*Systematics and Biodiversity* (2016), **14(1)**: 1–8

Taylor & Francis
Taylor & Francis Group

## Perspective

# Third generation sequencing: technology and its potential impact on evolutionary biodiversity research

CHRISTOPH BLEIDORN

Molecular Evolution and Systematics of Animals, Institute for Biology, University of Leipzig, Talstraße 33, D-04103 Leipzig, Germany
German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

(Received 28 April 2015; accepted 21 August 2015)

Next generation sequencing transformed the field of evolutionary biology and high throughput sequencing platforms are routinely used in phylogenomic, population genomic or metagenomic studies. Here I review the recent technical advancements of third generation sequencing instruments, thereby covering nanopore sequencing and single molecule real-time (SMRT) sequencing. The output and error rates are compared with sequencing platforms of the second generation (454 pyrosequencing, Illumina and Ion Torrent). Third generation sequencers produce sequence reads in hitherto unprecedented lengths and will help to strongly increase the quality of genome assemblies. Moreover, the speed of sequencing and ease of sample preparation enables sequencing in the field. Even though the output and error rate of the new generation of sequencer remains to be improved, new possibilities for evolutionary research will open up in the near future by these new techniques.

**Key words**: assembly, evolution, genome, metagenomics, nanopore sequencing, next generation sequencing, single molecule real-time sequencing



## Introduction

DNA sequencing was dominated by Sanger's chain termination method (Sanger, Nicklen, & Coulson, 1977) for decades. A milestone in the use of DNA sequencing was the sequencing of the human genome, which was announced in 2001 by two competing consortia. The costs for the human genome sequencing project were estimated to reach US$3 billion (Collins, Morgan, & Patrinos, 2003). Subsequent projects exploring the diversity and functionality of the human genome triggered a much needed advancement in DNA sequencing methods. Massively parallelized high throughput approaches were developed, known as next generation sequencing (NGS) techniques (Glenn, 2011; Metzker, 2010). 454 pyrosequencing (Margulies et al., 2005) became widely used in evolutionary studies early on, as a high number of reads which were comparable in length to those from Sanger sequences (now up to 1000 bp) were the output of single runs. Even higher outputs of literally billions of reads per run are available from Illumina sequencers based on reversible terminator sequencing (Bentley et al., 2008). However, read lengths are much shorter in this technique, with the newest generation of Illumina machines now outputting high quality reads approaching 250–300 bp (http://www.illumina.com/systems/sequencing.html). The most recent of the second generation technologies to become available is Ion Torrents' Ion Proton sequencer (Rothberg et al., 2011). This technique, which in contrast to Sanger, 454 and Illumina does not rely on an optical approach, enables fast and cheap sequencing with machines that are less expensive to purchase.

Correspondence to: Christoph Bleidorn. E-mail: bleidorn@uni-leipzig.de

## NGS transformed biodiversity research

In 2015, 454 pyrosequencing is ready to be retired and will not be supported by the distributing company (Roche) after 2016. However, 454 sequences are still commonly used in amplicon-based metabarcoding studies (Petrosino, Highlander, Luna, Gibbs, & Versalovic, 2009; Yu et al., 2012). Illumina as the leader on the DNA sequencing market announced the sequencing of human genomes for less than $1000 (doi: 10.1038/nature.2014.14530). Phylogenomic studies using transcriptome sequencing, target enrichment strategies or whole genome sequencing usually rely on this technique (Gerth, Gansauge, Weigert, &

ISSN 1477-2000 print / 1478-0933 online

http://dx.doi.org/10.1080/14772000.2015.1099575

**Table 1.** Comparison of the output of selected sequencing platforms. Numbers are according to companies or recent publications.

| Platform | Sequencer | Costs sequencing platform | Reads per run/lane | Output per run/lane | Maximal read lengths[1] | Average run duration |
|---|---|---|---|---|---|---|
| Sanger | ABI 3730xl | $100,000 | 96 | 100 kbp | 1000 bp | 2−3 hours |
| 454 | GS FLX | $450,000 | 1,000,000 | 700 mpb | 1000 bp | 24 hours |
| Illumina | HiSeq 3000 | $750,000 | 300,000,000[2] | 150 gbp[3] | 250 bp | 4 days |
| Illumina | NextSeq500 | $250,000 | 400,000,000 | 120 gbp[3] | 150 bp | 30 hours |
| Illumina | MiSeq | $100,000 | 25,000,000 | 15 gbp[3] | 300 bp | 24 hours |
| Ion Torrent | Proton II | $224,000 | 330,000,000 | 66 gbp | 200 bp | 4 hours |
| Ion Torrent | PGM 318 | $50,000 | 5,000,000 | 2 gbp | 400 bp | 7 hours |
| PacBio | RS II | $700,000 | 50,000 | 400 mbp | 54 kbp | 3 hours |
| Nanopore | MinION | $1,000 | 80,000[4] | 490 mbp[4] | 150 kbp | n.a.[4] |

[1]Estimated for high quality reads, individual reads could be longer.
[2]Single read run on one lane, capable of paired-end runs.
[3]Output for paired-end runs (in case of HiSeq a single lane).
[4]Machine run time is usually adjusted to need of sequencing depth, example given for an 48 hours run.



Bleidorn, 2014; Johnson, Walden, & Robertson, 2013; Lemmon, Emme, & Lemmon, 2012; Misof et al., 2014; Moroz et al., 2014; Weitemier et al., 2014). Likewise, affordable genome sequencing for non-model organisms helped to transform the field of population genetics to population genomics (Ellegren, 2014). Moreover, sequencing strategies reducing genome complexity as established in various RADseq-approaches provide an efficient method for genotyping large sample sizes (Andrews et al., 2014; Puritz et al., 2014). Adapted protocols for Ion Torrents' Ion Proton sequencer make such studies even more attractive for smaller labs, due to faster handling and less expensive (see Table 1) sequencing machines (Recknagel, Jacobs, Herzyk, & Elmer, 2015).

## Third generation sequencing

A third generation of sequencing techniques recently became available either commercially or at least for selected beta-tester. Sequencers distributed by Pacific Biosystems (PacBio) use Single Molecule Real-Time (SMRT) sequencing (Eid et al., 2009), whereas Oxford Nanopore Techniques (ONT) developed a device for nanopore sequencing (Branton et al., 2008). In contrast to second generation techniques (454, Illumina, Ion Torrent) these methods do not include an amplification step during sequencing library preparation, therefore enabling single molecule sequencing. Moreover, the expected read lengths are much higher than those of second generation techniques, with average read lengths exceeding 6−8 kbp and maximal read lengths exceeding 30−150 kbp (see Table 1 for a comparison of sequencing platforms).

Single Molecule Real-Time (SMRT) sequencing is based on monitoring polymerase activity while incorporating differently labelled nucleotides into the DNA strand (Eid et al., 2009). Each nucleotide carries a base-specific fluorescent label on its phosphate group, which is released when being incorporated by the polymerase. Incorporated nucleotides are detected by real-time imaging during strand synthesis. This basically means that the incorporation of nucleotides by a polymerase is filmed. The whole process takes place in a small well which is surrounded by aluminium walls called Zero-mode waveguide (ZMW). Single DNA polymerase molecules are attached on the surface of these wells, where their activity can be monitored. With a diameter of 70 nm and a depth of 100 nm these wells are extremely small and ∼150,000 ZMWs are comprised on so-called SMRT-cells for sequencing. For the sequencing process, fluorescent labelled nucleotides are flooded into these small cavities and their presence while floating in and out is measured as background noise. However, whenever a labelled nucleotide is associated with the template DNA in the polymerase active site a pulse of fluorescence intensity can be recorded for the corresponding dye (Fig. 1, step 2). Such an emission of light lasts some milliseconds and will be detected by the ZMW sensor. After incorporation, the fluorescent label is cleaved away by the normal activity of the polymerase, leading to a diffusion of the dye in the background and subsequent drop of the emission signal (Fig. 1, step 3). Now the next nucleotide can be incorporated and the emission signal of its dye can be recorded (Fig. 1, step 4). With this technique approximately 2−4 nucleotides are synthesized per second. The emission spectra not only help to unravel the DNA sequence of the template molecule, but also reveal possible epigenetic modifications due to subtle differences in emission patterns (Flusberg et al., 2010). The sequencing process is rather fast and takes about 4 hours per SMRT-cell. Reads are considerably longer than those of second generation sequencers, averaging more than 10 kbp, including reads as long as 54 kbp (Lee

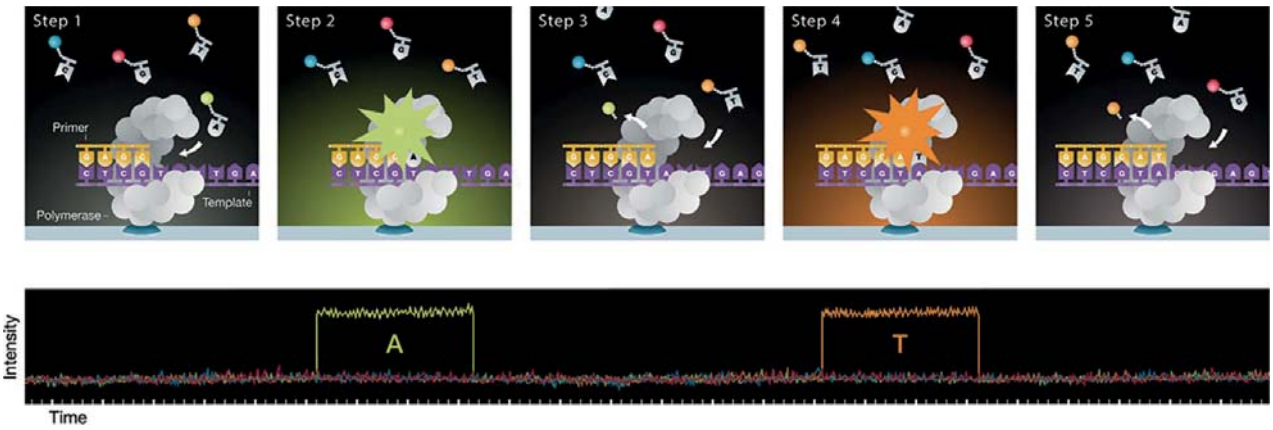


**Fig. 1.** In Single Molecule Real-Time (SMRT) sequencing the emission spectra of fluorescent labelled nucleotides are detected while being incorporated by the polymerase. Reprinted by permission from Pacific Biosciences.



et al., 2014). However, a caveat of the technique is the high error rate, which is estimated to be up to 20% (Hackl, Hedrich, Schultz, & Förster, 2014). The output per SMRT-cell is, at ∼400 mbp, also relatively small (Table 1).

With the MinION, ONT have developed the first available device for nanopore sequencing. These sequencing devices are the size of a small cell phone and can be plugged into the USB of a laptop (Fig. 2), making them ideal for 'sequencing to go'. The first MinIONs were distributed to selected laboratories for beta-testing in early 2014. The principle of the underlying technique is based on biological nanopores. Some toxins, such as the staphylococcal $\alpha$-hemolysin protein, are able to form a pore within a phospholipid bilayer, thereby creating a small channel measuring a few nm in diameter (Feng, Zhang, Ying, Wang, & Du, 2015). When placing the bilayer in a salt solution, electrodes can be used to form an ionic gradient. Negatively charged DNA can be forced by this gradient to pass the nanopore and each nucleotide passing the pore characteristically decreases the amplitude of the ion current, which can be detected by a sensor (Branton et al., 2008; Clarke et al., 2009). The MinION sequencing device is equipped with 512 channels containing nanopores, each detecting ∼10 bp per second. The sequencing is conducted by a method called strand sequencing. For sequencing library preparation the ends of each DNA molecule are modified and a hairpin adapter is ligated to one end of the molecule, while a motor protein is ligated to the other (Goodwin et al., in press). The motor protein ratchets the DNA molecule through the nanopore, by which it becomes single-stranded (Fig. 3). However, single strands of one molecule are not separated, as they are

**Fig. 2.** ONTs MinION sequencing device attached to a laptop computer. With permission of Oxford Nanopore Techniques.

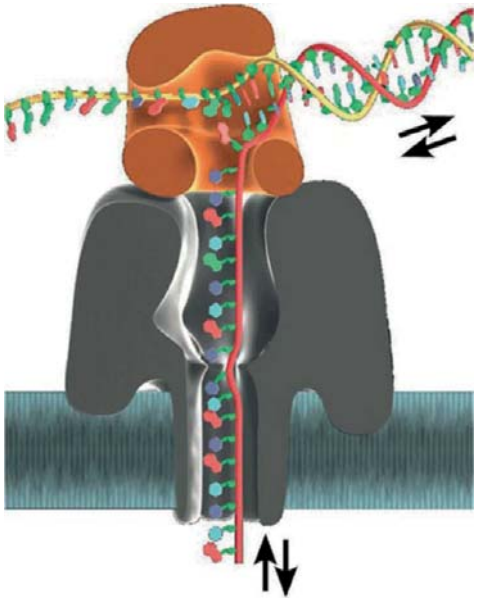

**Fig. 3.** Ratcheting of a DNA strand through a biological nanopore. Reprinted by permission from Macmillan Publishers Ltd: Nature Biotechnology (Schneider & Dekker, 2012), copyright (2012).

connected by the hairpin-adapter. In the ideal case, one strand passes through the nanopore, followed by the hairpin-adaptor and the second strand of the molecule. By this it is possible to generate a consensus sequence (or 2D-sequence) for the sequenced DNA double strand. Due to the length of the nanopore tunnel and the speed of the process, more than one nucleotide is always present in the pore at a time. Accordingly, usually the signal of overlapping 5-mers is recorded, which means that the cloud-based base-calling software MinKNOW needs to distinguish $4^5$ (1024) possible ionic current states for all possible 5-mers to generate the raw sequence. Not surprisingly a high error rate is reported for all reads produced by this technique so far, ranging from 25% to 40%. The output of the published studies ranged from 90 to 490 mbp per 48 hours, with average read lengths around 6 kbp and maximum read lengths of up to 150 kbp (Ashton et al., 2015; Goodwin et al., in press; Laver et al., 2015; Quick, Quinlan, & Loman, 2014). A benchmark study of the MinION Analysis and Reference Consortium reported that a typical experiment yields ~20,000 2D-reads comprising 115 mbp with an error of ~12% (Ip et al., 2015). A successor for the MinION sequencing device, the MkII, was announced at the 'London Calling 2015' conference organized by ONT (http://events.nanoporetech.com/events/london-calling-2015). This device should be available in 2016. Expected improvements include a higher output and a lower error rate. The exact pricing scheme has still to be announced. However, given that such devices may be used for $20 an hour it will surely have a gigantic impact on the field of evolutionary biodiversity research.

## Outlook

Third regeneration sequencing immediately offers three key advantages: it generates long reads, it is fast, and it is easy. The availability of long reads will have a major impact on evolutionary studies involving assembly or barcoding approaches. Assembling genomes solely based on short reads without any available reference genome remains a challenge (Koren & Phillippy, 2015). *De novo* assembly is usually conducted by *k*-mer assemblers (Miller, Koren, & Sutton, 2010). For these methods sequence reads are fragmented into even smaller pieces, so-called *k*-mers, which usually range between 21 and 101. However, if genomes bear repetitive regions exceeding the length of a *k*-mer it becomes impossible to reconstruct them correctly, leading to highly fragmented assemblies. Long reads (e.g., >8000 bp) allow significant improvements of genome assemblies (Koren & Phillippy, 2015). Instead of using k-mer assemblies, long reads can be assembled with an overlap-layout-consensus approach (Miller et al., 2010). Such methods were also used for Sanger sequence reads, e.g., for the initial assembly of the human genome. Long sequencing reads are overlapped and this information is subsequently used for error correction. Corrected overlapping reads are then connected in a graph, where nodes represent the sequence reads and edges correspond to the overlap. Based on this information reads are laid out and consensus sequences are created. The potential of SMRT long reads has recently been demonstrated as they are powerful in resolving long repeat regions and might soon become the gold standard in prokaryote genome sequencing (Chin et al., 2013). The resulting assemblies were comparable in quality and contiguity to assemblies from Sanger sequencing. Moreover, SMRT long reads have been used to resolve complex genomic regions of chimp and human (Chaisson et al., 2015; Huddleston et al., 2014). Initial results from nanopore sequencing also show that it is possible to reconstruct a bacterial genome sequence solely with this technique (Loman, Quick, & Simpson, 2015). Hybrid assembly strategies have been developed to overcome the high error rates of current nanopore derived sequences. Long nanopore reads were error corrected using high accuracy Illumina reads and subsequently assembled (Madoui et al., 2015). Currently, massive sequencing of genomes from organisms all over the tree of life is envisaged by different consortia, many of them driven by research questions in the field of evolution (e.g., GIGA Consortium of Scientists, 2014; Grigoriev et al., 2014; i5K Consortium, 2013; Koepfli, Paten, & O'Brien, 2015). In contrast to prokaryote (bacterial and archaeal) genomes, eukaryote genomes are often characterized by a high percentage of repetitive sequences (Brown, 2007), which are difficult to assemble with short reads. Long reads will significantly improve *de novo* genome assemblies for non-model organisms. This will allow the methodological gap between model and non-model species to be bridged, helping to elucidate the hidden biology of many organisms (Dunn, Leys, & Haddock, 2015).

Long sequence reads are not only desired for improving genome assemblies, but also for metabarcoding and metagenomic studies. In both cases automated identification of species from environmental DNA (e.g., soil, water, etc.), microbiomes, or from samples containing entire organisms is achieved (Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012). Sequenced DNA reads, usually *16S rRNA* (prokaryotes), *matK* or *rbcL* (plants), ribosomal ITS (fungi), or *cox1* (animals) amplicons, are compared with a reference database (e.g., Barcoding of Life Database, NCBI GenBank) for identification (Pawlowski et al., 2012). Longer reads will allow a more reliable assignment of DNA sequences to reference species or bacterial strains, as complete genes instead of less informative fragments could be targeted (Mitra, Schubach, & Huson, 2010). First studies using SMRT long reads for amplicon identification show that the longer reads improve the species assignment of individual

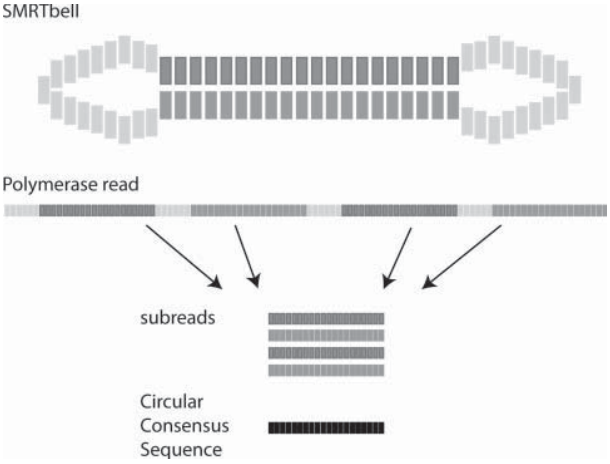


**Fig. 4.** Circular consensus sequencing (CCS). Sequencing libraries are constructed as SMRTbells, which consist of a double-stranded DNA fragment (separate strands in darker grey tones) which has single stranded hairpin-adapters (light grey) ligated to its ends. A primer site is within the hairpin-loop. Polymerase with strand-displacement activity can copy the strands several times, while sequence data are recorded. The resulting sequence consists of several copies of the sequence fragments (subreads), always interrupted by an adaptor sequence. As the sequencing error is randomly distributed, subreads can be used to create a high quality consensus sequence.

sequence reads, though the comparatively high error rate limits the usefulness of this approach as species richness will be overestimated (Mosher, Bernberg, Shevchenko, Kan, & Kaplan, 2013; Schloss, Westcott, Jenior, & Highlander, 2015). However, with SMRT sequencing it is possible to sequence the same molecule several times by circular consensus sequencing (CCS) (Fig. 4) (Travers, Chin, Rank, Eid, & Turner, 2010). For library preparation, sequence fragments are capped with single stranded hairpin-loops as adaptors, resulting in circular molecules. Starting from a primer site within the hairpin-loop, molecules can be sequenced multiple times. As the sequence error seems to be randomly distributed, these sequences are used to generate consensus sequences with an accuracy of 99%. Such sequences from CCS seem to be appropriate for metabarcoding studies (Fichot & Norman, 2013). Other applications for CCS include assembly free genotyping of transcripts from complex gene families. This has been demonstrated for MHC alleles from macaques and vomeronasal gene receptors of mouse lemurs (Larsen, Heilman, & Yoder, 2014; Westbrook et al., in press). Transcriptome analyses have become an important tool to understand the genomic basis of adaptations or to investigate phenotypic plasticity (Elmer & Meyer, 2011; Schlichting & Wund, 2014; Stapley et al., 2010). Third generation sequencing has the potential to eliminate the step of transcriptome sequence assembly in general (Martin & Wang, 2011).

Third generation sequencing is fast and sequences are retrieved in real-time. In the case of the MinION, raw data are analysed using cloud computing by internet access. MinION nanopore sequencing was used to genotype *Salmonella* strains from an outbreak in an English hospital (Quick et al., 2015). The time from DNA sequencing to strain typing was 6 hours, which included 1 hour of sequencing and two hours of data analysis. Similar results have been obtained for metagenomic analysis of human blood samples for an unbiased screen for viral pathogens (Greninger et al., 2015; Kilianski et al., 2015). Despite producing highly erroneous reads viral strains could be correctly assigned within less than one hour of sequencing. Nanopore sequencing is also easy to handle. The sequencer is the size of a small cell phone, making it highly portable. At the moment the bottleneck with the use of this approach is access to the internet. The library preparation is simple and takes around two hours. Using this setup it was possible to sequence Ebola strains in the field. The virus genomes of 14 patients in Guinea (West Africa) could be completely sequenced within 12 days, and in one case it took only 48 hours after sample collection (Hayden, 2015). All these studies benefit from the fact that the amount of sequencing needed can be controlled by parallel data analysis. That means, during sequencing it is possible to monitor if the desired sequences or sequence coverage are approached. Currently, cartridges for automatic sample preparation (Voltrax) are under development, which can be docked to the nanopore sequencing device (https://www.nanoporetech.com/products-services/voltrax). By using such a device the realization of the Star Trek inspired DNA barcoding tricoder would finally be possible. The speed of sequencing and ease of sample preparation will open up further interesting possibilities, e.g., monitoring of metagenomic communities over time, or the real-time detection of environmental DNA.

The comparatively high error rate and low output of third generation sequencers means that they cannot yet replace Illumina or Ion Torrent sequencing. For many evolutionary research questions in the field of population genomics, phylogenomics or metagenomics second generation techniques will remain state of the art for at least the next few years. It has been demonstrated for SMRT sequencing that methodological improvements strongly improved the quality and length of sequence reads (Mosher et al., 2014). A similar development is expected for nanopore sequencing. Besides handheld MinION sequence devices, ONT also announced the development of high throughput bench-top sequencers (PromethION). These machines will sequence with thousands of nanopores in parallel and output of several terabases per day is expected (http://blogs.nature.com/tradesecrets/2015/05/27/two-days-of-minion). A new sequencing revolution is underway and will further transform the field of evolutionary biodiversity research.

# Appendix

## Amplicon

An amplicon is a fragment of DNA or RNA which has been replicated by PCR. Usually specific genes for subsequent species or strain identification are targeted with next generation sequencing, e.g., *cox1* for animals, *rbcl* and *matk* for plants, ribosomal ITS for fungi, or *16s rRNA* for prokaryotes.

## Assembly

Assembly is the process of constructing contiguous sequences from sequence reads. A sequence derived from assembling several sequence reads is called contig. Three main types of assembly methods are currently in use: greedy, overlap-layout consensus methods and *k*-mer assemblies. *De novo* assemblies are those which do not rely on an available reference genome.

## Ion semiconductor sequencing

Ion semiconductor sequencing (as used in Ion Torrent sequencers) analyses changes of hydrogen ion concentration. Any time when a nucleotide is incorporated into a DNA strand by a polymerase, a hydrogen (or proton) is released. The release of these protons is measured in real-time by so-called ion-sensitive field-effective transistors. As in pyrosequencing, the wells where the sequencing takes place are flooded in cycles with one sort of nucleotide at a time. Detection of changes in pH allows inferring if and how many bases have been incorporated into a sequence read.

## *k*-mer

*k*-mers represent all possible substrings, of a given length, of a sequence. Every analysed sequence will be fragmented in all possible *k*-1 overlapping fragments of this size. For assembly, de Bruijn graphs are constructed, where nodes represent sequences (*k*-mers) which are connected by edges in case they show a *k*-1 overlap. Contigs are retrieved by traversing every edge of the graph exactly once (Euler path). *k*-mer assemblies are the standard for short read assemblies.

## Overlap-layout-consensus (OLC) assembly

The OLC assembly can be divided into three steps. In the first step pairwise alignments of all sequence reads are conducted. The information of overlapping reads is stored in graphs, the so-called overlap-graphs. The nodes in the graph represent sequence reads, whereas edges indicate which reads are connected by an overlap. The second step is the layout, where the relative position of sequence reads (nodes) of every overlap graph is determined and arranged accordingly into an alignment. This is conducted by searching for a mathematical path describing a way to go over every node exactly one time (Hamilton path). In the last step, the resulting alignments are used to determine consensus sequences which represent contigs.

## Paired-end

Sequencing of the same molecule from both ends is called paired-end sequencing. This can typically be conducted using the Illumina technique.

## Pyrosequencing

Pyrosequencing (as used in 454 sequencers) is a technique where nucleotides are released one after another and washed over the template DNA strand to be sequenced. A cascade of enzymatic reactions leads to the emission of detectable light signal which is in its strength proportional to the number of nucleotides incorporated during this step.

## RADseq

Restriction-site associated DNA sequencing (RADseq) is a method that sequences a reduced part of the target genome. Therefore, target DNA is sheared by the use of restriction enzymes and a size-selected subset of these fragments is later sequenced. Several modified RADseq protocols are in use and the choice of restriction enzymes allows controlling the number of fragments to be sequenced. RADseq is widely used for SNP (single nucleotide polymorphisms) discovery and genotyping in population genomic studies.

## Reversible terminator sequencing

Reversible terminator sequencing (as conducted by Illumina sequencing machines) takes place on a flow cell, where billions of sequences can be processed. Based on a sequencing-by-synthesis approach all four nucleotides are added simultaneously to the flow cell, together with polymerase. Every incorporated nucleotide is chemically blocked at its 3' OH-group and carries a removable fluorophore for identification by laser. After detection, the blocking and the fluorophore are removed and the process of sequencing is continued by incorporating the next nucleotide.

## Target enrichment

By target enrichment genomic regions are selectively captured from a DNA sample before sequencing. Enrichment

can be facilitated either by PCR, molecular inversion probes or hybridization to target specific probes. In evolutionary diversity research target enrichment is often used to generate datasets for phylogenomic studies or to enrich for complete mitochondrial genome sequencing.

## References


Andrews, K. R., Hohenlohe, P. A., Miller, M. R., Hand, B. K., Seeb, J. E., & Luikart, G. (2014). Trade-offs and utility of alternative RADseq methods: Reply to Puritz et al. *Molecular Ecology, 23*, 5943−5946.

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.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G.,... Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature, 456*, 53−59.

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.

Brown, T. (2007). *Genomes 3*. New York, USA: Garland Science Publisher.

Chaisson, M. J. P., Huddleston, J., Dennis, M. Y., Sudmant, P. H., Malig, M., Hormozdiari, F.,... Eichler, E. E. (2015). Resolving the complexity of the human genome using single-molecule sequencing. *Nature, 517*, 608−611.

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C.,... Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods, 10*, 563−569.

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.

Collins, F. S., Morgan, M., & Patrinos, A. (2003). The Human Genome Project: Lessons from Large-Scale Biology. *Science, 300*, 286−290.

Dunn, C. W., Leys, S. P., & Haddock, S. H. D. (2015). The hidden biology of sponges and ctenophores. *Trends in Ecology & Evolution, 30*, 282−291.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G.,... Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science, 323*, 133−138.

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution, 29*, 51−63.

Elmer, K. R., & Meyer, A. (2011). Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in Ecology & Evolution, 26*, 298−306.

Feng, Y., Zhang, Y., Ying, C., Wang, D., & Du, C. (2015). Nanopore-based Fourth-generation DNA Sequencing Technology. *Genomics, Proteomics & Bioinformatics, 13*, 4−16.

Fichot, E., & Norman, R. S. (2013). Microbial phylogenetic profiling with the Pacific Biosciences sequencing platform. *Microbiome, 1*, 10.

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A.,... Turner, S. W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods, 7*, 461−465.

Gerth, M., Gansauge, M. T., Weigert, A., & Bleidorn, C. (2014). Phylogenomic analyses uncover origin and spread of the *Wolbachia* pandemic. *Nature Communications, 5*, 5117.

GIGA Consortium of Scientists (2014). The Global Invertebrate Genomics Alliance (GIGA): developing community resources to study diverse invertebrate genomes. *Journal of Heredity, 105*, 1−18.

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources, 11*, 759−769.

Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M.C., & McCombie, W. R. (in press). Oxford nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Research*. doi:10.1101/gr.191395.115

Greninger, A. L., Naccache, S. N., Federman, S., Yu, G., Mbala, P., Bres, V.,... Chiu, C. Y. (2015). Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Medicine, 7*, 99.

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R.,... Shabalov, I. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research, 42*, D699−D704.

Hackl, T., Hedrich, R., Schultz, J., & Förster, F. (2014). *proovread*: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics, 30*, 3004−3011.

Hayden, E. (2015). Pint-sized DNA sequencer impresses first users. *Nature, 521*, 15−16.

Huddleston, J., Ranade, S., Malig, M., Antonacci, F., Chaisson, M., Hon, L.,... Eichler, E. E. (2014). Reconstructing complex regions of genomes using long-read sequencing technology. *Genome Research, 24*, 688−696.

i5k Consortium (2013). The i5K Initiative: Advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *Journal of Heredity, 104*, 595−600.

Ip, C. L. C., Loose, M., Tyson, J. R., de Cesare, M., Brown, B. L., Jain, M.,… MinION Analysis and Reference Consortium (2015) MinION Analysis and Reference Consortium: Phase 1 data release and analysis [version 1; referees: awaiting peer review]. *F1000 Research*, 4, 1075.

Johnson, K. P., Walden, K. K. O., & Robertson, H. M. (2013). Next-generation phylogenomics using a target restricted assembly method. *Molecular Phylogenetics and Evolution, 66*, 417−422.

Kilianski, A., Haas, J., Corriveau, E., Liem, A., Willis, K., Kadavy, D.,... Minot, S. (2015). Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. *GigaScience, 4*, 12.

Koepfli, K.-P., Paten, B., & O'Brien, S. J. (2015). The Genome 10K Project: A Way Forward. *Annual Review of Animal Biosciences, 3*, 57−111.

Koren, S., & Phillippy, A. M. (2015). One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Current Opinion in Microbiology, 23*, 110−120.

Larsen, P.A., Heilman, A.M., & Yoder, A.D. (2014). The utility of PacBio circular consensus sequencing for characterizing complex gene families in non-model organisms. *BioMed Central Genomics, 15*, 720.

Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., & Studholme, D. J. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomolecular Detection and Quantification, 3*, 1−8.

Lee, H., Gurtowski, J., Yoo, S., Marcus, S., McCombie, W. R., & Schatz, M. (2014). Error correction and assembly complexity of single molecule sequencing reads. *bioRxiv* doi. org/10.1101/006395.

Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology, 61*, 727−744.

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.

Madoui, M., Engelen, S., Cruaud, C., Belser, C., Bertrand, L., Alberti, A.,... Aury, J. (2015). Genome assembly using Nanopore-guided long and error-free DNA reads. *BioMed Central Genomics, 16*, 327.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A.,... Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature, 437*, 376−380.

Martin, J. A., & Wang, Z. (2011). Next-generation transcriptome assembly. Nature Reviews Genetics, 12, 671−682.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature Reviews Genetics, 11*, 31−46.

Miller, J., Koren, S., & Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics, 95*, 315.

Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C.,... Zhou, X. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science, 346*, 763−767.

Mitra, S., Schubach, M., & Huson, D. (2010). Short clones or long clones? A simulation study on the use of paired reads in metagenomics. *BioMed Central Bioinformatics, 11*, S12.

Moroz, L. L., Kocot, K. M., Citarella, M. R., Dosung, S., Norekian, T. P., Povolotskaya, I. S.,... Kohn, A. B. (2014). The ctenophore genome and the evolutionary origins of neural systems. *Nature, 510*, 109−114.

Mosher, J. J., Bernberg, E. L., Shevchenko, O., Kan, J., & Kaplan, L. A. (2013). Efficacy of a 3rd generation high-throughput sequencing platform for analyses of 16S rRNA genes from environmental samples. *Journal of Microbiological Methods, 95*, 175−181.

Mosher, J. J., Bowman, B., Bernberg, E. L., Shevchenko, O., Kan, J., Korlach, J., & Kaplan, L. A. (2014). Improved performance of the PacBio SMRT technology for *16S* rDNA sequencing. *Journal of Microbiological Methods*, 104, 59−60.

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C.,... de Vargas, C. (2012). CBOL Protist Working Group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *Public Library of Science Biology, 10*, e1001419.

Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A., & Versalovic, J. (2009). Metagenomic Pyrosequencing and Microbial Identification. *Clinical Chemistry, 55*, 856−866.

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., & Bird, C. E. (2014). Demystifying the RAD fad. *Molecular Ecology, 23*, 5937−5942.

Quick, J., Ashton, P., Calus, S., Chatt, C., Gossain, S., Hawker, J.,... Loman, N. (2015). Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biology, 16*, 114.

Quick, J., Quinlan, A., & Loman, N. (2014). A reference bacterial genome dataset generated on the MinIONTM portable single-molecule nanopore sequencer. *GigaScience, 3*, 22.

Recknagel, H., Jacobs, A., Herzyk, P., & Elmer, K. R. (2015). Double-digest RAD sequencing using Ion Proton semiconductor platform (ddRADseq-ion) with nonmodel organisms. *Molecular Ecology Resources*, *15*, 1316−1329.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M.,... Bustillo, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature, 475*, 348−352.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, 74*, 5463−5467.

Schlichting, C. D., & Wund, M. A. (2014). Phenotypic plasticity and epigenetic marking: an assessment of evidence for genetic accomodation *Evolution, 68*, 656−672.

Schloss, P., Westcott, S., Jenior, M., & Highlander, S. (2015). Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. *PeerJ PrePrints, 3*, e778v1. doi: https://dx.doi.org/10.7287/peerj.preprints.778v1.

Schneider, G. F., & Dekker, C. (2012). DNA sequencing with nanopores. *Nature Biotechnology, 30*, 326−328.

Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R.,... Slate, J. (2010). Adaptation genomics: the next generation. *Trends in Ecology & Evolution, 25*, 705−712.

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

Travers, K. J., Chin, C.-S., Rank, D. R., Eid, J. S., & Turner, S. W. (2010). A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Research, 38*, e159.

Weitemier, K., Straub, S. C. K., Cronn, R. C., Fishbein, M., Schmickl, R., McDonnell, A., & Liston, A. (2014). HybSeq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences, 2*, 1400042.

Westbrook, C. J., Karl, J. A., Wiseman, R. W., Mate, S., Koroleva, G., Garcia, K.,... Palacios, G. (in press). No assembly required: Full-length MHC class I allele discovery by PacBio circular consensus sequencing. *Human Immunology*, in press

Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution, 3*, 613−623.

**Associate Editor: Sven Buerki**