Journal of Biotechnology 136 (2008) 3–10

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Review

The Genome Sequencer FLX™ System—Longer reads, more applications, straight forward bioinformatics and more complete data sets

Marcus Droege[a,*], Brendon Hill[b]

[a] Roche Applied Science, Global Marketing, 82372 Penzberg, Germany
[b] 454 Life Sciences, Branford, USA

ARTICLE INFO

Article history:
Received 3 January 2008
Received in revised form 17 March 2008
Accepted 31 March 2008

Keywords:
2nd generation sequencing technology
Human re-sequencing
De novo sequencing
Transcriptome analysis
Microbial genome sequencing
Metagenomics
Plants sequencing
Viral variants

ABSTRACT

The Genome Sequencer FLX System (GS FLX), powered by 454 Sequencing, is a next-generation DNA sequencing technology featuring a unique mix of long reads, exceptional accuracy, and ultra-high throughput. It has been proven to be the most versatile of all currently available next-generation sequencing technologies, supporting many high-profile studies in over seven applications categories. GS FLX users have pursued innovative research in *de novo* sequencing, re-sequencing of whole genomes and target DNA regions, metagenomics, and RNA analysis. 454 Sequencing is a powerful tool for human genetics research, having recently re-sequenced the genome of an individual human, currently re-sequencing the complete human exome and targeted genomic regions using the NimbleGen sequence capture process, and detected low-frequency somatic mutations linked to cancer.

© 2008 Elsevier B.V. All rights reserved.

Contents

1. Introduction ... 4
2. The 454 Sequencing technology—an overview ... 4
3. Long reads, high throughput and superior single-read accuracy without filtering against reference sequences ... 4
   3.1. Average read length of 200–300 bases ... 4
   3.2. Single-read accuracy of more than 99.5%, substitution errors are exceedingly rare ... 4
   3.3. High throughput and cost-efficient split up of runs—up to 2304 samples per large run ... 6
4. Long reads and superior single-read accuracy ensures broadest application versatility ... 7
   4.1. *De novo* sequencing (e.g. plant BACs, microbes, viruses) ... 7
   4.2. Whole genome re-sequencing (e.g. targeted human genomic regions, structural variations) ... 7
   4.3. Amplicon sequencing (e.g. exon re-sequencing, virus variant detection, DNA methylation) ... 8
   4.4. Metagenomics and microbial diversity ... 8
   4.5. EST sequencing (e.g. transcriptome survey of organisms with unknown genomes) ... 8
   4.6. Full length/shotgun sequencing of the transcriptome ... 8
   4.7. ncRNAs (all classes of ncRNA, from miRNA to >200 nt transcripts of unknown function) ... 8
5. Bioinformatics without large investment in enterprise scale infrastructure and people ... 8
   5.1. GS Reference Mapper software ... 8
   5.2. GS *De Novo* Assembler software ... 8
   5.3. The Amplicon Variant Analyzer (AVA) ... 8
6. Recent breakthroughs achieved using the Genome Sequencer System ... 9
   6.1. Re-sequencing of the human genome ... 9
   6.2. Re-sequencing of the human exome and targeted gene regions ... 9

* Corresponding author. Tel.: +49 88566 06888.
E-mail address: Marcus.Droege@Roche.com (M. Droege).

0168-1656/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2008.03.021

6.3. Analysis of structural variations in the human genome 9
6.4. Studying the molecular basis for eusociality using expression profiling 9
6.5. Metagenomics and microbial diversity 9
6.6. Ancient DNA 9
7. Even longer reads, considerably higher throughput and significant reductions in the costs per base in 2008 9
References 9

1. Introduction

For the past 30 years, the Sanger sequencing process has been the standard method for sequencing DNA. Despite continued advances such as the introduction of capillary electrophoresis systems, and a continuing decrease in costs, this method has been shown to be prohibitively expensive and time consuming to perform routine high-throughput sequencing such as routine sequencing of the human genomes. Demand for faster, affordable DNA sequencing has led to the development of so-called "next-generation" sequencing technologies, which have the potential to sequence the human genome for several thousands of dollars in the coming years.

In October 2005, 454 Life Science, a member of the Roche group, was the first company to introduce such a next-generation sequencing system into the life science market. The 454 Sequencing technology has been enthusiastically adopted by researchers worldwide and has already been used to achieve several of the recent breakthrough discoveries in genomic research. This article provides an overview of the 454 Sequencing technology and summarizes some selected breakthroughs achieved using the system.

2. The 454 Sequencing technology—an overview

The 454 Sequencing System supports the analysis of samples from a wide variety of starting materials including genomic DNA, PCR products, BACs, and cDNA. Samples such as genomic DNA and BACs are fractionated into small, 300–800-basepair fragments using a mechanical sheering process (nebulization). For smaller samples, such as small non-coding RNA or PCR amplicons, fragmentation is not required. Using a series of standard molecular biology techniques (Fig. 1), short adaptors (A and B) are added to each fragment. The adaptors are used for purification, amplification, and sequencing steps. Single-stranded fragments (sstDNA) with A and B adaptors compose the sample library used for subsequent workflow steps.

During the following emulsion PCR procedure (Fig. 2) the sstDNA library is first mixed with an excess of sepharose beads carrying oligonucleotides complementary to, e.g. the B-adaptor sequence of the library fragments. As a result most of these beads carry a unique single-stranded DNA library fragment. The bead-bound library is then emulsified with amplification reagents in a water-in-oil mixture. Each bead is now captured within its own microreactor where clonal amplification of the single-stranded DNA fragments occurs. This results in bead-immobilized, clonally amplified DNA fragments (ca. 10 million identical DNA molecules per bead).

As preparation for the sequencing reaction, sstDNA library beads are added to the DNA Bead Incubation Mix (containing DNA polymerase) and are layered with Enzyme Beads (containing sulfurylase and luciferase) onto the 454 PicoTiterPlate™ device (Fig. 3). This 70 mm × 75 mm plate is an optical device containing 1.6 million wells at a diameter of 44 μm per well. Only one library bead (around 30 μm) fits into one well. The layer of Enzyme Beads ensures that the DNA beads remain positioned in the wells during the sequencing reaction. The bead-deposition process maximizes the number of wells that contain a single-amplified library bead (avoiding more than one sstDNA library bead per well).

The loaded PicoTiterPlate device is placed into the Genome Sequencer FLX™ Instrument (Fig. 4). The fluidics sub-system flows sequencing reagents (containing buffers and nucleotides) across the wells of the plate. Nucleotides are flowed sequentially in a fixed order across the PicoTiterPlate device during a sequencing run. During the nucleotide flow, hundreds of thousands of beads each carrying millions of copies of a unique single-stranded DNA molecule are sequenced in parallel. If a nucleotide complementary to the template strand is flowed into a well, the polymerase extends the existing DNA strand by adding nucleotide(s). Addition of one (or more) nucleotide(s) results in a reaction that generates a light signal that is recorded by the CCD camera in the instrument (Fig. 5). In a limited range, the signal strength is proportional to the number of nucleotides incorporated in a single-nucleotide flow (Fig. 6).

3. Long reads, high throughput and superior single-read accuracy without filtering against reference sequences

3.1. Average read length of 200–300 bases

Read length is one of the most important key factors in high-throughput sequencing. Gaps in the consensus sequence, caused by repeats, can be covered by longer reads, leading to a more comprehensive result. Furthermore, long reads enable haplotyping, the identification of low frequency viral quasi species, annotation of fragments isolated from the transcriptome, the more accurate quantification of microbial diversities and so forth. In fact, only the read lengths and single-read accuracies provided by conventional Sanger and the new 454 Sequencing technology allow high-quality *de novo* sequencing of genomes and transcriptomes. Fig. 7 shows the average read length generated using the Genome Sequencer FLX System.

3.2. Single-read accuracy of more than 99.5%, substitution errors are exceedingly rare

Compared to the Genome Sequencer 20 System, significant enhancements in the single-read accuracy were integrated into the Genome Sequencer FLX System (Fig. 8). Currently, single-read accuracies of >99.5% over the first 200 bases are typically achieved. Most notably, this error rate already includes small insertions and deletions caused by the presence of homopolymers. Substitution errors are exceedingly rare (down to 10^{-6}). The Genome Sequencer FLX™ System offers high-throughput sequencing with single-read accuracies equivalent, or better than traditional Sanger sequencing (99.5%). The vast majority of errors that make up the 0.5% single-read error rate are overcalling or undercalling the length of homopolymeric stretches. The magnitude of this error mode can be put into perspective with an example that comes from researchers at the University of Bielefeld (Tauch et al., 2006). A *Corynebacterium urealyticum* strain containing 451 oligonucleotides in the range of 6–13 nucleotides was sequenced using the 454 Sequencing technology (GS20). The lengths of 6 out of the 461 homopolymers were

Download English Version:

https://daneshyari.com/en/article/24554

Download Persian Version:

https://daneshyari.com/article/24554

Daneshyari.com