A lab in the field: applications of real-time, in situ metagenomic sequencing

Adriel Latorre-Pérez¹, Javier Pascual¹, Manuel Porcar¹,² and Cristina Vilanova¹,*

¹Darwin Bioprospecting Excellence SL, Valencia, Spain and ²Institute for Integrative Systems Biology, I2SysBio, University of Valencia-CSIC, Valencia, Spain

*Correspondence address. Darwin Bioprospecting Excellence SL, Valencia, Spain; E-mail: cristina@darwinbioprospecting.com

Abstract
High-throughput metagenomic sequencing is considered one of the main technologies fostering the development of microbial ecology. Widely used second-generation sequencers have enabled the analysis of extremely diverse microbial communities, the discovery of novel gene functions, and the comprehension of the metabolic interconnections established among microbial consortia. However, the high cost of the sequencers and the complexity of library preparation and sequencing protocols still hamper the application of metagenomic sequencing in a vast range of real-life applications. In this context, the emergence of portable, third-generation sequencers is becoming a popular alternative for the rapid analysis of microbial communities in particular scenarios, due to their low cost, simplicity of operation, and rapid yield of results. This review discusses the main applications of real-time, in situ metagenomic sequencing developed to date, highlighting the relevance of this technology in current challenges (such as the management of global pathogen outbreaks) and in the next future of industry and clinical diagnosis.

Keywords: third-generation sequencing; in situ metagenomics; microbial ecology

Introduction
For many years, culture-dependent approaches were the only tools available for the study of microorganisms, although the vast majority of microbial species (>99%) cannot be cultivated [1]. This limitation lasted until the development of molecular techniques, such as the automation of Sanger sequencing [2], molecular markers [3], cloning [4], or fluorescence in situ hybridization [5], among many others. However, these molecular techniques presented other weaknesses, like the inability to access low-abundance microorganisms, generating a bias towards the most abundant taxa.

The term metagenomics was proposed in the 1990s [6] to define the set of genomes that could be found in a given environment. The fundamental aim of metagenomics is the study of microorganisms in the context of their community by means of sequencing genomic fragments from the entire microbiome simultaneously. Nevertheless, this goal can be partially accomplished by sequencing marker genes, even though this approach should not be considered as true metagenomics [7]. In marker-gene studies, generic, relatively universal primers are used to amplify a fragment of a given gene through polymerase chain reaction (PCR) (e.g. 16S rRNA for bacteria/archaea, 18S/ITS for fungi) from all genomes present in a given sample, and the resulting pool of amplicons is sequenced. Next, the sequences are clustered into operational taxonomic units (OTUs), each OTU is taxonomically identified, and compared across samples. Traditionally, OTUs were constructed by grouping sequences according to a defined similarity threshold (typically 97%).
However, OTUs are being replaced by amplicon sequence variants, which group sequences that are completely identical [6]. While fast and inexpensive, this method does not give any information on the hundreds of thousands of functional genes encoded by other parts of the (meta)genomes as these remain unsequenced. Whole-genome sequencing (WGS)—or shotgun—metagenomics can offer an alternative and complementary method since it is based on the application of sequencing techniques to the entirety of the genomic material in the microbiome of an environmental sample. Sequencing the genomes of all microorganisms can provide information about the diversity of functional genes, and allow the assignment of each metabolic function to specific taxa, to identify novel genes or proteins so far unknown, and to assemble genomes in order to study evolutionary relationships.

The number of metagenomic studies has dramatically increased in the last years, mainly due to the emergence of high-throughput sequencing technologies and the development of bioinformatic tools that facilitate the assembly of data and the assignment of sequences through a process called binning [9]. The binning process consists of grouping assembled sequences (contigs) into discrete units (bins), which ideally represent draft genomes of individual microorganisms [10]. Overall, both high-throughput sequencing and bioinformatics have proven powerful tools that have generated, at a relatively low cost, a huge amount of genetic information [11].

High-throughput sequencing technologies can be divided into second- and third-generation ones. Two of the most widely used second-generation sequencing (SGS) technologies are Illumina and Ion Torrent. Although both techniques are based on sequence-by-synthesis, they have methodological differences. In Illumina sequencers, short DNA fragments are attached to a glass slide or micro-well and amplified to produce clusters. Fluorescence-labelled nucleotides are then washed attached to a glass slide or micro-well and amplified to produce differences. In Illumina sequencers, short DNA fragments are based on sequence-by-synthesis, they have methodological widely used second-generation sequencing (SGS) technologies. Third-generation sequencing (TGS), also known as long-read sequencing, is based on single-molecule sequencing, which speeds up the sequencing process. This technology is currently under active development and includes platforms such as Pacific Biosciences (PacBio) or Oxford Nanopore Technologies (ONT). PacBio is based on single-molecule, real-time sequencing technology. It can be directly plugged into a standard USB3 port on a computer with a simple configuration. Specifically, a computer with a solid-state drive, >8 GB of RAM, and >128 GB of hard disk space can be used for sequencing. The sequencer periodically outputs a group of reads in the form of raw current signals, which are then base-called on a laptop or on an ultra-portable ONT’s MinIT. Furthermore, sequence analyses (such as sequence alignment and genome polishing) can be performed on a mobile phone [24]. Therefore, the ultra-portability, affordability, and speed in data production make the MinION technology suitable for real-time sequencing in a variety of environments, such as Ebola surveillance in West Africa during the last outbreak [25], microbial communities inspection in the Arctic [26], DNA sequencing on the International Space Station (ISS) [27], and even the recently emerging pandemic coronavirus SARS-CoV-2 [28, 29]. This review describes a range of applications in which having portable, low-cost, fast, and robust technologies allowing an in situ analysis of samples is key to address important challenges.

**Portable sequencing in natural environments**

Exploring the microbial diversity of natural environments via DNA sequencing techniques has become a routine in the last decade. Large-scale studies like the Earth Microbiome Project have led to the massive characterization of microbial populations inhabiting different environments on our planet [30]. Moreover, metagenomic sequencing has proved to be very useful for a wide range of applications such as recovering new genomes from unculturable organisms, mining microbial enzymes with potential applications in the industry, or
discovering new biosynthetic gene clusters [31–33]. These studies have typically relied on next-generation sequencing platforms like Illumina, which usually requires shipping samples to a centralized sequencing facility. Nevertheless, biodiversity assessment studies are usually carried out in remote locations with limited access to DNA sequencing services, forcing scientists to design-intensive sampling expeditions and returning to their home institutions to perform the sequencing and the data analysis.

ONT sequencers have emerged as an alternative to these traditional approaches, allowing the creation of mobile, in-field laboratories. Figure 1 depicts a general workflow for the metagenomic analysis of samples using adapted protocols and a MinION device. Pomerantz et al. [34] and Menegon et al. [35] designed portable laboratories that included thermocyclers and centrifuges powered by external batteries, and a MinION device connected to a laptop to perform in situ DNA sequencing. Both works were not focused on metagenomic applications, but on evaluating the taxonomic identity of different animal specimens (reptiles and amphibians) via targeted sequencing of the 16S rRNA gene or other mitochondrial genes. However, the applied methodologies and lab configurations could be easily adapted to perform metataxonomic approaches relying on the amplification and massive sequencing of marker genes.

The feasibility of MinION-based metagenomic sequencing protocols has been specially tested in extremely cold environments. Edwards et al. [36] reported for the first time the use of mobile laboratories for the in situ characterization of the microbiota of a High Arctic glacier. They were able to adapt the widely used PowerSoil DNA Isolation kit (MoBio, Inc.) for its in-field use, and to perform the data analysis either online and offline. The report included new results from in situ metagenomes and 16S rRNA sequencing of different glaciers samples, and a benchmarking of the performance of in-field sequencing protocols by using mock communities as well as real samples. In the latter case, they compared the resulting taxonomic profiles with the microbial composition assessed by SGS platforms, describing strongly positive Pearson correlations at the phylum level. Goordial et al. [26] were also able to perform in situ MinION sequencing in the McGill Arctic Research Station. In this case, a permafrost sample was analysed using two different library preparation kits on the same extracted DNA. A similar percentage of Bacteria and Archaea was detected using both kits, but differences in the relative abundance of viruses and eukaryotic organisms were noted. The taxonomic profile of the same permafrost sample was also obtained by means of 16S rRNA illumina sequencing. Notably, similar taxonomic groups were identified in all the cases at the phylum level, although relative abundances varied among the different methodologies. In a parallel work, Johnson et al. [37] used portable field techniques to isolate DNA from desiccated microbial mats collected in the Antarctic Dry Valleys, construct metagenomic libraries, and sequence the samples outdoors (Taylor Valley; Temperature = −1°C) and in the McMurdo Station (Room Temperature, RT). Longer reads were achieved by sequencing at RT, but average and median read length did not depend on ambient temperature. The study also reported that cold temperatures (4°C) reduced the quality of the generated sequences, even when working with high-quality DNA (Lambda Phage).

Finally, Gowers et al. [38] designed and transported a miniaturized lab across Europe’s largest ice cap (Vatnajökull, Iceland) by ski and sledge. They adapted DNA extraction and sequencing protocols to be performed in a tent during the expedition, using solar energy and external batteries to power the hardware. Offline basecalling was achieved in situ by using Guppy (Oxford Nanopore, Oxford, UK), but the metagenomic data analysis could not be carried out due to code errors while running the local version of Kaiju [39].

In addition to cold environments, ONT sequencers have also been applied for sequencing a biofilm sample at a depth of 100 m within a Welsh coal mine [40]. This work presented the ‘MetageNomad’, a suite of off-the-shelf tools for metagenomic sequencing in remote areas using battery-powered equipment. The authors were able to perform the data analysis in situ by using Centrifuge [41] and a local database for characterizing the microbial composition of the sample.

Interestingly, MinION devices have allowed DNA sequencing off the Earth. A first study from Castro-Wallace et al. [27] compared the performance of nanopore sequencing in the ISS with experiments carried out on Ground Control, obtaining similar results. As a proof-of-concept, the authors used equimolar mixtures of genomic DNA from lambda bacteriophage, Escherichia coli (strain K12, MG1655) and Mus musculus (female BALB/c mouse) for the metagenomic sequencing. Data analysis could not be carried out at the ISS because of the lack of a laptop with the necessary tools installed, but it was demonstrated on the ground that sequencing analysis and microbial identification are completely feasible aboard the ISS. Recently, Burton et al. [42] have reported that the preparation and sequencing of 16S rRNA libraries are also achievable at the ISS. Specifically, the ZymoBIOMICS Microbial Community DNA Standard (Zymo Research) was used as the input DNA. Again, the results were comparable to the microbial profiles obtained on Earth. Remarkably, Carr et al. [43] determined that ONT sequencers performed consistently in reduced gravity environments, which would allow the use of nanopore sequencing in space expeditions to Mars or icy moons.

Although the viability of nanopore sequencing has been widely demonstrated even under extremely harsh conditions, the vast majority of the studies resulted in reduced yield compared to current MinION’s metagenomic output (Table 1), which could reach up to 27 Gbp using a single flowcell [48]. This highlights the need to optimize in-field protocols in order to maximize the use of sequencing resources and reduce the price per sample, which is a key factor in some applications. Recently, a work from Urban et al. [44] studied the microbial communities present in the surface water of Cam River (Cambridge). All the protocols were carried out in the lab, and the authors were able to achieve up to ~5.5 M 16S rRNA full-length sequences with exclusive barcode assignments in a single MinION run. Other groups have used MinION devices for characterizing river water [45], seawater [46], and marine sediments [47] through
metagenomic sequencing. Even though these experiments were not implemented in the field, they demonstrated the possibility of obtaining higher sequencing yields (Table 1). The described outputs are compatible with more ambitious metagenomic analyses, such as the de novo recovery of single genomes directly from complex environmental samples. For that reason, the adaptation of sequencing protocols to field conditions is still to be further optimized.

Supporting microbiome-driven industrial processes

Microbiology has been present in the industry for centuries. In fact, human beings already used microorganisms for their own benefits long before they even knew that microscopic life existed. Nowadays, most of the microbiome-driven industrial processes are still not completely understood. Metagenomic sequencing has been widely applied in order to shed light on the microbial and metabolic transitions occurring on these industrial transformations. Some examples include the investigation of the link between microorganisms and their key roles or prevalence in microbial-based food products [49, 50]; the interaction of plants and root-associated bacteria for enhancing plant mineral nutrition [51]; or the description of the adverse effects of industrial subproducts used as soil fertilizers [52].

ONT portable sequencers are not only a valuable tool for characterizing industrial microbiomes, but for detecting and monitoring crucial microorganisms in real time (Fig. 2). Hardegen et al. [53] used full-length 16S rRNA sequencing for analysing changes in the archaean community present in anaerobic digesters operating under different conditions. Higher proportions of Methanosarcina spp. were detected in the reactors achieving elevated biogas production. Although the sequencing was not carried out in situ, the suitability of MinION for monitoring and evaluating an industrial process through a microbial marker was demonstrated. Bacteriomes involved in the biogas production have been also studied through nanopore sequencing [54, 55], producing results which could be coupled with the Lotka–Volterra model for analysing the microbial interactions occurring in the reactor [56].

Water quality and wastewater management is another area of great interest for microbial monitoring. In fact, it has been proposed that sewage could serve for tracking infectious agents excreted in urine or faeces, such as SARS-CoV-2 [57]. In this particular context, the in situ and real-time assessment of pathogenic microorganisms by means of MinION sequencing would be especially advantageous. Hu et al. [58] reported correlations between E. coli culturing counts and the proportion of nanopore reads mapping a comprehensive human gut microbiota gene dataset, highlighting the potential use of this molecular technique as an indicator of faecal contamination. ONT metagenomic sequencing results were similar to those obtained with Illumina 16S rRNA sequencing, but a reduced time was achieved using MinION. Nanopore sequencing could be also employed for evaluating antibiotic resistance genes (ARGs) and antimicrobial-resistant pathogens present in wastewater treatment plants [59]. In this case, both Illumina and nanopore shotgun sequencing revealed comparable abundances of major ARG types. The agreement between the two platforms has been also described for the analysis of different water sources in Nepal through 16S rRNA sequencing [60]. Although long-reads allowed the classification of 59.41% of the reads down to the species level—no Illumina reads were classified at this level—a significant number of false-positives arose. These results were consistent with observations from [61], which showed that the bacterial identification at the genus level was reliable. Species-level misclassifications could be partially addressed by employing different—and optimized—bioinformatic approaches for the taxonomic classification [65, 62], by sequencing the complete 16S-ITS-23S region of the ribosomal operon [63, 64], or by coupling MinION sequencing with complementary quantitative PCR assays [60].

Agro-food industry would also benefit from real-time sequencing. For instance, nanopore metagenomic sequencing could be useful for the quick detection of plant pathogens infecting crops. Hu et al. [65] were able to identify the fungal species causing diseases on wheat plants, which were previously infected with known microbes. Co-occurrences between fungal and bacterial genera were also detected. Viral infectious diseases could be in situ monitored by using this technology, allowing rapid and improved response to outbreaks [66]. Other successful applications of ONT in the food industry included the characterization of the microbiome of a salmon ectoparasite (Caligus rogercresseyi), revealing its potential role as a reservoir for fish pathogens [67]; and the determination of the fish species present in complex mixtures, which would help to prevent—and rapidly detect—food fraud [68].

Overall, nanopore results generally agreed with those obtained by Illumina sequencing when available, thus validating the use of this technology for the vast majority of applications. Despite the huge potential shown, the suitability of MinION sequencing in an industrial context has yet to be ascertained, since all the discussed works were not carried out under field conditions. In fact, there are some critical points to be addressed before this technique could become a standard in the industry: (i) sequencing cost should be reduced; (ii) rapid and reliable in situ DNA extraction and library preparation protocols should be designed and validated; (iii) minimal sequencing yields should be determined for each specific application; (iv) fast and real-time pipelines should be created and tested; and (v) level of expertise for managing the data and the samples should be notably reduced.

Real-time analysis of clinical samples

Microbial infections are an increasingly relevant problem in intensive care units worldwide. Especially, the emergence of multi-drug resistance microorganisms is one of the main threats our society is facing from a clinical point of view [69]. Current diagnostics for pathogen identification in hospitals is still mainly dependent on culture- and molecular-based approaches, which have several limitations regarding specificity, bias, sensitivity, and time to diagnosis. The revolution of high-throughput sequencing and the decreasing costs associated to SGS has strongly empowered clinical diagnostics and other aspects of medical care [70]. In the particular case of clinical infections, high-throughput metagenomic sequencing allowed for the first time the precise strain-level identification of multiple pathogenic agents in single, all-inclusive diagnostic tests [71]. However, the limitations of SGS regarding cost and time to results (as described in previous sections) hamper its application when a fast analysis is needed. For instance, in the case of sepsis, patients are usually treated with broad-spectrum antibiotics until the first results of culture-based analysis (including determination of antibiotic susceptibility) are obtained 36–48 h later. When available, SGS approaches can speed up the process to ~24 h, but result is expensive, labour intensive, and
| References          | Samples                                      | Library type | Lab type         | Equipment                                                                                   | Yield (no of reads) |
|---------------------|----------------------------------------------|--------------|------------------|--------------------------------------------------------------------------------------------|--------------------|
| Edwards et al. [36] | Cryoconite and mock communities              | Metagenomic  | In-field         | Vortex, Microcentrifuge, Fluorometer, PCR cycler (optional), Laptop, and Miscellaneous power sources | 3514–52 000        |
| Edwards et al. [36] | Cryoconite and mock communities              | 16S rRNA     | In-field         | Vortex, Microcentrifuge, Fluorometer, MiniPCR, Laptop, and Miscellaneous power sources      | 20,000–220 051     |
| Menegon et al. [35] | Animal tissue                                | 16S rRNA and other marker genes | In-field         | Microcentrifuge, Fluorometer, Thermocycler, Laptop, Portable Refrigerator, and 12 V portable batteries | 5039               |
| Pomerantz et al. [34] | Animal tissue                                | 16S rRNA and other marker genes | In-field         | Benchtop centrifuge, MiniPCR, Laptop, and External batteries                               | 16 663             |
| Goordial et al. [26] | Permafrost                                    | Metagenomic  | Research station | Vortex, Microcentrifuge, Fluorometer, Magnetic rack, and Computer                           | 6348–9530          |
| Johnson et al. [37] | Microbial mats                                | Metagenomic  | In-field         | Laptop and Insulating materials                                                            | 573–6026           |
| Gowers et al. [38]  | Soil                                          | Metagenomic  | In-field         | USB Vortex, Hand-powered Centrifuge, Fluorometer, Laptop, Solar panels, and External batteries | 19 839–133 538     |
| Edwards et al. [40] | Sediment                                      | Metagenomic  | In-field         | TerraLyser (Zymo Research, Inc.), 12 V Microcentrifuge, Fluorometer, MiniPCR, two Laptops, and External batteries | 1184               |
| Castro-Wallace et al. [27] | Mock DNA                                     | Metagenomic  | ISS              | Microcentrifuge, Fluorometer, MiniPCR, and Computer                                        | 14 903–60 864 (libraries prepared in a regular lab on Earth) |
| Burton et al. [42]  | Mock DNA and pure cultures                    | 16S rRNA     | ISS              | MiniPCR, Computer, Refrigerator, and Freezer                                               | >15 000            |
| Urban et al. [44]   | River water                                   | 16S rRNA     | Regular lab      | Fully equipped                                                                              | 737 164-5 491 510  |
| Hammer et al. [45]  | River water                                   | Metagenomic  | Regular lab      | Fully equipped                                                                              | 397 884-1 261 165  |
| Liem et al. [46]    | Seawater                                      | Metagenomic  | Regular lab      | Fully equipped                                                                              | 225 200-1 316 823  |
| Cáceres et al. [47] | Marine sediments                              | Metagenomic  | Regular lab      | Fully equipped                                                                              | 1 500 000          |
Metagenomic applications are often limited by the nature of the samples to be analysed. For instance, the characterization of prokaryotes or viruses present in a sample dominated by host cDNA conversion and random amplification prior to sequencing. Despite the notable error rate observed in the sequences, all viruses (chikungunya virus, Ebola virus, and hepatitis C virus) were correctly identified and most of their genomes were recovered with high accuracy (97–99%). A similar approach was reported for the rapid identification of mosquito-borne arbovirus [79], and other viruses causing co-infections, including dengue, from human serum samples [80].

On the other hand, an extensive number of reports have been focused on the analysis of infections caused by bacterial communities (Table 2), using different approaches which resulted in different analysis times. Even though a range of PCR-free protocols have been developed for MinION sequencing, one of the main problems associated to the analysis of microbial communities in clinical samples is the overwhelming concentration of host DNA, which hampers the detection of bacterial sequences during the first hours of the sequencing runs [89, 90]. Several strategies have been applied to partially overcome this limitation. On the one hand, PCR-based approaches targeting the 16S rRNA gene proved the most rapid methods to identify pathogenic agents from human samples. Particular examples of this are the metagenomic analysis in empyema patients with pleural effusion [83] and the metagenomic analysis of patients with acute respiratory distress syndrome [84], both studies reporting the obtention of the first results in only 2 h after the collection of samples. On the other hand, the use of human cell-free samples allows the application of WGS protocols for the analysis of the communities, yielding not only taxonomic information but also the identification of putative antimicrobial resistance genes, which are of outstanding relevance for the selection of effective treatments. Pendleton et al. analysed in 2017 [86] lavage fluids from patients with pneumonia and managed to identify the bacterial pathogens in the lungs in <9h using a WGS strategy. Similar approaches performed on urine samples [87] and resected valves from patients with endocarditis [85] yielded a diagnosis in 4 h. For the analysis of bacterial sepsis, recent reports describe the application of MinION metagenomic sequencing on cell-free samples (<6h from samples to results) [81] and on faecal samples from preterm infants (obtaining results in <5h) [82]. The depletion of human DNA prior to metagenomic sequencing proved also a useful alternative to reduce total analysis time [88].

In the current SARS-CoV-2 outbreak, MinION sequencing is proposed not only as a rapid tool for WGS, but also as a metagenomics-based approach for the rapid diagnosis of poly-microbial/viral infections associated to coronavirus disease COVID19. This is especially relevant to optimize the treatment of patients suffering severe symptoms of the disease.

Finally, other advantages of MinION sequencing besides the reduction of analysis are also to be highlighted. Given the low price of the devices and consumables (in comparison to SGS equipment), MinION has enabled the metagenomic analysis of clinical samples in areas with limited resources [25, 91]. Also, from a technical point of view, the generation of long reads increases the resolution of the taxonomic analysis of the samples, reaching in most cases a species-level identification of the most abundant members of the communities [92, 93].

The ‘read until’ strategy: towards cost-effective in situ metagenomics

Metagenomic applications are often limited by the nature of the samples to be analysed. For instance, the characterization of prokaryotes or viruses present in a sample dominated by host
be 450 bp/s [98–100], and it is relatively common to achieve another DNA fragment. ONT sequencing speed is estimated to sequences longer than 100 kbp [24, 101]. Theoretically, to dis-

fective enrich multiple loci of interest within a bacterial ge-

tome, enabling up to 5-fold coverage improvement. In the field

of metagenomics, BOSS-RUNS could be applied for improving
the characterization of samples by ensuring the deep sequenc-
ing of clade-specific genetic markers [104]. On the other hand,
Kovaka et al. [99] recently developed UNCALLED, a tool able
to directly map ONT raw signals in order to detect wanted/
unwanted sequences. They used this approach for sequencing a mock community (ZymoBIOMICS high molecular weight) con-
taining seven bacteria and one yeast. The objective was to map
the generated signals to a database containing the references
for the bacterial genomes (29 Mbp), rejecting DNA strands when
a match was detected. Bacterial sequencing depletion resulted

DNA via direct shotgun sequencing could be really challenging,
and would require high sequencing depth, thus increasing the
cost of the analysis [94, 95]. Although it is possible to enrich
samples in particular fractions (i.e. differential centrifugation
and filtration) or DNA fragments (i.e. PCR amplification and
DNA hybridization) [96, 97], several factors should be taken into
account when considering a fast, in situ application. Mainly, it
would be especially difficult to adapt enrichment protocols to
field conditions, and they could cause substantial losses of ge-
netic material, add extra time to sample preparation, and result
in a significant bias.

In this context, targeted or selective real-time sequencing—
also known as ‘Read Until’—is a new approach for focusing the
sequencing process to specific DNA fragments of interest. Read
Until is based on the ability of programming nanopore
sequencers to reject individual DNA molecules while they are
being read [98], releasing the individual nanopore to sequence
another DNA fragment. ONT sequencing speed is estimated to
be 450 bp/s [98–100], and it is relatively common to achieve
sequences longer than 100 kbp [24, 101]. Theoretically, to dis-
card a read for being read after a few seconds of translocation
through the nanopore would prevent wasting sequencing ca-
pacity, which could be saved for sequencing targeted DNA frag-
ments [99]. In a metagenomic context, the Read Until strategy
could be used to deplete sequencing of undesirable DNA (i.e.
host DNA) or for enriching specific genes/genomes. This
depletion/enrichment procedures would not require any experi-
mental steps, thus facilitating their use under field conditions.

Selective sequencing was first demonstrated by Loose et al.
[102]. Later, Edwards et al. [103] showed the ability of Read Until
strategies to enrich E. coli genomic sequences over human DNA.
However, the actual revolution in targeted ONT sequencing is
taking place in the recent months, with three different
approaches being simultaneously released (Table 3). The first
one, named BOSS-RUNS, introduced the dynamic selection of
DNA regions of interest [100]. This method consists of focusing
sequencing efforts on areas that have achieved low coverage
during the run, thus leading to the compensation of sequencing
bias. With this methodology, De Maio et al. [100] were able to ef-
ectively enrich multiple loci of interest within a bacterial ge-
nome, enabling up to 5-fold coverage improvement. In the field
of metagenomics, BOSS-RUNS could be applied for improving
the characterization of samples by ensuring the deep sequenc-
ing of clade-specific genetic markers [104]. On the other hand,
Kovaka et al. [99] recently developed UNCALLED, a tool able
to directly map ONT raw signals in order to detect wanted/
unwanted sequences. They used this approach for sequencing a mock community (ZymoBIOMICS high molecular weight) con-
taining seven bacteria and one yeast. The objective was to map
the generated signals to a database containing the references
for the bacterial genomes (29 Mbp), rejecting DNA strands when
a match was detected. Bacterial sequencing depletion resulted
in up to 4.46-fold of yeast genome enrichment. Finally, another strategy based on DNA sequences comparison has been proposed by Payne et al. [98]. In this work, the same ZymoBIOMICS mock community was used, but the enrichment of the yeast genome was achieved in a different way. Briefly, sequencing started with default parameters, but when a pre-defined coverage was reached for a specific microorganism, its genome sequence was given to the Read Until application in order to reject DNA strands coming from this microorganism. Interestingly, the pipeline was adapted to incorporate a metagenome classifier (Centrifuge) [41], allowing the use of this strategy without prior knowledge of the sample.

Overall, selective sequencing has proved useful for different metagenomic applications. Nevertheless, an associated reduced total yield per flowcell has been reported [98, 99]. This could be explained by two main reasons: (i) rejecting DNA strands increase the time that a nanopore is not reading a molecule and (ii) voltage changes needed for rejecting the fragments may produce pore blockages [98]. Nuclease flush could potentially help to overcome this situation, although current throughputs are enough for enriching DNA sequences and reducing the time needed to reach the desired coverage [98, 99], which is a key factor in many in situ applications.

### Concluding remarks

In this work, we have reviewed the state-of-the-art, current research, and applications of real-time, in situ metagenomics. The spectacular development of metagenomic technologies in the last years as well as the number and importance of current and new challenges—including biomedical hazards—that could be addressed with portable metagenomic sequencing, reveals the importance of further developing this technology to match a variety of riches that we can, already, forecast. For example, we can envisage a close future in which microbial ecologists will be equipped with small, MinION-like devices that will allow both extract DNA, carry out a fast sequencing, and yield the results in a very short time. The understandability of the results and the minimization of the—visible—bioinformatic background will be very important to allow non-specialized staff to use such portable devices. The recent COVID-19 outbreak as well as the surveillance of Ebola, Zika, and many other emergent diseases will need an army of—not necessarily specialized—detectors, for which easy-to-run, easy-to-understand platforms will be needed. Alternatively, raw sequencing data will have to be transmitted through secure Internet-based applications to centralized points, in which specialist staff will further process and finally analyse the information. Such portable, easy-to-use, cheap devices will be used in quality control of all sorts of foods and ingredients; in the identification of crop pathogens on an individual plant basis; in forensic investigations; in the assessment of the energetic potential of different substrates or batches for biogas production; or for the identification of the best soils for specific crops, as deduced by the soil microbial (either taxonomic or functional) profile. In order to meet all these possibilities (which we have ambitiously described in future and not conditional tense), the combination of five traits will have to take place. The in situ, portable platform of the future will (have to be): inexpensive, robust, fast, easy to use, and connectable. A platform with these features will have a game-changing effect on the way we perform—and understand—microbial ecology.

### Acknowledgements

The authors want to thank all their colleagues from Darwin Bioprospecting Excellence who contributed to the development of MinION sequencing protocols and field applications. Adriel Latorre is a recipient of a Doctorado Industrial fellowship from the Ministerio de Ciencia, Innovación y Universidades (reference DI-17-09613).

### Conflict of interest statement

None declared.

### References

1. Rappé MS, Giovannoni SJ. The uncultured microbial majority. Annu Rev Microbiol 2003;57:1–11
2. Sanger F, Air GM, Barrell BG et al. Nucleotide sequence of bacteriophage \(\phi\)X174 DNA. Nature 1977;265:687–95.

3. Pace NR, Stahl DA, Lane DJ et al. The analysis of natural microbial populations by ribosomal RNA sequences. Advances in Microbial Ecology. Boston, MA: Springer, 1986, 1–55

4. Pace NR. Analyzing natural microbial populations by rRNA sequences. ASM News 1985;51:4–12

5. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Mol Biol Rev 1995;59:143–69

6. Handelsman J, Rondon MR, Brady SF et al. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol 1998;5:824–9

7. Quince C, Walker A, Simpson J et al. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 2017;35:833–44

8. Callahan B, McMurdie P, Holmes S. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J 2017;11:2639–43

9. Sharon I, Morowitz MJ, Thomas BC et al. Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. Genome Res 2013;23:111–20

10. Wu Y, Tang Y, Tringe S et al. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome 2014;2:25

11. Alneberg J, Karlsson C, Diven A et al. Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-assembled genomes. Microbiome 2018;6:173

12. Moore SC, Penrice-Randal R, Alruwaili M et al. Amplicon based MinION sequencing of SARS-CoV-2 and metagenomic characterisation of nasopharyngeal swabs from patients with COVID-19. medRxiv 2020

13. Menzel P, Ng K, Krogh A. Fast and sensitive taxonomic classification of metagenomic sequences. Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-assembled genomes. Microbiome 2018;6:173

14. Gowers G, Vince O, Charles J et al. On site DNA barcoding by nanopore sequencing. PLoS One 2017;12:e0184741

15. Edwards A, Debbonaire A, Nicholls S et al. In-field metagenome and 16S rDNA gene amplicon nanopore sequencing robustly characterize glacier microbiota. bioRxiv 2016

16. Johnson S, Zaikeva E, Goerlich D et al. Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. GigaScience 2018;7:ggy033

17. Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing and life detection in remote (79°26’N) Canadian high arctic permafrost ice wedge microbial communities. Front Microbiol 2017;8:2594

18. Harcourt J, Tamin A, Lu X et al. Isolation and characterization of SARS-CoV-2 from the first US COVID-19 patient. bioRxiv 2020

19. Screen-faceted proteomics to reveal Earth’s multiscale microbial diversity. Nature 2017;551:457–63

20. Alneberg J, Karlsson C, Divne A et al. Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-assembled genomes. Microbiome 2018;6:173

21. Moore SC, Penrice-Randal R, Alruwaili M et al. Amplicon based MinION sequencing of SARS-CoV-2 and metagenomic characterisation of nasopharyngeal swabs from patients with COVID-19. medRxiv 2020

22. Thompson L, Sanders J, McDonald D et al. A communal catalogue reveals Earth’s multiscale microbial diversity. Nature 2017;551:457–63

23. Alneberg J, Karlsson C, Divne A et al. Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-assembled genomes. Microbiome 2018;6:173

24. Edwards A, Debbonaire A, Nicholls S et al. In-field metagenome and 16S rDNA gene amplicon nanopore sequencing robustly characterize glacier microbiota. bioRxiv 2016

25. Johnson S, Zaikeva E, Goerlich D et al. Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. GigaScience 2018;7:ggy033

26. Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing and life detection in remote (79°26’N) Canadian high arctic permafrost ice wedge microbial communities. Front Microbiol 2017;8:2594

27. Harcourt J, Tamin A, Lu X et al. Isolation and characterization of SARS-CoV-2 from the first US COVID-19 patient. bioRxiv 2020

28. Moore SC, Penrice-Randal R, Alruwaili M et al. Amplicon based MinION sequencing of SARS-CoV-2 and metagenomic characterisation of nasopharyngeal swabs from patients with COVID-19. medRxiv 2020

29. Thompson L, Sanders J, McDonald D et al. A communal catalogue reveals Earth’s multiscale microbial diversity. Nature 2017;551:457–63

30. Alneberg J, Karlsson C, Divne A et al. Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-assembled genomes. Microbiome 2018;6:173

31. Moore SC, Penrice-Randal R, Alruwaili M et al. Amplicon based MinION sequencing of SARS-CoV-2 and metagenomic characterisation of nasopharyngeal swabs from patients with COVID-19. medRxiv 2020

32. Thompson L, Sanders J, McDonald D et al. A communal catalogue reveals Earth’s multiscale microbial diversity. Nature 2017;551:457–63

33. Menzel P, Ng K, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 2016;7:11257

34. Edwards A, Soares A, Rassner S et al. Deep sequencing: intraterrestrial metagenomics illustrates the potential of off-grid Nanopore DNA sequencing. bioRxiv 2017

35. Kim D, Song I, Breitwieser F et al. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res 2016;26:1721–9

36. Burton A, Stahl S, John K et al. Off earth identification of bacterial populations using 16S rDNA nanopore sequencing. Genes 2020;11:76

37. Carr C, Bryan N, Saboda K et al. Nanopore sequencing at mars, Europa and microgravity conditions. bioRxiv 2020
82. Leggett RM, Alcon-Giner C, Heavens D et al. Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. Nat Microbiol 2020;5:430–42
83. Mitsuhashi S, Kryukov K, Nakagawa S et al. A portable system for rapid bacterial composition analysis using a nanopore-based sequencer and laptop computer. Sci Rep 2017;7:5657
84. Tanaka H, Matsuo Y, Nakagawa S et al. Real-time diagnostic analysis of MinION™-based metagenomic sequencing in clinical microbiology evaluation: a case report. JA Clin Rep 2019;5:24
85. Cheng J, Hu H, Kang Y et al. Identification of pathogens in culture-negative infective endocarditis cases by metagenomic analysis. J Antimicrob Chemother 2017;72:104–14
86. Charalampous T, Kay GL, Richardson H et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol 2019;37:783–92
87. Sanderson ND, Street TL, Foster D et al. Real-time analysis of nanopore-based metagenomic sequencing from infected orthopaedic devices. BMC Genom 2018;19:714
88. Gong L, Huang YT, Wong CH et al. Culture-independent analysis of liver abscess using nanopore sequencing. PLoS One 2018;13:e0190853
89. Nakagawa S, Inoue S, Kryukov K et al. Rapid sequencing-based diagnosis of infectious bacterial species from meningitis patients in Zambia. Clin Transl Immunol 2019;8:e1087
90. Ibironke O, McGuinness L, Lu S et al. Species-level evaluation of the human respiratory microbiome. GigaScience 2020;9:giaa038
91. D’Andreano S, Cuscó A, Francino O. Rapid and real-time identification of fungi up to the species level with long amplicon Nanopore sequencing from clinical samples. bioRxiv 2020.
92. Nelson M, Pope C, Marsh R et al. Human and extracellular DNA depletion for metagenomic analysis of complex clinical infection samples yields optimized viable microbiome profiles. Cell Rep 2019;26:2227–40.e5
93. Feigelman R, Kahlert C, Baty F et al. Sputum DNA sequencing in cystic fibrosis: non-invasive access to the lung microbiome and to pathogen details. Microbiome 2017;5:20
94. O’Flaherty B, Li Y, Tao Y et al. Comprehensive viral enrichment enables sensitive respiratory virus genomic identification and analysis by next generation sequencing. Genome Res 2018;28:869–77
95. Covey L, Fan Y, Pullan S et al. Metagenomic nanopore sequencing of influenza virus direct from clinical respiratory samples. J Clin Microbiol 2019;58:e00963–19
96. Payne A, Holmes N, Clarke T et al. Nanopore adaptive sequencing for mixed samples. Whole exome capture and targeted panels. bioRxiv 2020.
97. Kovaka S, Fan Y, Ni B et al. Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. bioRxiv 2020.
98. De Maio N, Manser C, Munro R et al. BOSS-RUNS: a flexible and practical dynamic read sampling framework for nanopore sequencing. bioRxiv 2020.
99. Jain M, Olsen H, Turner D et al. Linear assembly of a human centromere on the Y chromosome. Nat Biotechnol 2018;36:321–3
100. Loose M, Malla S, Stout M. Real-time selective sequencing using nanopore technology. Nat Methods 2016;13:751–4
101. Edwards H, Krishnakumar R, Sinha A et al. Real-time selective sequencing with RUBRIC: read until with basecall and reference-informed criteria. Sci Rep 2019;9:11475
102. Tickle T et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods 2015;12:902–3