Received: 18 March 2019 | Revised: 26 March 2019 | Accepted: 26 March 2019

DOI: 10.1111/dgd.12608

REVIEW ARTICLE

WILEY Development, Growth & Differentiation

Nanopore sequencing: Review of potential applications in functional genomics

Nobuaki Kono | Kazuharu Arakawa

Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, Japan

Correspondence
Kazuharu Arakawa, Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, Japan.
Email: gaou@sfc.keio.ac.jp

Funding information
ImPACT Program of Council for Science, Technology and Innovation (Cabinet Office, Government of Japan); Yamagata Prefectural Government and Tsuruoka City, Japan

Abstract

Molecular biology has been led by various measurement technologies, and increased throughput has developed omics analysis. The development of massively parallel sequencing technology has enabled access to fundamental molecular data and revealed genomic and transcriptomic signatures. Nanopore sequencers have driven such evolution to the next stage. Oxford Nanopore Technologies Inc. provides a new type of single molecule sequencer using protein nanopore that realizes direct sequencing without DNA synthesizing or amplification. This nanopore sequencer can sequence an ultra-long read limited by the input nucleotide length, or can determine DNA/RNA modifications. Recently, many fields such as medicine, epidemiology, ecology, and education have benefited from this technology. In this review, we explain the features and functions of the nanopore sequencer, introduce various situations where it has been used as a critical technology, and expected future applications.

KEYWORDS
long reads, nanopore sequencing, next generation sequencing

1 | INTRODUCTION

Novel technologies that visualize the unseen or detect the undetectable have always contributed to breakthroughs in scientific discoveries, and the rapid advent of high-throughput and affordable DNA sequencing technologies has undoubtedly been the key driving force in the progress of life sciences over the last decade (Goodwin, Mcpherson, & Mccombie, 2016). Genomic information has also been one of the cores of molecular biology, providing and assisting tools to probe the genome, its structure, epigenetics, gene expression, and a multitude of other applications. Latest in the line of DNA sequencers are the nanopore sequencers (Heather & Chain, 2016; Jain, Olsen, Paten, & Akeson, 2016; Mardis, 2013), successfully commercialized by Oxford Nanopore Technologies (ONT) (Brown & Clarke, 2016). While a variety of sequencing approaches exist such as pyrosequencing, sequencing by synthesis, sequencing by ligation, and single molecule real time (SMRT) sequencing, DNA sequencing methods since Sanger sequencing predominantly relied on the process of DNA synthesis (Clarke et al., 2009; Eid et al., 2009; Schadt, Turner, & Kasarskis, 2010). Nanopore sequencing distinguishes itself from these previous approaches, in that it directly detects the nucleotides without active DNA synthesis, as a long stretch of single stranded DNA passes through a protein nanopore that is stabilized in an electrically resistant polymer membrane (Branton et al., 2008; Feng, Zhang, Ying, Wang, & Du, 2015). By setting a voltage across this membrane, sensors detect the ionic current changes shifted by nucleotides occupying the pore in real time as the DNA molecule passes through. History of the development of nanopore sequencing of DNA, as well as detailed methods for sequencing, is already reviewed in depth elsewhere (review in (Deamer, Akeson, & Branton, 2016)) and is beyond the scope of this review. Here, we focus on the current advantages as well as the limitations of the ONT nanopore sequencer, reviewing research publications available thus far. It should be noted, however, that in a rapidly evolving technology like nanopore sequencing, reviewing efforts can also quickly become outdated.

Nanopore's unique sequencing method provides multiple advantages over existing technologies. Firstly, nanopore sequencing

does not require imaging equipment to detect the nucleotides, allowing the system to scale down in size to a portable level. The current MinION Mk1B device only weighs 90 g, measuring only 3.3 × 10.5 × 2.3 cm in size, fitting in the palm of one's hand. The cost of the device is also much lower compared to other massively parallel sequencers, the initial cost being only around $1,000, including the device and initial set of reagents. MinION devices can be powered through the USB (Universal Serial Bus) port of laptop computers, such that sequencing can be conducted anywhere, even in the field. Lack of an image analysis step also allows real time base calling during sequencing, realizing rapid detection of target DNA for the screening of pathogens from clinical samples, for example. Secondly, since nanopore sequencing directly detects the input molecule without DNA amplification or synthesis, there is no apparent limit to the length of DNA that can be sequenced. The challenge in read length using nanopore sequencers therefore is not in the sequencing technology itself, but in the library preparation step, which needs to extract and load intact extremely high molecular weight (HMW) DNA into the flow cell of the sequencer. Sequenced reads exceeding a mega-base have been reported, demonstrating the extraordinary capabilities of the nanopore device in sequencing extremely long stretches of the DNA molecule (Jain et al., 2018). The extreme long reads enable de novo genome assembly without preparation of complex mate-pair libraries typically required with short read sequencing. The extremely long reads are also useful to determine sequences of genomic region containing long repetitive sequences, which is difficult with short read sequencing. Long reads also allow the study of structural variations within the genome (Cretu Stancu et al., 2017). Moreover, detection of the ionic current changes shifted by the nucleotides passing through the nanopore is not limited to the canonical four nucleobases of adenine, guanine, cytosine and thymine. Taking advantage of the nature of unamplified direct sequencing, nanopore sequencing can directly observe base modifications such as methylations (Rand et al., 2017; Simpson et al., 2017), and even direct sequencing of RNA molecules containing uracil bases (Garalde et al., 2018).

The main current drawback of nanopore sequencing, as with other long read sequencers, is the relatively high error rate compared to short read sequencing. Current error rates range from 5% to 20%, dependent on the type of molecules and library preparation methods, and the errors include both insertions and deletions (Rang, Kloosterman, & De Ridder, 2018). Unlike SMRT sequencing of PacBio, there seems to be systematic error in nanopore sequencing (Jain et al., 2018), so error correction typically requires additional short read sequence data. Data yield also varies rather significantly depending on the input, which is also difficult to predict. On the other hand, active developments in bioinformatics for base-calling and error correction, as well as optimization on library preparation steps and new pore developments by ONT, are rapidly improving on these issues. Throughput of the system is greatly enhanced; for example, by the recent introduction of the PromethION system, where a single flow cell can yield 50–100 Gbp (typical yield of MinION system is 5–10 Gbp) and 24 flow cells can be run in parallel. Improvements in base-caller software from Hidden Markov Model (HMM) based methods to Recurrent Neural Network (RNN) based algorithms enhanced base-level accuracy by 2%–5% (Rang et al., 2018; Teng et al., 2018). In the following sections, we review each of the above points in detail.

2 | ULTRA-LONG READS - "WHALE WATCHING"

The official library preparation protocol with adapter ligation provided by ONT recommends an optional fragmentation step for the input DNA to average in the 8 kbp range, in order to have optimal molar concentration of adaptor-ligated DNA ends to meet the nanopores during sequencing to gain maximal throughput, which is around 0.2 pM. However, since there is no apparent technical limit for the size of DNA to be sequenced, the nanopore community primarily led by Loman et al. called out for "whale spotting" with nanopore (http://lab.loman.net/2017/03/09/ultrareads-for-nanopore/). A "whale" is a comparative measurement system that relates ultra-long DNA bp to the nearly equivalent numbers of grams: for example, one of the smallest whales is the narwhal, weighing around 940 kg, and read lengths of 940 kbp to 1 mbp were the initial challenge. In order to load such HMW DNA into the sequencer, every step of DNA handling required reconsideration. Typical DNA extraction methods using commercial kits such as the silica spin column usually result in relatively shorter DNA molecules, typically below 50 kbp. Drying of DNA to eliminate residual ethanol during purification, clean-up and concentration with paramagnetic Solid Phase Reversible Immobilization (SPRI) beads, and even the pipetting of DNA solution result in fragmentation of ultra-long DNA molecules. Therefore, Quick developed a protocol based on the classic Sambrook method (Sambrook & Russell, 2001) of using phenol-chloroform-isoamyl alcohol with wide-bore pipette tips to extract HMW DNA from cultured cells, and then prepared a nanopore library with a Rapid Sequencing Kit (ONT) that utilized transposons to add the sequencing adapters with minimal pipetting steps (Quick J. 2018. Ultra-long read sequencing protocol for RAD004. https://www.protocols.io/view/ultra-long-read-sequencing-protocol-for-rad004-mrxc57n). A larger amount of input DNA was also required due to the decreased molarity of DNA ends with ultra-long molecules. One of the very first "whales" was successfully spotted with this protocol for human culture cells, where an ultra-long read of length 882 kbp was mapped to the reference spanning a 950 kbp region (Jain et al., 2018). With a sequencing speed of 450 bp/s for 1D kits using R9.4 flow cells, the "whale watching" of this single read took more than 30 min as it passed through the pore. Bioinformatics methods were also required to observe the entirety of such ultra-long sequencing, for the software controlling the sequencing (MinKNOW) erroneously subdivided some of the continuous data streams of ultra-long reads into shorter multiple reads. A software to observe such discontinuity, designated BulkVis, was developed, which observed the largest "whale" yet at 2,272,580 bp (Payne, Holmes, Rakyan, &

Loose, 2018). This single ultra-long read is comparable to the size of smaller bacterial chromosomes such as those of *Bifidobacterium longum* NCC2705, with 2,256,640 bp (Schell et al., 2002), or nearly one half of the *Escherichia coli* genome, demonstrating the potential to sequence chromosome-level reads with nanopore technology. It is also interesting that this kind of new technology reevaluates the rather classic methods of DNA handling, such as the use of the phenol-chloroform-isoamyl alcohol method, vacuum concentration, pulse-field gel electrophoresis (PFGE) with gel plugs and tapping or gentle rotation as opposed to spin columns, the use of wide bore pipette tips, pipette tip cutting, vortexing, SPRI clean-up, or regular submarine gel electrophoresis. In order to maximize the read length in nanopore sequencing, one needs to take extra care for such HMW DNA handling throughout the library preparation steps.

The extraction protocol should be adjusted according to the target species, and cell lysis and homogenization procedures have a profound influence on the extracted nucleotide quality for tissue samples, unlike Gram-negative bacteria or cultured cells that lack a thick cell wall or exoskeleton. Therefore, homogenization and DNA extraction methods should be optimized for each species, especially for plants or arthropods that harbor thick cell walls and cuticular exoskeletons. In the case of plant genomes such as the fastest-growing angiosperm (greater duckweed) or tropical timber trees (teak), HMW genomic DNA extraction has been successfully performed by a CTAB method (Hoang et al., 2018; Yasodha et al., 2018), following homogenization steps to break the cell wall, where grinding with liquid nitrogen is generally used to crush the cell wall while cellular enzymes remain inactivated. After sufficient homogenization, samples are then resuspended in detergent-based extraction buffers containing cetyltrimethylammonium bromide (CTAB). Due to the abundance of polysaccharides in plants, the CTAB method is established as the best detergent for purifying DNA from plant material (Murray & Thompson, 1980). Likewise, in insects such as harlequin ladybird or fruit fly, a dissected tissue or a whole specimen frozen with liquid nitrogen is ground using a pestle, and HMW DNA extraction is subsequently performed with Genomic-tip 500/G kit (QIAGEN) or Blood & Cell Culture kit (QIAGEN) (Gautier et al., 2018; Miller, Staber, Zeitlinger, & Hawley, 2018). For vertebrates such as clownfish or eel, suitable tissue samples such as muscles or livers are selected (Jansen et al., 2017; Tan et al., 2018).

Coupled with the portable, real time, single molecule nature of the nanopore sequencing technology, it is worth noting that the long reads are expected to contribute to translational applications in genomic medicine and clinical testing (Ameur, Kloosterman, & Hestand, 2019). Notably, long reads are critical in structural variation analysis and the complete sequencing of repetitive DNA contents of clinical utility. Norris and colleagues demonstrated that nanopore sequencers could detect large-scale structural variations, including large deletions, inversions, and translocations related to the inactivation of tumor suppressor genes in pancreatic cancer, with very few reads (Norris, Workman, Fan, Eshleman, & Timp, 2016). Another group developed a new bioinformatics tool, NanoSV, for structural variation detection with nanopore data using split read mapping (Cretu Stancu et al., 2017). Mitsuhashi and colleagues successfully identified a full-length microsatellite repeat spanning 49,877 bp in the D4Z4 array responsible for facioscapulohumeral muscular dystrophy (Mitsuhashi, Nakagawa, et al., 2017). In the case of HLA genotyping, the latest MinION accuracy has reached a quality that offers cost-effective and scalable genotyping with minor shortcomings (Lang et al., 2018). Furthermore, hybrid assembly of combined short- and long-reads could resolve the structure and chromosomal insertion site of the antibiotic resistance island in *Salmonella* Typhi Haplotype 58 (Ashton et al., 2015), opening up numerous possibilities for real-time genomic epidemiology.

3 | LONG READ ASSEMBLY STRATEGIES

In short read assembly, the de Bruijn graph (DBG) algorithm has been most commonly used due to its accuracy and speed (Lu, Giordano, & Ning, 2016). The DBG algorithm splits sequenced reads into their k-mer components and constructs a graph with connecting pairs of k-mers based on whether they have k-1 common nucleotides. This approach, however, is highly dependent on the determination of the k-mer graph, which is not suitable for long reads with innately high error rates, and cannot fully take advantage of the longer lengths of the reads, to take account of complex genomic features including structural variation or non-random elements (Sohn & Nam, 2018). Long read assembly therefore revived interest in an overlap-layout-consensus (OLC) algorithm with all-against-all pairwise alignment, which is flexible in read length and robust to error, allowing for a decrease in chimeric contigs or assembly bubbles. Canu has been the most popular assembly software used in numerous nanopore assembly projects, which includes tf-idf weighted MinHas for fast and accurate adaptive overlapping Nanopore-only error correction and using the raw electric current signal information can significantly improve base-level accuracies to around 99%, as demonstrated in the whole genome assembly of *E. coli* K-12 MG1655 (Loman, Quick, & Simpson, 2015). This approach, however, still ends with a higher base-level error rate in comparison to short read assembly, and is highly computationally intensive. A hybrid strategy utilizing short Illumina reads is therefore commonly adopted to polish the assembly (Figure 1a) (Giordano et al., 2017; Istace et al., 2017). Multiple rounds of polishing are often necessary, where the error correction process is validated using Benchmarking Universal Single-Copy Orthologs (BUSCO) completeness scores (Jain et al., 2018; Simao, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015; Tan et al., 2018; Tyson et al., 2018). Several assemblers incorporate the hybrid error correction step prior to the assembly, in place of the initial overlap-based error correction (Bankevich et al., 2012; Zimin et al., 2013). While hybrid error correction is computationally efficient and generally results in highly accurate sequences, it should be noted that the disadvantage of short reads remains in this

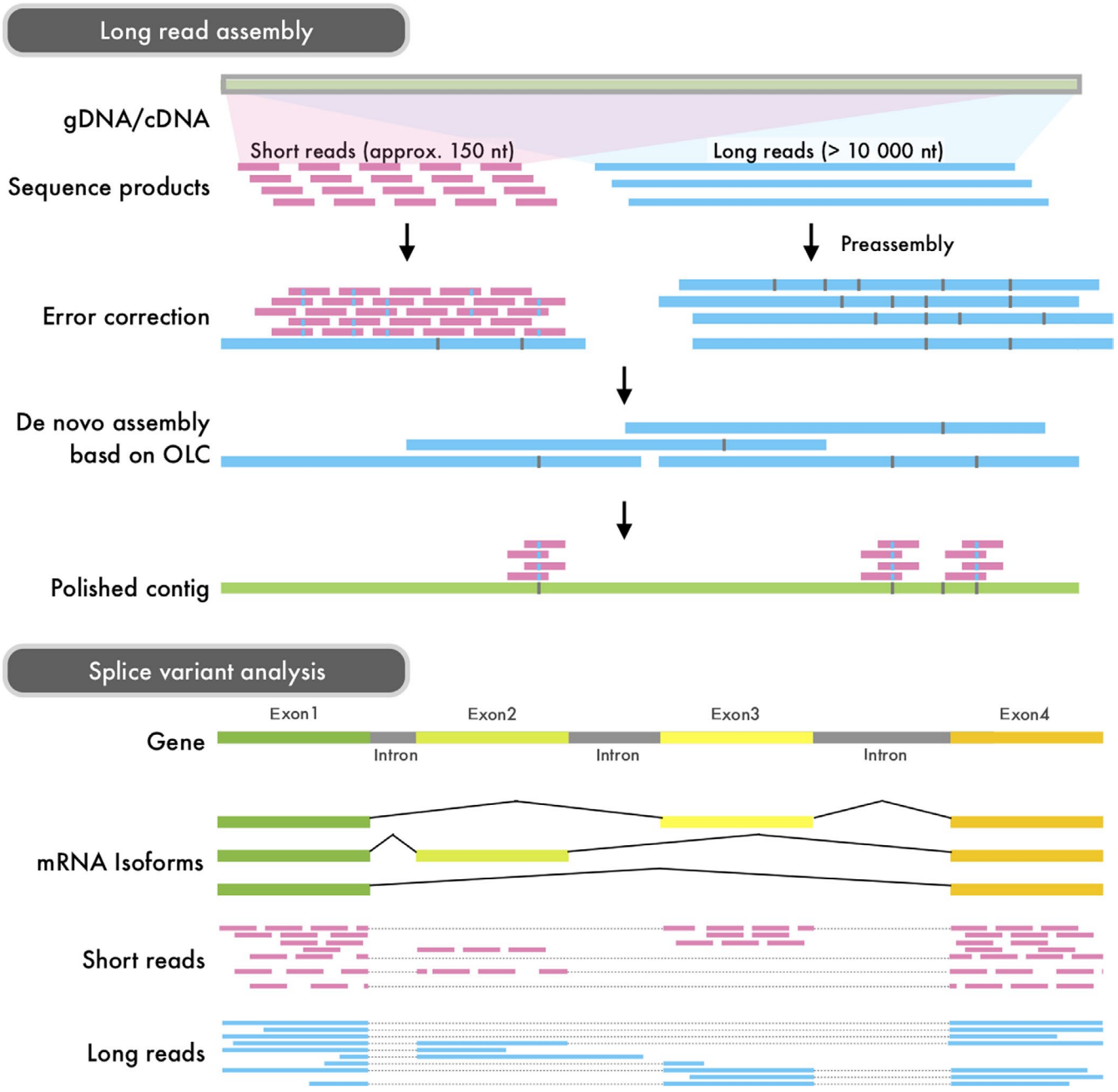

FIGURE 1 (Long read assembly) The nanopore read assembly requires error correction, assembly, and polishing processes. Error correction can be performed by hybrid or long read only approaches. Hybrid methods use high-accuracy short read. Assembled contigs are polished to improve the consensus sequence by using high-accuracy read or raw current data obtained by nanopore sequencer. (Splice variant analysis) Long read obtained by direct sequencing allows detection of structural variation as it is. Therefore, the splice variant analysis does not require assembly

approach, where error correction of non-random or repetitive regions is difficult.

4 | TRANSCRIPTOME ANALYSIS AND DIRECT RNA SEQUENCING

The advantages of long reads are not limited to the study of genomic DNA, but are also useful in the study of RNA sequences. It is especially pertinent to uncover the diversity of alternative splicing isoforms and their expression levels, which are often difficult to delineate with short reads (Figure 1b). Consequently, the combination of long and short reads is reported to increase transcriptome analysis performance (Weirather et al., 2017). Use of long reads for the analysis of splice isoforms has clear advantages over short reads, in that there is no need for read mapping or assembly to figure out the isoform structures, but one can instead simply look at the entire sequence to pinpoint the isoform. The *Dscam1* (Down syndrome

cell adhesion molecule 1, (Schmucker et al., 2000)) gene structure in *Drosophila*, which produces a total of 38,016 different isoforms via alternative splicing arising from four variable exon clusters, is regarded as "the most complicated alternatively spliced gene known in nature" (Bolisetty, Rajadinakaran, & Graveley, 2015). Nanopore full-length cDNA sequencing analysis, however, successfully identified 7,899 full-length isoforms of this gene. The accurate identification of incorporated exon structures using long reads is also powerful in its application to evidence-based gene structural annotation. Cook and colleagues developed an automated gene structure annotation pipeline named LoReAn (long-read annotation), in which they demonstrated that a full-length cDNA sequence could assist correct annotation of the gene structures in *Arabidopsis thaliana* and *Oryza sativa* (Cook et al., 2019).

Long read cDNA sequencing is also possible with other platforms, but the nanopore sequencer is the only device that has made the direct sequencing of long stretches of RNA molecule possible (Garalde et al., 2018). Although tRNA detection by nanopore has previously been explored and reported (Henley et al., 2016; Smith, Abu-Shumays, Akeson, & Bernick, 2015), the ONT nanopore sequencer commercialized direct RNA sequencing kits in 2017. Conventional "RNA-seq" with short reads, which is rather inappropriately termed so since the RNA molecules are not directly sequenced (Hrdlickova, Toloue, & Tian, 2017), requires reverse transcription (RT) of the template RNA molecule into complementary DNA (cDNA), which is typically further amplified by PCR. Direct RNA sequencing (again, it is unfortunate that true RNA sequencing has to add the word "direct" to differentiate itself from conventional "RNA-Seq"), on the other hand, directly detects the ribonucleobases passing through the pore without RT or PCR amplification, and this approach is free of possible biases or misamplifications introduced during such steps. The gene annotation process is simplified by the direct RNA sequencing, and it has therefore allowed the identification of more complex or novel transcript isoforms genome-wide (Byrne et al., 2017; Krizanovic, Echchiki, Roux, & Sikic, 2018), and the ability to differentiate transcript haplotypes as well as to identify 3′ poly (A) tail lengths (Workman et al., 2018). Such an application is also practical for the study of RNA viral genomes, because the genome has multiple reading frames, anti-sense gene locations, inefficient termination signals, and complex splice forms, and gene annotation is challenging using the conventional RNA-seq method (Depledge et al., 2018).

Moreover, another strength of unamplified or non-reverse-transcribed direct RNA sequencing is the detection of nucleotide modifications. Since the presence of modified bases results in altered ionic current signal from the unmodified base as it passes through the pore, nanopore sequencing is able to detect the modifications without any additional sample preparation. Rand and colleagues demonstrated that three cytosine variants within DNA molecules could be detected with an accuracy of 80% using an HMM-HDP (HMM with a Hierarchical Dirichlet Process) (Rand et al., 2017), and Simpson and colleagues also proved that the nanopore sequencer could detect the 5-methylcytosine (5-mC) at up to 95% accuracy using HMM (Simpson et al., 2017). For RNA modification, N6-methyladenosine (m6A) or 5-mC is the most common internal modifications in mRNA that are implicated in various RNA metabolism and regulations, and nanopore sequencing can identify these modifications (Garalde et al., 2018). A comprehensive study of the human poly (A) transcriptome using direct RNA sequencing has recently been reported, including the detection of these modifications as well as A-to-I RNA editing (Workman et al., 2018). Although it should be noted that the error rate of direct RNA sequencing exceeds the already high error rate of nanopore DNA sequencing, the ability to directly study RNA molecules could open up a wide range of research that are otherwise not possible.

5 | REAL TIME AND PORTABLE

The excellent portability of a MinION device that can be bus-powered from a mobile PC can make a significant contribution in the context of in-the-field sequencing (Ameur et al., 2019; Quick et al., 2017), which has already been demonstrated in extreme places such as an academic conference venue (Ii et al., 2019), coal fields (Edwards et al., 2017), the arctic (Edwards, Debbonaire, Sattler, Mur, & Hodson, 2016), Antarctica (Johnson, Zaikova, Goerlitz, Bai, & Tighe, 2017), and even on the International Space Station (Castro-Wallace et al., 2017). For these purposes, mobile lab approaches for DNA extraction and library preparation have been explored using devices like VolTRAX V2 or Bento Lab (all-in-one DNA laboratory including thermal cycler: https://www.bento.bio), and experimental protocol developments are also underway, such as those using LAMP (loop-mediated isothermal amplification) methods instead of conventional PCR to eliminate the need for a thermal cycler. LAMP only requires a water bath to conduct its isothermal amplification, and Imai and colleagues demonstrated an identification system for human *Plasmodium* species (Imai et al., 2017). Yamagishi and colleagues also reported the serotyping system for dengue virus in a single day (Yamagishi et al., 2017) using this approach.

Moreover, the real time nature of the device, where DNA molecules passing through the pore are immediately base-called on a per-read basis as they are sequenced, becomes a more rapid and highly robust alternative to conventional clinical testing such as chemiluminescence, ELISA, immune chromatographic tests (ICT), and real-time PCR. Greninger and colleagues implemented a web-based pipeline for real-time analysis (MetaPORE) and made it possible to detect Chikungunya virus, Ebola virus, and hepatitis C virus simultaneously from a human blood sample with nanopore sequencing (Greninger et al., 2015). Likewise, Mitsuhashi and colleagues developed a real-time identification of bacterial composition within 2 hr of obtaining a sample (Mitsuhashi, Kryukov, et al., 2017), and Quick et al. (2015) showed that genotyping of a hospital outbreak of Salmonella could be carried out in less than half a day. Furthermore, combining the portable and real-time nature of the MinION device, it has been adopted for the surveillance and monitoring of viral outbreaks in resource-limited locations. The Zika virus outbreak was declared an international public health emergency by the World

Health Organization (WHO) in 2016, and the Zika in Brazil Real-time Analysis (ZiBRA) project was established to sequence a thousand genomes from Brazil to monitor the epidemiological information (Faria et al., 2016). Quick et al. (2017) thus developed a protocol to rapidly obtain the complete genome of the Zika virus from clinical samples using the MinION sequencer, encompassing sample preparation to bioinformatics analysis. The same research group also developed a genomic surveillance system using the nanopore sequencer for the Ebola outbreak in West Africa (Quick et al., 2016). This system takes as little as 15–60 min sequencing after receiving the blood sample, realizing an on-the-spot virus testing without the need for the transportation of hazardous specimens. Furthermore, Runtuwene and colleagues demonstrated a new MinION application for genotyping analysis of the malaria parasite *Plasmodium falciparum* in Indonesia (Runtuwene et al., 2018), stressing the potential of the nanopore genotyping approach as an effective and practical alternative for the diagnosis of the malaria parasite.

6 | USE CASE: HIGHLY REPETITIVE SPIDER SILK GENES

The improved accessibility of sequencing technology makes it easier to access genomic data, and contributes significantly to molecular biology in non-classical model organisms (Russell et al., 2017). Long read assemblies are especially useful in this regard in complex large eukaryotic genomes, typically with numerous non-random features such as highly repetitive sequences and polyploidy (Kono, Nakamura, Ohtoshi, Pedrazzoli Moran, et al., 2019; Kono, Nakamura, Ohtoshi, Tomita, Numata, et al., 2019). As one example, we here introduce the de novo sequencing of a spider silk protein (spidroin) gene. The spider silk exhibits extraordinary physio-chemical properties for potential industrial applications as a multifunctional protein material, such as high toughness, high tensile strength per density, and thermostability (Blamires, Blackledge, & Tso, 2017; Omenetto & Kaplan, 2010). Multiple spidroin families exist, corresponding to different types of silk with various mechanical properties, but they all share conserved non-repetitive N/C-terminal domains (Hayashi, Blackledge, & Lewis, 2004), and a highly complex sequence organization that is extremely long: typically exceeding 10kbp, and is almost entirely comprised of repetitive sequences. The multi-giga-base size of spider genomes (Babb et al., 2017) and this characteristic spidroin gene structure (Xu & Lewis, 1990) makes the determination of the full length of the sequence challenging, as even from the PCR amplification step of the gene there is a risk of producing chimeric artifacts. Unamplified single molecule long nanopore sequencing is one of a very limited number of possible approaches.

In order for an untargeted, comprehensive search for spidroin fragments, the non-repetitive N/C-terminal domains as well as repeat unit sequences were computationally searched from a short read transcriptome assembly. These fragments were then extended by an OLC of short reads until the contigs met unresolvable bifurcations in the DBG. Then, the collected extended fragments were ordered based on nanopore long reads directly obtained from mRNA or genomic DNA to correct for base-level accuracies, so that the gene length and the arrangement of repeat units were guaranteed, avoiding compressed alignment (Ricker, Qian, & Fulthorpe, 2012). Here, we illustrate a specific example of the spidroin gene sequence obtained by nanopore sequencer (Figure 2). *Araneus ventricosus* is an orb-weaving spider and has seven spidroin gene families with a combination of repetitive and non-repetitive unit structures as described above. One of the spidroin genes, the minor ampullate spidroin (*MiSp*) gene, was determined by low-coverage (5.5M reads) direct DNA sequencing. The collection of long reads covering the full-length of *MiSp* gene sequence from numerous nanopore reads was performed based on similarity with the *MiSp* gene fragments constructed by short read assembly in advance. In the previous study, it was reported that the *MiSp* gene sequence was over 5,000 bp in length (Chen et al., 2012), and nanopore sequencing showed over 8.3 kbp in length. The same case of another gene (aciniform spidroin:

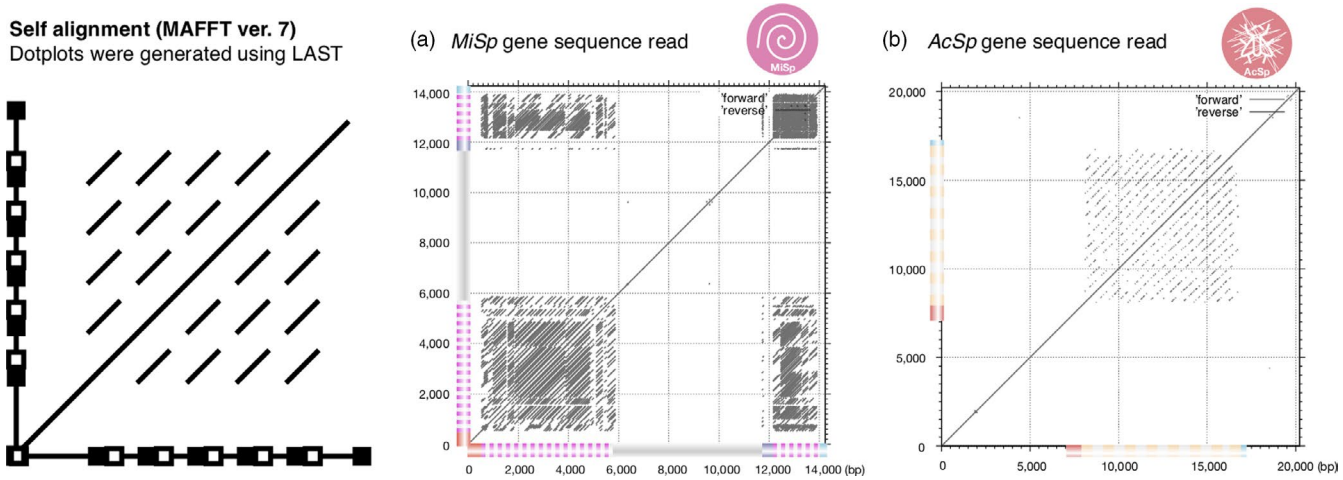

FIGURE 2 The dot plots represent the nanopore reads covering full length gene. The area where dot concentrates indicates the receptive region. Red and blue box mean N- and C-terminus domain. Gray box means intron region. (a) and (b) show *MiSp* and *AcSp* gene reads obtained by *Araneus ventricosus* genomic DNA. These dotplots were generated using LAST (Katoh & Standley, 2013)

TABLE 1 List of bioinformatics software commonly used in the analysis of nanopore data

	Software	Description	Reference
Basecaller	Albacore	Basecaller developed by Oxford Nanopore Technologies Inc.	https://nanoporetech.com
	Chiron	Establish end-to-end basecalling using a deep learning CNN+RNN+CTC structure	(Teng et al., 2018)
	Metrichor	Cloud-based basecaller that cannot be run locally	https://nanoporetech.com
	Tombo	Detection tool for modified nucleotides such as methylation	https://nanoporetech.com
Polishing	Nanopolish	HMM-based polishing tool using raw signal data of reads	(Louis et al., 2016)
	Racon	Polishing tool for contigs assembled by Canu (Koren et al., 2017) with raw long reads.	(Vaser, Sovic, Nagarajan, & Sikic, 2017)
	Pilon	Polishing tool for contigs assembled by Canu (Koren et al., 2017) with Illumina short reads.	(Walker et al., 2014)
Assembler	DBG2OLC	Hybrid large genome assembler using short read contigs as anchor points for overlap graph construction with long reads	(Ye, Hill, Wu, Ruan, & Ma, 2016)
	Wtdbg2	Fast long-read assembler based on fuzzy-Bruijn graph (FBG) that is analogy to de Bruijn graph but permits mismatches and gaps	(Ruan & Li, 2019)
	RaGOO	Scaffolding tool using reference-guided contig ordering and orienting	(Alonge, Soyk, Ramakrishnan, & Wang, 2019)
	MaSuRCA	Hybrid genome assembler using long and short reads	(Zimin et al., 2013)
	Canu	Canu is one of the leading Celera assembler that designed for SMRT and ONT reads. A successor of PBcR (Miller et al., 2008)	(Koren et al., 2017)
	SMARTdenovo	Fast and reasonably accurate OLC assembler for long reads without prior error correction step	(Giordano et al., 2017)
	TULIP	Efficient assembler for large and complex genome using seed extension	(Jansen et al., 2017)
	Falcon	Long read assembler for phased diploid genome	(Chin et al., 2016)
	Miniasm	Efficient assembler for SMRT and ONT reads without an error correction	(Li, 2016)
	Unicycler	Assembler specially designed for hybrid assembly of small (bacterial or viral) genomes based on SPAdes	(Wick, Judd, Gorrie, & Holt, 2017)
	ABruijn	De novo assembler for long and noisy read using de Bruijn graph	(Lin et al., 2016)
	HINGE	Long read OLC assembler based without pre-assembly or read correction step	(Kamath, Shomorony, Xia, Courtade, & Tse, 2017)
SV/SNV caller	NanoSV	Breakpoint junction detection of structural variant using mapping BAM file produced by LAST	(Cretu Stancu et al., 2017)
	Sniffles	Structural variation caller using automatic filtering of false events based on coverage data	(Sedlazeck et al., 2018)
	npInv	Specially designed tool for non-allelic homologous recombination (NAHR) inversions	(Shao et al., 2018)
	marginCaller	SNV detection program in marginAlign package	https://github.com/benedictpaten/marginAlign
Alignment	minimap2	Fast mapping and alignment program for long noisy reads	(Li, 2018)
	minialign	Fast and moderately accurate alignment tool for long reads	https://github.com/ocxtal/minialign
	LAST	Genome-scale sequence comparison tool to find similar region	(Kielbasa, Wan, Sato, Horton, & Frith, 2011)
	NGMLR	Long read aligner with a new convex gap-cost scoring model for SNV caller (Sniffles)	(Sedlazeck et al., 2018)
	marginAlign	Package to align reads to a reference genome for SNV calling (marginCaller)	https://github.com/benedictpaten/marginAlign
	BWA-MEM	Leading short read mapping tool tuned to work with ONT recently	https://github.com/lh3/bwa
Others	NanoPipe	Interactive web-based tool for fast and easy processing and analysis of the ONT data	(Shabardina et al., 2019)
	BulkVis	Visualization tool for bulk FAST5 files obtained from ONT sequencing	(Payne et al., 2018)

AcSp) is also shown in Figure 2. Because of the unamplified single molecule long read sequencing of genomic DNA, not only the gene length, but also the repetitive or exon/intron structure were clearly uncovered.

7 | LIMITATIONS

The largest limitation of the nanopore sequencer is the comparatively lower read accuracy when compared to short read sequencers. Because insertions and deletions are included in the errors, nanopore reads per se are not optimal for single nucleotide variation (SNV) detection. Although SNV genotyping with nanopore-sequenced reads has been demonstrated, high coverage reads were required (Ebler, Haukness, Pesout, Marschall, & Paten, 2018; Koren et al., 2017), and HLA genotyping also has problems in that it cannot distinguish specific alleles due to a lack of read accuracy (Jain et al., 2018).

Improving the accuracy of nanopore sequencers depends on both the pore chemistry and the base-calling algorithm (Patel et al., 2018). The latest R9.X nanopore is derived from the *E. coli* Curlin sigma S-dependent growth (CsgG) gene (Goyal et al., 2014), achieving significantly reduced error rate (Patel et al., 2018; Rang et al., 2018). According to the Nanopore Community Meeting 2018 (https://nanoporetech.com/resource-centre/videos/community-meeting), a new protein, R10 nanopore, is due to be released in 2019, further improving the accuracy for homopolymers. So far, R9.X nanopore has a sharp reader that can only dominate five bases, making accurate determination of the homopolymers difficult. However, the new chemistry has long and multiple "readers" which can dominate more base signal, so homopolymer sequence accuracy is projected to improve from Q34 of R9.4 to Q40 at 75× coverage.

Basecalling algorithms that convert raw ionic current signal to nucleotide sequences based on a deep learning approach, combining a convolutional neural network (CNN), connectionist temporal classification, and a RNN, improved the read accuracy dramatically (Rang et al., 2018) over the previous HMM-based approach from segmented raw data according to k-mer (Boza, Brejova, & Vinar, 2017; Stoiber & Brown, 2017). As described above, many limitations can be overcome by improved software fine-tuned for use with nanopore data. Many bioinformatics tools specialized or optimized for nanopore sequencing have been released, and many more are in active development. Table 1 lists representative software tools useful in the analysis of nanopore data.

ACKNOWLEDGMENTS

The authors would like to thank Dr. James Fleming for his diligent proofreading of this manuscript. This work was funded by the ImPACT Program of Council for Science, Technology and Innovation (Cabinet Office, Government of Japan) and in part by research funds from the Yamagata Prefectural Government and Tsuruoka City, Japan.

ORCID

Nobuaki Kono https://orcid.org/0000-0001-5960-7956

Kazuharu Arakawa https://orcid.org/0000-0002-2893-4919

REFERENCES

Alonge, M., Soyk, S., Ramakrishnan, S., & Wang, X. (2019). Fast and accurate reference-guided scaffolding of draft genomes. *bioRxiv*.

Ameur, A., Kloosterman, W. P., & Hestand, M. S. (2019). Single-molecule sequencing: Towards clinical applications. *Trends in Biotechnology*, *37*, 72–85. https://doi.org/10.1016/j.tibtech.2018.07.013

Ashton, P. M., Nair, S., Dallman, T., Rubino, S., Rabsch, W., Mwaigwisya, S., … O'Grady, J. (2015). MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nature Biotechnology*, *33*, 296–300. https://doi.org/10.1038/nbt.3103

Babb, P. L., Lahens, N. F., Correa-Garhwal, S. M., Nicholson, D. N., Kim, E. J., Hogenesch, J. B., … Voight, B. F. (2017). The *Nephila clavipes* genome highlights the diversity of spider silk genes and their complex expression. *Nature Genetics*, *49*, 895–903. https://doi.org/10.1038/ng.3852

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*, 455–477. https://doi.org/10.1089/cmb.2012.0021

Blamires, S. J., Blackledge, T. A., & Tso, I. M. (2017). Physicochemical property variation in spider silk: Ecology, evolution, and synthetic production. *Annual Review of Entomology*, *62*, 443–460. https://doi.org/10.1146/annurev-ento-031616-035615

Bolisetty, M. T., Rajadinakaran, G., & Graveley, B. R. (2015). Determining exon connectivity in complex mRNAs by nanopore sequencing. *Genome Biology*, *16*, 204. https://doi.org/10.1186/s13059-015-0777-z

Boza, V., Brejova, B., & Vinar, T. (2017). DeepNano: Deep recurrent neural networks for base calling in MinION nanopore reads. *PLoS ONE*, *12*, e0178751. https://doi.org/10.1371/journal.pone.0178751

Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., … Schloss, J. A. (2008). The potential and challenges of nanopore sequencing. *Nature Biotechnology*, *26*, 1146–1153. https://doi.org/10.1038/nbt.1495

Brown, C. G., & Clarke, J. (2016). Nanopore development at Oxford Nanopore. *Nature Biotechnology*, *34*, 810–811. https://doi.org/10.1038/nbt.3622

Byrne, A., Beaudin, A. E., Olsen, H. E., Jain, M., Cole, C., Palmer, T., … Vollmers, C. (2017). Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nature Communications*, *8*, 16027. https://doi.org/10.1038/ncomms16027

Castro-Wallace, S. L., Chiu, C. Y., John, K. K., Stahl, S. E., Rubins, K. H., McIntyre, A. B. R., … Turner, D. J. (2017). Nanopore DNA sequencing and genome assembly on the international space station. *Scientific Reports*, *7*, 18022. https://doi.org/10.1038/s41598-017-18364-0

Chen, G., Liu, X., Zhang, Y., Lin, S., Yang, Z., Johansson, J., … Meng, Q. (2012). Full-length minor ampullate spidroin gene sequence. *PLoS ONE*, *7*, e52293. https://doi.org/10.1371/journal.pone.0052293

Chin, C. S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., … Ecker, J. R. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, *13*, 1050–1054. https://doi.org/10.1038/nmeth.4035

Clarke, J., Wu, H. C., Jayasinghe, L., Patel, A., Reid, S., & Bayley, H. (2009). Continuous base identification for single-molecule nanopore

DNA sequencing. *Nature Nanotechnology*, *4*, 265–270. https://doi.org/10.1038/nnano.2009.12

Cook, D. E., Valle-Inclan, J. E., Pajoro, A., Rovenich, H., Thomma, B., & Faino, L. (2019). Long-read annotation: Automated eukaryotic genome annotation based on long-read cDNA sequencing. *Plant Physiology*, *179*, 38–54. https://doi.org/10.1104/pp.18.00848

Cretu Stancu, M., Van Roosmalen, M. J., Renkens, I., Nieboer, M. M., Middelkamp, S., de Ligt, J., … Espejo, Valle.-Inclan. J. (2017). Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nature Communications*, *8*, 1326. https://doi.org/10.1038/s41467-017-01343-4

Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nature Biotechnology*, *34*, 518–524. https://doi.org/10.1038/nbt.3423

Depledge, D. P., Kalanghad Puthankalam, S., Sadaoka, T., Bready, D., Mori, Y., Placantonakis, D. G., … Wilson, A. C. (2018). Native RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *bioRxiv*.

Ebler, J., Haukness, M., Pesout, T., Marschall, T., & Paten, B. (2018). Haplotype-aware genotyping from noisy long reads. *bioRxiv*.

Edwards, A., Debbonaire, A. R., Sattler, B., Mur, L. A. J., & Hodson, A. J. (2016). Extreme metagenomics using nanopore DNA sequencing: A field report from Svalbard, 78 °N. *bioRxiv*.

Edwards, A., Soares, A., Rassner, S., Green, P., Felix, J., & Mitchell, A. (2017). Deep sequencing: Intra-terrestrial metagenomics illustrates the potential of off-grid nanopore DNA sequencing. *bioRxiv*.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., … Dalal, R. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, *323*, 133–138. https://doi.org/10.1126/science.1162986

Faria, N. R., Sabino, E. C., Nunes, M. R., Alcantara, L. C., Loman, N. J., & Pybus, O. G. (2016). Mobile real-time surveillance of Zika virus in Brazil. *Genome Medicine*, *8*, 97. https://doi.org/10.1186/s13073-016-0356-2

Feng, Y., Zhang, Y., Ying, C., Wang, D., & Du, C. (2015). Nanopore-based fourth-generation DNA sequencing technology. *Genomics Proteomics Bioinformatics*, *13*, 4–16. https://doi.org/10.1016/j.gpb.2015.01.009

Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., … Serra, S. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods*, *15*, 201–206. https://doi.org/10.1038/nmeth.4577

Gautier, M., Yamaguchi, J., Foucaud, J., Loiseau, A., Ausset, A., Facon, B., … Lopez-Roques, C. (2018). The genomic basis of color pattern polymorphism in the Harlequin Ladybird. *Current Biology*, *28*(3296–3302), e3297.

Giordano, F., Aigrain, L., Quail, M. A., Coupland, P., Bonfield, J. K., Davies, R. M., … Ning, Z. (2017). De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms. *Scientific Reports*, *7*, 3935. https://doi.org/10.1038/s41598-017-03996-z

Goodwin, S., Mcpherson, J. D., & Mccombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*, 333–351. https://doi.org/10.1038/nrg.2016.49

Goyal, P., Krasteva, P. V., Van Gerven, N., Gubellini, F., Van den Broeck, I., Troupiotis-Tsaïlaki, A., … Chapman, M. R. (2014). Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature*, *516*, 250–253. https://doi.org/10.1038/nature13768

Greninger, A. L., Naccache, S. N., Federman, S., Yu, G., Mbala, P., Bres, V., … Mulembakani, P. (2015). Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Medicine*, *7*, 99. https://doi.org/10.1186/s13073-015-0220-9

Hayashi, C. Y., Blackledge, T. A., & Lewis, R. V. (2004). Molecular and mechanical characterization of aciniform silk: Uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. *Molecular Biology and Evolution*, *21*, 1950–1959. https://doi.org/10.1093/molbev/msh204

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*, 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

Henley, R. Y., Ashcroft, B. A., Farrell, I., Cooperman, B. S., Lindsay, S. M., & Wanunu, M. (2016). Electrophoretic deformation of individual transfer RNA molecules reveals their identity. *Nano Letters*, *16*, 138–144. https://doi.org/10.1021/acs.nanolett.5b03331

Hoang, P. N. T., Michael, T. P., Gilbert, S., Chu, P., Motley, S. T., Appenroth, K. J., … Lam, E. (2018). Generating a high-confidence reference genome map of the Greater Duckweed by integration of cytogenomic, optical mapping, and Oxford Nanopore technologies. *Plant Journal*, *96*, 670–684. https://doi.org/10.1111/tpj.14049

Hrdlickova, R., Toloue, M., & Tian, B. (2017). RNA-Seq methods for transcriptome analysis. *Wiley Interdisciplinary Reviews: RNA*, *8*(1). https://doi.org/10.1002/wrna.1364

Ii, K. M., Kono, N., Paulino-Lima, I. G., Tomita, M., Rothschild, L. J., & Arakawa, K. (2019). Complete genome sequence of Arthrobacter sp. Strain MN05-02, a UV-Resistant Bacterium from a manganese deposit in the Sonoran Desert. *Journal of Genomics*, *7*, 18–25. https://doi.org/10.7150/jgen.32194

Imai, K., Tarumoto, N., Misawa, K., Runtuwene, L. R., Sakai, J., Hayashida, K., … Suzuki, Y. (2017). A novel diagnostic method for malaria using loop-mediated isothermal amplification (LAMP) and MinION nanopore sequencer. *BMC Infectious Diseases*, *17*, 621. https://doi.org/10.1186/s12879-017-2718-9

Istace, B., Friedrich, A., D'agata, L., Faye, S., Payen, E., Beluche, O., … Engelen, S. (2017). de novo assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience*, *6*, 1–13.

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., … Dilthey, A. T. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, *36*, 338–345. https://doi.org/10.1038/nbt.4060

Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biology*, *17*, 239. https://doi.org/10.1186/s13059-016-1103-0

Jansen, H. J., Liem, M., Jong-Raadsen, S. A., Dufour, S., Weltzien, F. A., Swinkels, W., … Palstra, A. P. (2017). Rapid de novo assembly of the European eel genome from nanopore sequencing reads. *Scientific Reports*, *7*, 7213. https://doi.org/10.1038/s41598-017-07650-6

Johnson, S. S., Zaikova, E., Goerlitz, D. S., Bai, Y., & Tighe, S. W. (2017). Real-Time DNA sequencing in the Antarctic dry valleys using the Oxford Nanopore Sequencer. *Journal of Biomolecular Techniques*, *28*, 2–7.

Kamath, G. M., Shomorony, I., Xia, F., Courtade, T. A., & Tse, D. N. (2017). HINGE: Long-read assembly achieves optimal repeat resolution. *Genome Research*, *27*, 747–756. https://doi.org/10.1101/gr.216465.116

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*, 772–780. https://doi.org/10.1093/molbev/mst010

Kielbasa, S. M., Wan, R., Sato, K., Horton, P., & Frith, M. C. (2011). Adaptive seeds tame genomic sequence comparison. *Genome Research*, *21*, 487–493. https://doi.org/10.1101/gr.113985.110

Kono, N., Nakamura, H., Ohtoshi, R., Tomita, M., Numata, K., & Arakawa, K. (2019). The bagworm genome reveals a unique fibroin gene that provides high tensile strength. *Communications Biology*. in press.

Kono, N., Nakamura, H., Ohtoshi, R., Pedrazzoli Moran, D.A., Shinohara, A., Yoshida, Y., … Arakawa, K. (2019). Orb-weaving spider *Araneus ventricosus* genome elucidates the spidroin gene catalogue. *Scientific Reports*. in press.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via

adaptive k-mer weighting and repeat separation. *Genome Research*, *27*, 722–736. https://doi.org/10.1101/gr.215087.116

Krizanovic, K., Echchiki, A., Roux, J., & Sikic, M. (2018). Evaluation of tools for long read RNA-seq splice-aware alignment. *Bioinformatics*, *34*, 748–754. https://doi.org/10.1093/bioinformatics/btx668

Lang, K., Surendranath, V., Quenzel, P., Schofl, G., Schmidt, A. H., & Lange, V. (2018). Full-Length HLA Class I genotyping with the MinION nanopore sequencer. *Methods in Molecular Biology*, *1802*, 155–162. https://doi.org/10.1007/978-1-4939-8546-3

Li, H. (2016). Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*, *32*, 2103–2110. https://doi.org/10.1093/bioinformatics/btw152

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Lin, Y., Yuan, J., Kolmogorov, M., Shen, M. W., Chaisson, M., & Pevzner, P. A. (2016). Assembly of long error-prone reads using de Bruijn graphs. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, E8396–E8405. https://doi.org/10.1073/pnas.1604560113

Loman, N. J., Quick, J., & Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods*, *12*, 733–735. https://doi.org/10.1038/nmeth.3444

Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., … Kleihues, P. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A summary. *Acta Neuropathologica*, *131*, 803–820. https://doi.org/10.1007/s00401-016-1545-1

Lu, H., Giordano, F., & Ning, Z. (2016). Oxford nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics*, *14*, 265–279. https://doi.org/10.1016/j.gpb.2016.05.004

Mardis, E. R. (2013). Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*, *6*, 287–303. https://doi.org/10.1146/annurev-anchem-062012-092628

Miller, J. R., Delcher, A. L., Koren, S., Venter, E., Walenz, B. P., Brownley, A., … Sutton, G. (2008). Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics*, *24*, 2818–2824. https://doi.org/10.1093/bioinformatics/btn548

Miller, D. E., Staber, C., Zeitlinger, J., & Hawley, R. S. (2018). Highly contiguous genome assemblies of 15 *Drosophila* species generated using nanopore sequencing. *G3 (Bethesda)*, *8*, 3131–3141. https://doi.org/10.1534/g3.118.200160

Mitsuhashi, S., Kryukov, K., Nakagawa, S., Takeuchi, J. S., Shiraishi, Y., Asano, K., & Imanishi, T. (2017). A portable system for rapid bacterial composition analysis using a nanopore-based sequencer and laptop computer. *Scientific Reports*, *7*, 5657. https://doi.org/10.1038/s41598-017-05772-5

Mitsuhashi, S., Nakagawa, S., Takahashi Ueda, M., Imanishi, T., Frith, M. C., & Mitsuhashi, H. (2017). Nanopore-based single molecule sequencing of the D4Z4 array responsible for facioscapulohumeral muscular dystrophy. *Scientific Reports*, *7*, 14789. https://doi.org/10.1038/s41598-017-13712-6

Murray, M. G., & Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*, *8*, 4321–4325. https://doi.org/10.1093/nar/8.19.4321

Norris, A. L., Workman, R. E., Fan, Y., Eshleman, J. R., & Timp, W. (2016). Nanopore sequencing detects structural variants in cancer. *Cancer Biology & Therapy*, *17*, 246–253. https://doi.org/10.1080/15384047.2016.1139236

Omenetto, F. G., & Kaplan, D. L. (2010). New opportunities for an ancient material. *Science*, *329*, 528–531. https://doi.org/10.1126/science.1188936

Patel, A., Belykh, E., Miller, E. J., George, L. L., Martirosyan, N. L., Byvaltsev, V. A., & Preul, M. C. (2018). MinION rapid sequencing: Review of potential applications in neurosurgery. *Surgical Neurology International*, *9*, 157.

Payne, A., Holmes, N., Rakyan, V., & Loose, M. (2018). BulkVis: A graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/bty841

Quick, J., Ashton, P., Calus, S., Chatt, C., Gossain, S., Hawker, J., … Nye, K. (2015). Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella. *Genome Biology*, *16*, 114. https://doi.org/10.1186/s13059-015-0677-2

Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., … Beutler, N. A. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature Protocols*, *12*, 1261–1276. https://doi.org/10.1038/nprot.2017.066

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., … Boettcher, J. P. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, *530*, 228–232. https://doi.org/10.1038/nature16996

Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akeson, M., & Paten, B. (2017). Mapping DNA methylation with high-throughput nanopore sequencing. *Nature Methods*, *14*, 411–413. https://doi.org/10.1038/nmeth.4189

Rang, F. J., Kloosterman, W. P., & De Ridder, J. (2018). From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biology*, *19*, 90. https://doi.org/10.1186/s13059-018-1462-9

Ricker, N., Qian, H., & Fulthorpe, R. R. (2012). The limitations of draft assemblies for understanding prokaryotic adaptation and evolution. *Genomics*, *100*, 167–175. https://doi.org/10.1016/j.ygeno.2012.06.009

Ruan, J., & Li, H. (2019). Fast and accurate long-read assembly with wtdbg2. *bioRxiv*.

Runtuwene, L. R., Tuda, J. S. B., Mongan, A. E., Makalowski, W., Frith, M. C., Imwong, M., … Maeda, R. (2018). Nanopore sequencing of drug-resistance-associated genes in malaria parasites, *Plasmodium falciparum*. *Sci. Rep.*, *8*, 8286. https://doi.org/10.1038/s41598-018-26334-3

Russell, J. J., Theriot, J. A., Sood, P., Marshall, W. F., Landweber, L. F., Fritz-Laylin, L., … Umen, J. (2017). Non-model model organisms. *BMC Biology*, *15*, 55. https://doi.org/10.1186/s12915-017-0391-5

Sambrook, J., & Russell, D. W. (2001). *Molecular cloning : A laboratory manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Schadt, E. E., Turner, S., & Kasarskis, A. (2010). A window into third-generation sequencing. *Human Molecular Genetics*, *19*, R227–R240. https://doi.org/10.1093/hmg/ddq416

Schell, M. A., Karmirantzou, M., Snel, B., Vilanova, D., Berger, B., Pessi, G., … Arigoni, F. (2002). The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 14422–14427. https://doi.org/10.1073/pnas.212527599

Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., … Zipursky, S. L. (2000). Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell*, *101*, 671–684. https://doi.org/10.1016/s0092-8674(00)80878-8

Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., & Schatz, M. C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. *Nature Methods*, *15*, 461–468. https://doi.org/10.1038/s41592-018-0001-7

Shabardina, V., Kischka, T., Manske, F., Grundmann, N., Frith, M. C., Suzuki, Y., & Makałowski, W. (2019). NanoPipe-a web server for nanopore MinION sequencing data analysis. *Gigascience*, *8*. https://doi.org/10.1093/gigascience/giy169

Shao, H., Ganesamoorthy, D., Duarte, T., Cao, M. D., Hoggart, C. J., & Coin, L. J. M. (2018). npInv: Accurate detection and genotyping of

inversions using long read sub-alignment. *BMC Bioinformatics*, *19*, 261. https://doi.org/10.1186/s12859-018-2252-9

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., & Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods*, *14*, 407–410. https://doi.org/10.1038/nmeth.4184

Smith, A. M., Abu-Shumays, R., Akeson, M., & Bernick, D. L. (2015). Capture, unfolding, and detection of individual tRNA molecules using a nanopore device. *Frontiers in Bioengineering and Biotechnology*, *3*, 91.

Sohn, J. I., & Nam, J. W. (2018). The present and future of de novo whole-genome assembly. *Briefings in Bioinformatics*, *19*, 23–40.

Stoiber, M., & Brown, J. (2017). BasecRAWller: Streaming nanopore basecalling directly from raw signal. *bioRxiv*.

Tan, M. H., Austin, C. M., Hammer, M. P., Lee, Y. P., Croft, L. J., & Gan, H. M. (2018). Finding Nemo: Hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (*Amphiprion ocellaris*) genome assembly. *Gigascience*, *7*, 1–6.

Teng, H., Cao, M. D., Hall, M. B., Duarte, T., Wang, S., & Coin, L. J. M. (2018). Chiron: Translating nanopore raw signal directly into nucleotide sequence using deep learning. *Gigascience*, *7*. https://doi.org/10.1093/gigascience/giy037

Tyson, J. R., O'neil, N. J., Jain, M., Olsen, H. E., Hieter, P., & Snutch, T. P. (2018). MinION-based long-read sequencing and assembly extends the *Caenorhabditis elegans* reference genome. *Genome Research*, *28*, 266–274. https://doi.org/10.1101/gr.221184.117

Vaser, R., Sovic, I., Nagarajan, N., & Sikic, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, *27*, 737–746. https://doi.org/10.1101/gr.214270.116

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., … Earl, A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, *9*, e112963. https://doi.org/10.1371/journal.pone.0112963

Weirather, J. L., De Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X. J., … Au, K. F. (2017). Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res*, *6*, 100. https://doi.org/10.12688/f1000research

Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology*, *13*, e1005595. https://doi.org/10.1371/journal.pcbi.1005595

Workman, R. E., Tang, A., Tang, P. S., Jain, M., Tyson, J. R., Zuzarte, P. C., Gilpatrick, T., … Timp, W. (2018). Nanopore native RNA sequencing of a human poly(A) transcriptome. *bioRxiv*.

Xu, M., & Lewis, R. V. (1990). Structure of a protein superfiber: Spider dragline silk. *Proceedings of the National Academy of Sciences of the United States of America*, *87*, 7120–7124. https://doi.org/10.1073/pnas.87.18.7120

Yamagishi, J., Runtuwene, L. R., Hayashida, K., Mongan, A. E., Thi, L. A. N., Thuy, L. N., … Frith, M. (2017). Serotyping dengue virus with isothermal amplification and a portable sequencer. *Scientific Reports*, *7*, 3510. https://doi.org/10.1038/s41598-017-03734-5

Yasodha, R., Vasudeva, R., Balakrishnan, S., Sakthi, A. R., Abel, N., Binai, N., … Dev, S. A. (2018). Draft genome of a high value tropical timber tree, Teak (*Tectona grandis* L. f): Insights into SSR diversity, phylogeny and conservation. *DNA Research*, *25*, 409–419. https://doi.org/10.1093/dnares/dsy013

Ye, C., Hill, C. M., Wu, S., Ruan, J., & Ma, Z. S. (2016). DBG2OLC: Efficient Assembly of Large Genomes Using Long Erroneous Reads of the Third Generation Sequencing Technologies. *Scientific Reports*, *6*, 31900. https://doi.org/10.1038/srep31900

Zimin, A. V., Marcais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, *29*, 2669–2677. https://doi.org/10.1093/bioinformatics/btt476

How to cite this article: Kono N, Arakawa K. Nanopore sequencing: Review of potential applications in functional genomics. *Develop Growth Differ*. 2019;61:316–326. https://doi.org/10.1111/dgd.12608