Roche

Diagnostics

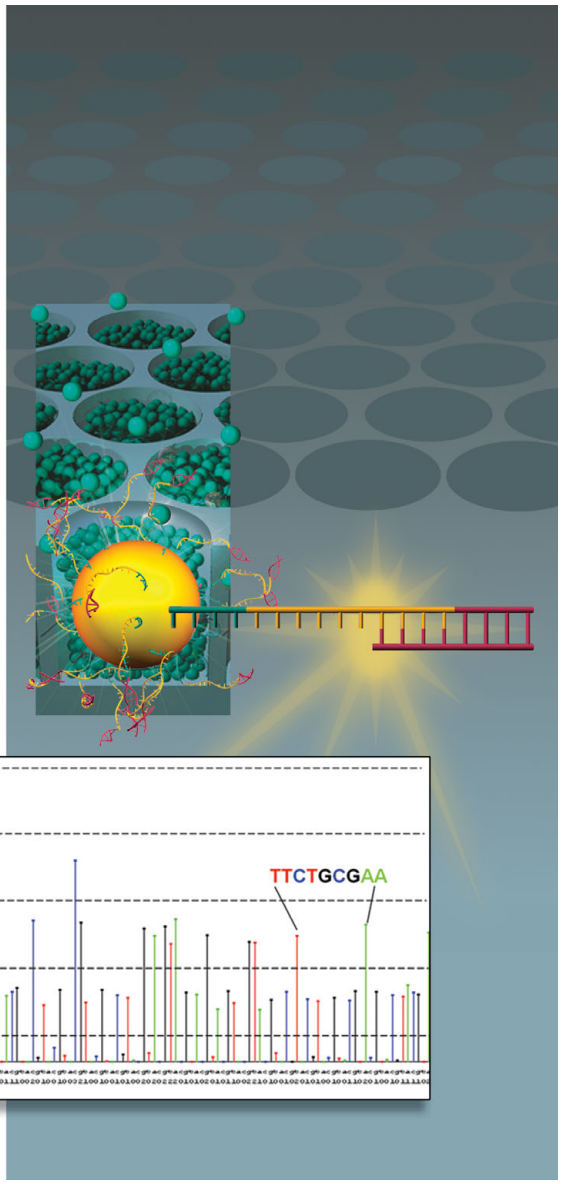
Genome Sequencer FLX System

First to the Finish

www.roche-applied-science.com

454 LIFE SCIENCES

Roche Applied Science introduces the second-generation sequencing technology – the Genome Sequencer FLX System – with more flexibility, supporting more applications. Expanding upon the proven technology and solid performance of the Genome Sequencer 20 System, this new platform enables longer read lengths and more reads per instrument run, providing throughput that is unparalleled in today's market.



Genome Sequencer FLX System

More flexibility, more applications

The Genome Sequencer FLX System is the newest addition to Roche's next-generation sequencing platform, building upon proven technology and peer-reviewed publications.

- Generate over 100 million bases per 7.5-hour instrument run.
- Achieve longer reads, averaging 200 to 300 bases.
- Attain higher throughput with over 400,000 reads per run.
- Generate single-read accuracy that is greater than 99.5% for over 200 base pair reads.
- Benefit from consensus accuracy that is greater than 99.99%.

The Genome Sequencer FLX System features complete software packages, including easy-to-use graphical user interfaces (GUI) for mapping, assembly, and amplicon variation detection.

- Use the project management software tool to group multiple instrument runs for analysis, and add instrument runs over time for additional analysis power.
- Obtain higher-confidence results using next-generation algorithms — from base calling to assembly and mapping, along with variation analysis.

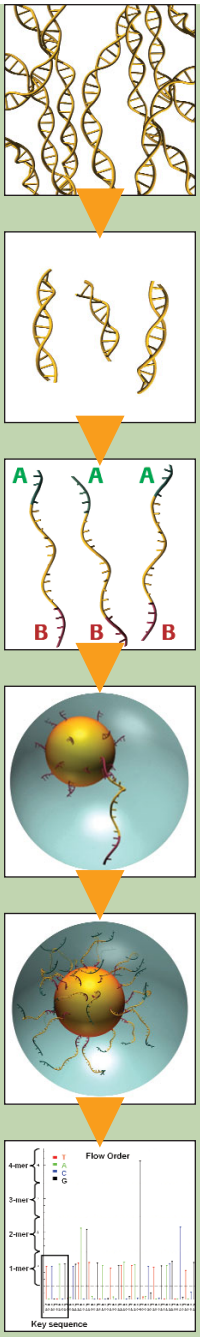
Perform breakthrough science with more applications using the flexibility of the Genome Sequencer FLX System.

- *De Novo* Sequencing Supported by Paired-End Applications and BAC Resequencing
- Comparative Genomics using Whole Genome Sequencing
- Amplicon Resequencing
- Transcriptome Analysis
- Gene Regulation Studies

Genome Sequencer FLX System Workflow

One fragment = One bead = One read

The Genome Sequencer FLX System provides a complete solution — from sample preparation through digital data analysis.



1) Sample Input: The Genome Sequencer FLX System supports the sequencing of samples from a variety of starting materials, including genomic DNA, PCR products, BACs, and cDNA.

2) Sample Fragmentation: Samples such as genomic DNA and BACs are fractionated into small 300 to 800 base-pair fragments. For some samples, such as small non-coding RNA, fragmentation is not required. Short PCR products can be amplified using Genome Sequencer fusion primers to go directly to Step 4, shown below.

3) Adaptor Ligation: Using a series of standard molecular biology techniques, short adaptors (A and B) – specific for both the 3' and 5' ends – are added to each fragment. The adaptors will also be used for purification, amplification, and sequencing steps. Single-stranded fragments, shown here, are used in subsequent steps in the workflow.

4) One Fragment = One Bead: The first step in emPCR (emulsion PCR) is shown. The adaptors enable hundreds of thousands of single-stranded fragments to bind to their own unique beads. The beads are then encapsulated into individual droplets formed by a water-in-oil emulsion, creating a microreactor containing one bead with one unique fragment. Each unique fragment is amplified without the introduction of competing or contaminating sequences. The entire fragment collection is amplified in parallel.

5) One Bead = One Read: The emPCR process amplifies each fragment to a copy number of several million per bead. Subsequently, the emulsion is broken while the fragments remain bound to their specific beads. After enrichment, the clonally amplified bead is ready to load onto the PicoTiterPlate device for sequencing.

6) Sequence Generation and Data-Analysis Tools: The Genome Sequencer FLX System produces over 400,000 reads per 7.5-hour instrument run. For sequencing-data analysis, three different bioinformatics tools are available for the following applications: resequencing up to 3 gigabases; amplicon variant detection by comparison with a known reference sequence; and *de novo* assembly up to 120 megabases.

Figure 1: Genome Sequencer FLX System Workflow Overview.

Performance Data and Flexibility

	E. coli	*T. thermophilus*
Genome Size (bases)	4,639,675	2,127,575
Mean Read Length	250	300
Oversampling	22	26
Assembly Contigs	105	59
Overall Accuracy	99.998%	99.994%
Single-Read Accuracy	>99.5%	>99.5%

Table 1: The table shows the results of *de novo* assembly of *E. coli* and *T. thermophilus*. Mean read length is approximately 250 or 300 base pairs. The single-read accuracy is unprecedented (greater than 99.5%) and the overall accurcy rates are greater than 99.99 %.

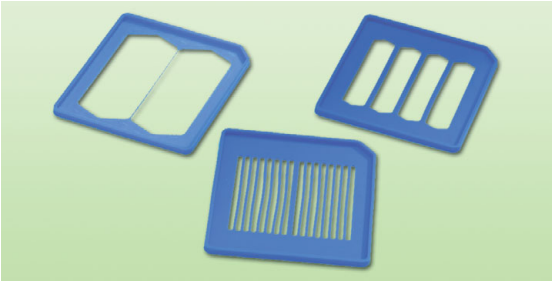


Figure 2: More flexibility with different sample loading options. For projects in which higher numbers of samples are needed, with fewer reads per sample, a series of gaskets are provided. These provide physical barriers to divide the PicoTiterPlate device into smaller segments, facilitating the loading of 2 to 16 samples onto one PicoTiterPlate device.

	Gasket	Number of reads per region	Number of bases per region*	Total reads for PicoTiterPlate device (all regions)	Total bases/ PicoTiterPlate device (all regions)*
70x75 PicoTiterPlate Device	2 regions	210,000	50,400,000	420,000	100,800,000
	4 regions	70,000	16,800,000	280,000	67,200,000
	16 regions	12,000	2,880,000	192,000	46,080,000

Table 2: Number of reads per PicoTiterPlate device configuration.

* Assumes read length of 240 bases per read. Results can vary depending upon average read length observed for a specific genome.

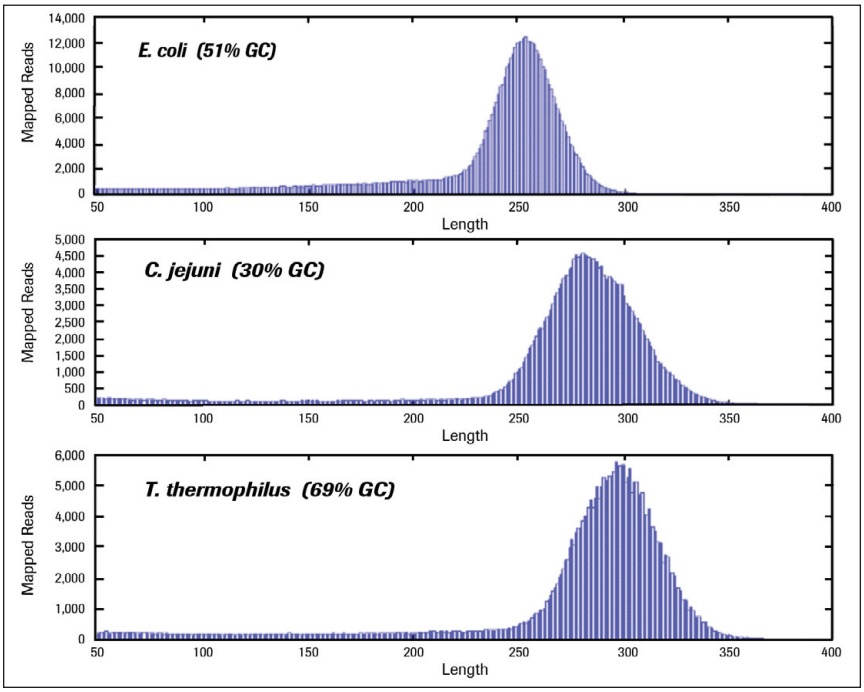


Figure 3: Average read lengths from different genomes. The read length is dependent on the characteristics of DNA sequence in the genomes. For example, genomes that are more AT or GC rich (and therefore more likely to have homopolymer repeats) yield a longer read length distribution as compared to an AT/GC neutral genome.

Applications and References

The Genome Sequencer FLX System supports an increasing number of applications, from characterizing genomic DNA to identification of transcripts, and gene regulation studies using RNA.

De Novo, Whole Genome, and BAC Sequencing

Perform whole genome shotgun sequencing of unknown genomes from bacteria and viruses in a single run, to more complex animal and plant genomes using BAC clones. Assembly of contigs into scaffolds is supported by a paired-end application with dedicated reagents and software.

References

Andries, K. *et al.*, "A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*," *Science* **307,** 223-227 (2005).

Velicer, G.J. *et al.*, "Comprehensive mutation identification in an evolved bacterial cooperator and its cheating ancestor," *PNAS* **103** (21), 8107-8112 (2006).

Jarvie, T., "Whole genome assembly using paired end reads in *E. coli, B. licheniformis,* and *S. cerevisiae,*" Genome Sequencer System Application Note No. 1/July 2006. Available at ***www.genome-sequencing.com***

Comparative Genomics using Whole Genome Sequencing

Perform comparative genomics in an unprecedented fashion to support the identification of genetic variations, including heterozygous SNPs, insertion-deletions, and mutation hotspots. Epidemiological studies, evolutionary studies, and metagenomics are also supported.

References

Poinar, H.N. *et al.*, "Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA," *Science* **311** (5759), 392-394 (2006).

Green, R.E. *et al.*, "Analysis of one million base pairs of Neanderthal DNA," *Nature*, **444** (7117), 330-336 (2006).

Amplicon Resequencing

Utilize the new amplicon sequencing tools to address the identification of somatic mutations in complex cancer samples or the analysis of genetic variability in human, animal, plant, or microbial populations. Due to the high number of reads that can be generated, the Genome Sequencer FLX System has a sensitivity that can detect genetic variations to 5% with just 1000x oversampling and 1% with 5000x coverage.

References

Thomas, R.K. *et al.*, "Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing," *Nature Medicine* **12**, (7) 852-855 (2006); advance online publication, doi:10.1038/nm1437 (2006).

Sogin, M.L. *et al.*, "Microbial diversity in the deep sea and the underexplored 'rare biosphere'," *PNAS* **103** (32), 12115-12120 (2006).

Transcriptome Analysis

Detect and characterize an equal number of expressed sequence tags (ESTs) – regardless of whether the genome is known or unknown. Both novel and known ESTs can be identified, providing information on the types of genes expressed, and whether translocation and trans-splicing events have occurred.

References

Gowda, M. *et al.*, "Robust analysis of 5'-transcript ends (5'-RATE): A novel technique for transcriptome analysis and genome annotation," *Nucleic Acids Research* Advance Access published online September 29, 2006, doi:10.1093/nar/gkl522 (2006).

Bainbridge, M.N. *et al.*, "Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach," *BMC Genomics* **7**, 246 doi:10.1186/1471-2164-7-246 (2006).

Nielsen, K.L. *et al.*, "DeepSAGE-digital transcriptomics with high sensitivity, simple experimental protocol and multiplexing of samples," *Nucleic Acids Research*, Advance Access published online October 5, 2006 doi:10.1093/nar/gkl714 (2006).

Gene Regulation Studies

- Discover new classes of small non-coding RNA (sncRNA) along with known sncRNAs. Perform genome-wide identification of sncRNA to better understand the controls for gene expression.
- Complete a whole-genome map of transcription factor binding sites using the Genome Sequencer FLX System in combination with chromatin immunoprecipitation (ChIP) studies.
- Quantitatively characterize the methylation state of each CpG dinucleotide in a target genomic sequence by utilizing a known bisulfite procedure and the Genome Sequencer FLX System. Loss of methylation, as well as hypermethylation of CpG islands within promoter regions is known to be an important regulation mechanism for many genes.

References

Axtell, M.J. *et al.*, "A two-hit trigger for siRNA biogenesis in plants," *Cell* **127**, 565-577 (2006).

Berezikov, E. *et al.*, "Diversity of microRNAs in human and chimpanzee brain," *Nature Genetics*, published online October 29, 2006, doi:10.1038/ng1914 (2006).

Lu, C. *et al.*, "MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase-2 mutant," *Genome Research* online before print September 5, 2006, doi:10.1101/gr.5530106 (2006).

Girard, A. *et al.*, "A germline-specific class of small RNAs binds mammalian Piwi proteins," *Nature* **442**, 199-202 (2006) advance online publication June 4, 2006, doi:10.1038/nature04917 (2006).

Turcott, C. and Jarvie, T., "Detection and quantitation of methylation patterns using amplicon sequencing," Genome Sequencer Application Note No. 3/September 2006. Available at ***www.genome-sequencing.com***

For additional references and detailed information, visit
www.genome-sequencing.com
or contact your local sales representative.

Limited Label Licenses, Disclaimers, and Trademarks

454, GENOME SEQUENCER, and PICOTITERPLATE are trademarks of 454 Life Sciences Corporation, Branford, CT, USA.

Other brands or product names are trademarks of their respective holders.

RESTRICTION ON USE: Purchaser is only authorized to use the Genome Sequencer Instrument with PicoTiterPlate devices supplied by 454 Life Sciences Corporation and in conformity with the procedures contained in the Operator's Manual.

For life-science research only.
Not for use in diagnostic procedures.

© 2006 Roche Diagnostics GmbH. All rights reserved.

454 LIFE SCIENCES

Roche Diagnostics

Roche Diagnostics GmbH
Roche Applied Science
68298 Mannheim
Germany

05003075001 • 12/06

www.roche-applied-science.com