-eukaryotes in Antarctic Dry Valley soils with $<5\%$ soil moisture. *Soil Biology & Biochemistry*, **38**, 3107–3119.

Feng, Y.-J., Liu, Q.-F., Chen, M.-Y., Liang, D. & Zhang, P. 2015. Parallel tagged amplicon sequencing of relatively long PCR products using the Illumina HiSeq platform and transcriptome assembly. *Molecular Ecology Resources*, **16**, 10.1111/1755-0998.12429.

Fernández-Mendoza, F., Domaschke, S., García, M.A, Jordan, P., Martín, M.P. & Printzen, C. 2011. Population structure of mycobionts and photobionts of the widespread lichen *Cetraria aculeata*. *Molecular Ecology*, **20**, 1208–1232.

Ficetola, G.F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., Gielly, L., Lopes, C.M., Boyer, F., Pompanon, F., Raye, G. & Taberlet, P. 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*, **15**, 543–556.

Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S., Gilbert, J.A., Wall, D.H. & Caporaso, J.G. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 10.1073/pnas.1215210110.

Frenot, Y., Chown, S.L., Whinam, J., Selkirk, P.M., Convey, P., Skotnicki, M. & Bergstrom, D.M. 2005. Biological invasions in the Antarctic: extent, impacts and implications. *Biological Reviews*, **80**, 45–72.

Gagné, S.A. & Proulx, R. 2009. Accurate delineation of biogeographical regions depends on the use of an appropriate distance measure. *Journal of Biogeography*, **36**, 561–562.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W.J. & Nekrutenko, A. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Research*, **15**, 1451–1455.

Gilbert, J.A, Meyer, F., Antonopoulos, D., Balaji, P., Brown, C.T., Desai, N., Eisen, J.A., Evers, D., Field, D., Feng, W., Huson, D., Jansson, J., Knight, R., Knight, J., Kolker, E., Konstantindis, K., Kostka, J., Kyrpides, N., Mackelprang, R., McHardy, A., Quince, C., Raes, J., Sczyrba, A., Shade, A. & Stevens, R. 2010. Meeting report: the terabase metagenomics workshop and the vision of an Earth microbiome project. *Standards in Genomic Sciences*, **3**, 243–248.

Glenn, T.C. 2011. Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, **11**, 10.1111/j.1755-0998.2011.03024.x.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D.B., Lander, E.S. & Nusbaum, C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, **27**, 182–189.

Gokul, J.K., Valverde, A., Tuffin, M., Cary, S.C. & Cowan, D.A. 2013. Micro-eukaryotic diversity in hypolithons from Miers Valley, Antarctica. *Biology*, **2**, 331–340.

Goordial, J., Davila, A., Greer, C.W., Cannam, R., DiRuggiero, J., McKay, C.P. & Whyte, L.G. 2016. Comparative activity and functional ecology of permafrost soils and lithic niches in a hyper-arid polar desert. *Environmental Microbiology*, 10.1111/1462-2920.13353.

Gutt, J., Zurell, D., Bracegridle, T.J., Cheung, W., Clark, M.S., Convey, P., Danis, B., David, B., De Broyer, C., di Prisco, G., Griffiths, H., Laffont, R., Peck, L.S., Pierrat, B., Riddle, M.J., Saucede, T., Turner, J., Verde, C., Wang, Z.M. & Grimm, V. 2012. Correlative and dynamic species distribution modelling for ecological predictions in the Antarctic: a cross-disciplinary concept. *Polar Research*, **31**, 10.3402/polar.v31i0.11091.

Hajibabaei, M., Spall, J.L., Shokralla, S. & van Konynenburg, S. 2012. Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecology*, **12**, 10.1186/1472-6785-12-28.

Hamming, R.W. 1950. Error detecting and error correcting codes. *Bell System Technical Journal*, **29**, 10.1002/j.1538-7305.1950.tb00463.x.

Headland, R.K. 2009. *A chronology of Antarctic exploration*, 2nd ed. Bernard Quaritch, 722 pp.

Hogg, I.D., Cary, S.C., Convey, P., Newsham, K.K., O'Donnell, A.G., Adams, B.J., Aislabie, J., Frati, F., Stevens, M.I. & Wall, D.H. 2006. Biotic interactions in Antarctic terrestrial ecosystems: are they a factor? *Soil Biology & Biochemistry*, **38**, 3035–3040.

Howard-Williams, C., Peterson, D., Lyons, W.B., Cattaneo-Vietti, R. & Gordon, S. 2006. Measuring ecosystem response in a rapidly changing environment: the Latitudinal Gradient Project. *Antarctic Science*, **18**, 465–471.

Hughes, K.A. & Convey, P. 2010. The protection of Antarctic terrestrial ecosystems from inter- and intra-continental transfer of non-indigenous species by human activities: a review of current systems and practices. *Global Environmental Change - Human and Policy Dimensions*, **20**, 10.1016/j.gloenvcha.2009.09.005.

Hughes, K.A. & Convey, P. 2012. Determining the native/non-native status of newly discovered terrestrial and freshwater species in Antarctica – current knowledge, methodology and management action. *Journal of Environmental Management*, **93**, 52–66.

Hughes, K.A., Convey, P., Maslen, N.R. & Smith, R.I.L. 2010. Accidental transfer of non-native soil organisms into Antarctica on construction vehicles. *Biological Invasions*, **12**, 875–891.

Hui, F.K.C., Taskinen, S., Pledger, S., Foster, S.D. & Warton, D.I. 2015. Model-based approaches to unconstrained ordination. *Methods in Ecology and Evolution*, **6**, 10.1111/2041-210X.12236.

Huson, D.H. & Weber, N. 2013. Microbial community analysis using MEGAN. *Microbial Metagenomics, Metatranscriptomics, and Metaproteomics*, **531**, 10.1016/B978-0-12-407863-5.00021-6.

Jarman, S.N., McInnes, J.C., Faux, C., Polanowski, A.M., Marthick, J., Deagle, B.E., Southwell, C. & Emmerson, L. 2013. Adélie penguin population diet monitoring by analysis of food DNA in scats. *PLoS ONE*, **8**, 10.1371/journal.pone.0082227.

Jungblut, A.D., Vincent, W.F. & Lovejoy, C. 2012. Eukaryotes in Arctic and Antarctic cyanobacterial mats. *FEMS Microbiology Ecology*, **82**, 416–428.

Kanagawa, T. 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of Bioscience and Bioengineering*, **96**, 317–323.

Kennicutt, M.C., Chown, S.L., Cassano, J.J., Liggett, D., Massom, R., Peck, L.S., Rintoul, S.R., Storey, J.W.V., Vaughan, D.G., Wilson, T.J. & Sutherland, W.J. 2014. Polar research: six priorities for Antarctic science. *Nature*, **512**, 10.1038/512023a.

Kennicutt, M.C., Chown, S.L., Cassano, J.J., *et al*. 2015. A roadmap for Antarctic and Southern Ocean science for the next two decades and beyond. *Antarctic Science*, **27**, 10.1017/S0954102014000674.

Khan, N., Tuffin, M., Stafford, W., Cary, C., Lacap, D.C., Pointing, S.B. & Cowan, D. 2011. Hypolithic microbial communities of quartz rocks from Miers Valley, McMurdo Dry Valleys, Antarctica. *Polar Biology*, **34**, 1657–1668.

Kircher, M., Sawyer, S. & Meyer, M. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, **40**, 10.1093/nar/gkr771.

Kwok, S. 1990. Procedures to minimize PCR-product carry-over. *In* Innis, M.A., Gelfand, D.H. & Sninsky, J.J., *eds*. *PCR protocols: a guide to methods and applications*. San Diego, CA: Academic Press, 142–145.

Lan, Y., Wang, Q., Cole, J.R. & Rosen, G.L. 2012. Using the RDP classifier to predict taxonomic novelty and reduce the search space for finding novel organisms. *PLoS ONE*, **7**, 10.1371/journal.pone.0032491.

Lawley, B., Ripley, S., Bridge, P. & Convey, P. 2004. Molecular analysis of geographic patterns of eukaryotic diversity in Antarctic soils. *Applied and Environmental Microbiology*, **70**, 5963–5972.

Lee, C.K., Barbier, B.A., Bottos, E.M., McDonald, I.R. & Cary, S.C. 2012a. The Inter-Valley Soil Comparative Survey: the ecology of Dry Valley edaphic microbial communities. *ISME Journal*, **6**, 10.1038/ismej.2011.170.

Lee, C.K., Herbold, C.W., Polson, S.W., Wommack, K.E., Williamson, S.J., McDonald, I.R. & Cary, S.C. 2012b. Groundtruthing next-gen sequencing for microbial ecology – biases and errors in community structure estimates from PCR amplicon pyrosequencing. *PLoS ONE*, **7**, 10.1371/journal.pone.0044224.

Legendre, P. & Andersson, M.J. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, **69**, 10.1890/0012-9615(1999)069[0001:DBRATM]2.0.CO;2.

Lemmon, A.R., Emme, S.A. & Lemmon, E.M. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, **61**, 10.1093/sysbio/sys049.

Lenz, T.L. & Becker, S. 2008. Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci – implications for evolutionary analysis. *Gene*, **427**, 117–123.

Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T. & Machida, R.J. 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, **10**, 10.1186/1742-9994-10-34.

Lindgreen, S. 2012. AdapterRemoval: easy cleaning of next generation sequencing reads. *BMC Research Notes*, **5**, 10.1186/1756-0500-5-337.

Lindgreen, S., Adair, K.L. & Gardner, P.P. 2016. An evaluation of the accuracy and speed of metagenome analysis tools. *Scientific Reports*, **6**, 19233, 10.1038/srep19233.

Liu, S., Li, Y.Y., Lu, J.L., Su, X., Tang, M., Zhang, R., Zhou, L.L., Zhou, C.R., Yang, Q., Ji, Y.Q., Yu, D.W. & Zhou, X. 2013. SOAPBarcode: revealing arthropod biodiversity through assembly of Illumina shotgun sequences of PCR amplicons. *Methods in Ecology and Evolution*, **4**, 1142–1150.

Lohse, M., Bolger, A.M., Nagel, A., Fernie, A.R., Lunn, J.E., Stitt, M. & Usadel, B. 2012. RobiNA: A user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Research*, **40**, W622–W627.

López-Bueno, A., Tamames, J., Velázquez, D., Moya, A., Quesada, A. & Alcamí, A. 2009. High diversity of the viral community from an Antarctic lake. *Science*, **326**, 858–861.

Machida, R.J. & Knowlton, N. 2012. PCR primers for metazoan nuclear 18S and 28S ribosomal DNA sequences. *PLoS ONE*, **7**, 10.1371/journal.pone.0046180.

Magalhaes, C., Stevens, M.I., Cary, S.C., Ball, B.A., Storey, B.C., Wall, D.H., Tuerk, R. & Ruprecht, U. 2012. At limits of life: multidisciplinary insights reveal environmental constraints on biotic diversity in Continental Antarctica. *PLoS ONE*, **7**, 10.1371/journal.pone.0044578.

Makhalanyane, T.P., Valverde, A., Birkeland, N.K., Cary, S.C., Tuffin, I.M. & Cowan, D.A. 2013. Evidence for successional development in Antarctic hypolithic bacterial communities. *ISME Journal*, **7**, 2080–2090.

Marchant, D.R. & Head, J.W. 2007. Antarctic dry valleys: microclimate zonation, variable geomorphic processes, and implications for assessing climate change on Mars. *Icarus*, **192**, 187–222.

McGaughran, A., Stevens, M.I., Hogg, I.D. & Carapelli, A. 2011. Extreme glacial legacies: a synthesis of the Antarctic springtail phylogeographic record. *Insects*, **2**, 62–82.

McGeoch, M.A., Shaw, J.D., Terauds, A., Lee, J.E. & Chown, S.L. 2015. Monitoring biological invasion across the broader Antarctic: a baseline and indicator framework. *Global Environmental Change - Human and Policy Dimensions*, **32**, 108–125.

McMurdie, P.J. & Holmes, S. 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, **8**, 10.1371/journal.pone.0061217.

McMurdie, P.J. & Holmes, S. 2014. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Computational Biology*, **10**, 10.1371/journal.pcbi.1003531.

Meyer, M., Stenzel, U. & Hofreiter, M. 2008. Parallel tagged sequencing on the 454 platform. *Nature Protocols*, **3**, 267–278.

Nakai, R., Abe, T., Baba, T., Imura, S., Kagoshima, H., Kanda, H., Kohara, Y., Koi, A., Niki, H., Yanagihara, K. & Naganuma, T. 2012. Eukaryotic phylotypes in aquatic moss pillars inhabiting a freshwater lake in East Antarctica, based on 18S rRNA gene analysis. *Polar Biology*, **35**, 1495–1504.

Niederberger, T.D., Sohm, J.A., Gunderson, T.E., Parker, A.E., Tirindelli, J., Capone, D.G., Carpenter, E.J. & Cary, S.C. 2015. Microbial community composition of transiently wetted Antarctic Dry Valley soils. *Frontiers in Microbiology*, **6**, 10.3389/fmicb.2015.00009.

Nielsen, U.N. & Wall, D.H. 2013. The future of soil invertebrate communities in polar regions: different climate change responses in the Arctic and Antarctic? *Ecology Letters*, **16**, 409–419.

O'Neill, E.M., Schwartz, R., Bullock, C.T., Williams, J.S., Shaffer, H.B., Aguilar-Miguel, X., Parra-Olea, G. & Weisrock, D.W. 2013. Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. *Molecular Ecology*, **22**, 111–129.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. 2015. . *Vegan: community ecology package*. Available at: https://cran.r-project.org/package=vegan.

Orgiazzi, A., Bianciotto, V., Bonfante, P., Daghino, S., Ghignone, S., Lazzari, A., Lumini, E., Mello, A., Napoli, C., Perotto, S., Vizzini, A., Bagella, A., Murat, C. & Girlanda, M. 2013. 454 pyrosequencing analysis of fungal assemblages from geographically distant, disparate soils reveals spatial patterning and a core mycobiome. *Diversity*, **5**, 73–98.

Pedersen, M.W., Overballe-Petersen, S., Ermini, L., Sarkissian, C.D., Haile, J., Hellstrom, M., Spens, J., Thomsen, P.F., Bohmann, K., Cappellini, E., Schnell, I.B., Wales, N.A., Caroe, C., Campos, P.F., Schmidt, A.M.Z., Gilbert, M.T.P., Hansen, A.J., Orlando, L. & Willerslev, E. 2014. Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society - Biological Sciences*, **B370**, 10.1098/rstb.2013.0383.

Pointing, S.B. & Belnap, J. 2012. Microbial colonization and controls in dryland systems. *Nature Reviews Microbiology*, **10**, 551–562.

Powell, S.M., Bowman, J.P., Snape, I. & Stark, J.S. 2003. Microbial community variation in pristine and polluted nearshore Antarctic sediments. *FEMS Microbiology Ecology*, **45**, 10.1016/S0168-6496(03)00135-1.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W.G., Peplies, J. & Gloeckner, F.O. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, **35**, 7188–7196.

R Development Team. 2016. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at: http://www.r-project.org/.

Ratnasingham, S. & Hebert, P.D.N. 2007. BOLD: the barcode of life data system (www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P. & Coissac, E. 2011. ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, **39**, 10.1093/nar/gkr732.

Robertson, C.E., Harris, J.K., Wagner, B.D., Granger, D., Browne, K., Tatem, B., Feazel, L.M., Park, K., Pace, N.R. & Frank, D.N. 2013. Explicet: graphical user interface software for metadata-driven management, analysis and visualization of microbiome data. *Bioinformatics*, **29**, 3100–3101.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K. 2009. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Roesch, L.F.W., Fulthorpe, R.R., Pereira, A.B., Pereira, C.K., Lemos, L.N., Barbosa, A.D., Suleiman, A.K.A., Gerber, A.L., Pereira, M.G., Loss, A. & da Costa, E.M. 2012. Soil bacterial community abundance and diversity in ice-free areas of Keller Peninsula, Antarctica. *Applied Soil Ecology*, **61**, 7–15.

Rogers, A.D. 2007. Evolution and biodiversity of Antarctic organisms: a molecular perspective. *Philosophical transactions of the Royal Society - Biological sciences*, **B362**, 2191–2214.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. & Erlich, H.A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase. *Science*, **239**, 487–491.

Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J. & Walker, A.W. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, **12**, 10.1186/s12915-014-0087-z.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A, Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., van Horn, D.J. & Weber, C.F. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, **75**, 7537–7541.

Shaw, J.D., Terauds, A., Riddle, M.J., Possingham, H.P. & Chown, S.L. 2014. Antarctica's protected areas are inadequate, unrepresentative, and at risk. *PLoS Biology*, **12**, 10.1371/journal.pbio.1001888.

Shokralla, S., Singer, G.A.C. & Hajibabaei, M. 2010. Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol. *BioTechniques*, **48**, 10.2144/000113362.

Sipos, R., Székely, A.J., Palatinszky, M., Révész, S., Márialigeti, K. & Nikolausz, M. 2007. Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targeting bacterial community analysis. *FEMS Microbiology Ecology*, **60**, 10.1111/j.1574-6941.2007.00283.x.

Smith, D.P. & Peay, K.G. 2014. Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. *PLoS ONE*, **9**, 10.1371/journal.pone.0090234.

Stiller, M., Knapp, M., Stenzel, U., Hofreiter, M. & Meyer, M. 2009. Direct multiplex sequencing (DMPS) – a novel method for targeted high-throughput sequencing of ancient and highly degraded DNA. *Genome Research*, **19**, 1843–1848.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. 2012a. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, **21**, 2045–2050.

Taberlet, P., Prud'Homme, S.M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., Choler, P., Clement, J.C., Melodelima, C., Pompanon, F. & Coissac, E. 2012b. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Molecular Ecology*, **21**, 1816–1820.

Tang, C.Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T.G. & Fontaneto, D. 2012. The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 16 208–16 212.

Terauds, A., Chown, S.L., Morgan, F., Peat, H.J., Watts, D.J., Keys, H., Convey, P. & Bergstrom, D.M. 2012. Conservation biogeography of the Antarctic. *Diversity and Distributions*, **18**, 726–741.

Turrill, W.B. 1938. The expansion of taxonomy with special reference to spermatophyta. *Biological Reviews*, **13**, 10.1111/j.1469-185X.1938.tb00522.x.

Van Dijk, E.L., Auger, H., Jaszczyszyn, Y. & Thermes, C. 2014. Ten years of next-generation sequencing technology. *Trends in Genetics*, **30**, 10.1016/j.tig.2014.07.001.

Van Rossum, G. & Drake, F.L. 1995. *Python tutorial*. Amsterdam: Centrum voor Wiskunde en Informatica.

Velasco-Castrillón, A. & Stevens, M.I. 2014. Morphological and molecular diversity at a regional scale: a step closer to understanding Antarctic nematode biogeography. *Soil Biology & Biochemistry*, **70**, 272–284.

Velasco-Castrillón, A., Gibson, J.A.E. & Stevens, M.I. 2014a. A review of current Antarctic limno-terrestrial microfauna. *Polar Biology*, **37**, 1517–1531.

Velasco-Castrillón, A., Page, T.J., Gibson, J.A.E. & Stevens, M.I. 2014b. Surprisingly high levels of biodiversity and endemism amongst Antarctic rotifers uncovered with mitochondrial DNA. *Biodiversity*, **15**, 130–142.

Velasco-Castrillón, A., Schultz, M.B., Colombo, F., Gibson, J.A.E., Davies, K.A, Austin, A.D. & Stevens, M.I. 2014c. Distribution and diversity of soil microfauna from East Antarctica: assessing the link between biotic and abiotic factors. *PLoS ONE*, **9**, 10.1371/journal.pone.0087529.

Wang, Y., Naumann, U., Wright, S.T. & Warton, D.I. 2012. mvabund – an R package for model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution*, **3**, 10.1111/j.2041-210X.2012.00190.x.

Whittaker, R.H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, **30**, 10.2307/1943563.

Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K.P., Paczian, T., Trimble, W.L., Bagchi, S., Gram, A., Chaterji, S. & Meyer, F. 2016. The MG-RAST metagenomics database and portal in 2015. *Nucleic Acids Research*, **44**, 10.1093/nar/gkv1322.

Willerslev, E., Hansen, A.J. & Poinar, H.N. 2004. Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology & Evolution*, **19**, 10.1016/j.tree.2003.11.010.

Wish, M. & Carroll, J.D. 1982. Multidimensional scaling and its applications. *In* Krishnaiah, P.R. & Kanal, L.N., *eds. Handbook of statistics 2*. North-Holland: Elsevier, 317–345.

Wu, T.H., Ayres, E., Bardgett, R.D., Wall, D.H. & Garey, J.R. 2011. Molecular study of worldwide distribution and diversity of soil animals. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 17 720–17 725.

Zhan, A.B., Bailey, S.A., Heath, D.D. & Macisaac, H.J. 2014. Performance comparison of genetic markers for high-throughput sequencing-based biodiversity assessment in complex communities. *Molecular Ecology Resources*, **14**, 10.1111/1755-0998.12254.

Zhou, X.F. & Rokas, A. 2014. Prevention, diagnosis and treatment of high-throughput sequencing data pathologies. *Molecular Ecology*, **23**, 10.1111/mec.12680.