

MILESTONE 3

Sequencing — the next generation

Credit: Zoran Obradovic / Alamy Stock Photo



Following successful efforts to sequence bacterial, animal and human genomes throughout the 1990s using Sanger sequencing, investigators began looking for cheaper and faster approaches. Sanger sequencing was limited in speed and cost owing to its reliance on dideoxynucleotide (ddNTP) ‘terminators’ and the need to use electrophoresis, limiting sequencing to a single DNA fragment at a time.

In 1998, Pal Nyrén’s laboratory described a sequencing-by-synthesis method known as pyrosequencing. This method is based on the measurement of pyrophosphate — which is released following the addition of a nucleotide to a growing DNA strand by DNA polymerase — using a two-enzyme, luciferase-based system. By sequentially adding different deoxynucleotides (dNTPs), the sequence of a template DNA molecule can be inferred by detecting light released at the site of nucleotide incorporation, allowing real-time sequencing and avoiding lengthy electrophoresis.

In parallel with these advances, researchers were investigating how to increase the throughput of preparing templates for sequencing, specifically cloning and amplifying DNA. In 1999, Mitra and Church described a method for amplifying DNA by performing PCR in a polyacrylamide film, producing an array of PCR colonies, or ‘colonies’, consisting of

thousands of clonal amplification products localized with their respective templates. These arrays set the stage for systems capable of interrogating many clonal populations in parallel. A study by Brenner et al. in 2000 described a method for interrogating gene expression by cloning cDNA derived from *Saccharomyces cerevisiae* transcripts onto microbeads, distributing these beads in an array on a flow cell and sequencing the attached cDNA. This study used a sequencing-by-ligation method involving rounds of restriction enzymes and the addition of a library of fluorescent adaptors. In 2003, Dressman et al. improved on this bead system, attaching biotinylated PCR primers to streptavidin-coated beads before dispersing the beads in a microemulsion and carrying out PCR. This emulsion PCR (ePCR) method allowed spatial isolation of each template to a single bubble and enabled clonal amplification of the template on each bead.

In 2005, two ground-breaking studies built on these discoveries and described high-throughput methods for rapidly and cheaply sequencing a whole bacterial genome. Jay Shendure, Greg Porreca and colleagues working in George Church’s laboratory described a workflow based on ePCR and sequencing by ligation to sequence a tryptophan-deficient derivative of the *Escherichia coli* strain MG1655. Clonal

“making sequencing high-throughput and affordable”

amplification was achieved with the ePCR system of Dressman et al., and the beads then packed to the surface of a glass slide for sequencing. A library of ePCR-amplified DNA fragments was sequenced using a strategy involving fluorophore-tagged degenerate nonamers, the ligation of which to the template DNA and subsequent fluorescent signal depended on the complementarity of a single base. Mapping of reads to a MG1655 reference genome showed a high degree of accuracy. The estimated cost of US\$0.11 per kb of sequence generated was one-ninth of the cost of electrophoretic sequencing methods at that time. The high accuracy, low cost and speed of this method pointed to its potential as a high-throughput technique for resequencing organisms to interrogate genetic variation.

The second paper, by a group of researchers at 454 Life Sciences led by Jonathan Rothberg, introduced a high-throughput sequencing platform based on ePCR and pyrosequencing. Amplification and sequencing of the fragmented *Mycobacterium genitalium* genome was performed in wells containing a single fragment, bead and picolitre reaction volumes on a fibre-optic slide over which dNTPs were added in waves. The blinking light pattern produced as rounds of dNTPs passed over the beads was used to determine the sequence of the bacterial genome.

These massively parallel sequencing techniques set the stage for rapid, low-cost sequencing of genomes. Soon after publication of the 2005 studies, the first high-throughput sequencing machines were marketed, with products based on a range of massively parallel sequencing technologies (MILESTONE 5). By making sequencing high-throughput and affordable, these technologies ushered in a new age of next-generation sequencing.

Joseph Willson,
Nature Reviews Cross-Journal Team

ORIGINAL ARTICLES Shendure, J. et al. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005) | Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005)

FURTHER READING Ronaghi, M., Uhlén, M. & Nyrén, P. A sequencing method based on real-time pyrophosphate. *Science* **281**, 363–365 (1998) | Mitra, R. D. & Church, G. M. In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res.* **27**, e34 (1999) | Brenner, S. et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **18**, 630–634 (2000) | Dressman, D. et al. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl Acad. Sci. USA* **100**, 8817–8822 (2003)