

# Solid Race

Novel sequencing technologies enable new dimensions in genetic analysis and product development. With Applied Biosystem's recently launched SOLiD™ System sequencer, competition is heating up as a workshop by GATC Biotech showed.

In 1998, about 200 Mb of human DNA fragments were sequenced. Data was generated with gel electrophoresis. Five years later, US Joint Genome Institute alone had an output of 1.5 billion bases per month. This was based on capillary array electrophoresis, and Applied Biosystems (AB) was the undisputed leader in automated DNA-sequencers. Since then, reputed

petition among instrument providers is heating up.

## Tremendous Need for Sequencing: Demand for Higher Quality

After termination of the Human Genome Project, there is a pressing need to follow up promising clues and probe in depth first results. The total number of genomes completed and newly started projects is rapidly increasing (see chart). There is a clear need for improved instrumentation and a business opportunity for all involved.

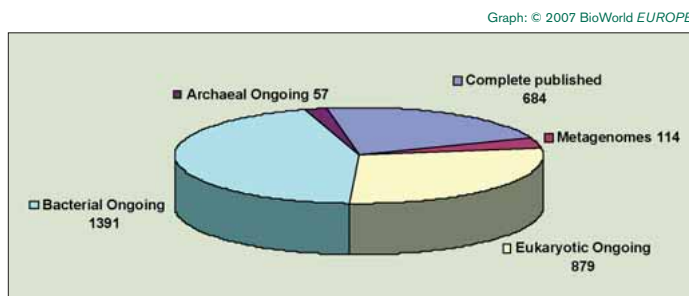
Experts estimate the current sequencing market in the US alone at about \$700m (€456m). The other half of revenues are generated outside the US, primarily in Europe. Next-generation

technologies expand the current market to include gene expression and genotyping. According to Jay Flatley, CEO of Illumina, today that market worldwide totals about \$2,25bn (€1,5bn) expected to rise to \$3,7bn (€2,5bn) in 2010.

A major attractiveness of next-generation sequencers – for academic customers, clinical labs and SME biotechs, in particular – is in massively reduced cost, which is kindling further demand. Whereas sequencing a whole genome with traditional technologies costs between three to seven million Euros – depending on the project size, one may get it now for just around 70.000 €, on AB's SOLiD™ System sequencer with an output up to 4 gigabases per run. In addition to whole genome sequencing, the new AB sequencer offers flexibility for applications ranging from targeted re-sequencing, methylation studies,

microRNA analysis up to chromatin immuno precipitation (ChIP). For example, two SOLiD runs on the Columbia Breast Cancer Database resulted in 66.6 million CpG-islands analyzed. The instrument is scalable and robust – defined as the ratio of system downtime to run success. Thierry Scarcez, a molecular biologist working for Applied Biosystems in Belgium said at the GATC Workshop that in contrast to other instruments the SOLiD™ System enables users to re-enter the workflow at multiple steps. If a run is incomplete or failed at some step, users do not need to regenerate the sample or slide.

In the end, it is the combination of reliable high throughput with accuracy, which distinguishes the ligation approach. High fidelity and quality of read lengths as short as 25 base pairs are achieved through 2-base encoding (see chart). 16 dinucleotide combinations with 4 fluorescent dyes are probed in two different reactions allowing to distinguish between false positives and true polymorphisms. This is especially important, if re-sequencing short reads and targeting human mutations or rare disease vari-



Genomes sequenced worldwide (Numbers up to Nov 27, 2007).

Source: GOLD Genomes OnLine Database, 27-11-2007.

competitors emerged as development of methods and instruments greatly improved. During the last two years Solexa (now Illumina) and 454 Life Sciences (now Roche) both launched new sequencers based on pyrosequencing resp. reversible terminator sequencing-by-synthesis. What is termed «next-generation sequencing» is offering greater depth of coverage and reduced read length compared to the traditional Sanger method. The new technologies enable ever faster, more accurate and cost-efficient sequencing. Now, with its SOLiD™ System sequencer based on ligation of oligonucleotides, AB takes up the challenge. The new system (De la Vega et al. 2007) was launched at the American Society of Human Genetics meeting in October. ABI is in the running again – in the US, but what about Europe? A user workshop on novel sequencers organised by GATC Biotech shows that com-

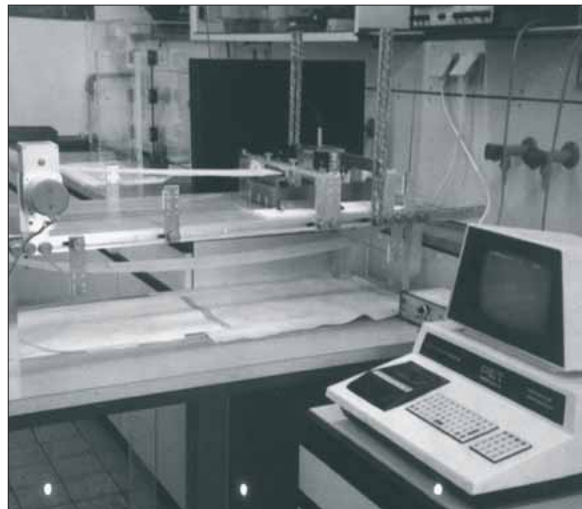


Applied Biosystem's SOLiD™ System sequencer at GATC Biotech.

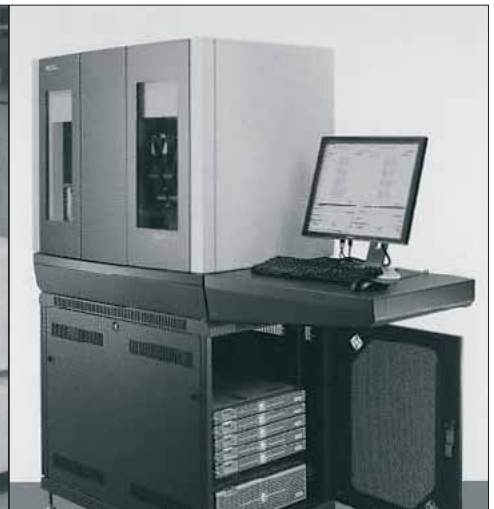
ants. A specific algorithm (McLaughlin et al. 2007) assigns to each single base in the read a quality score that allows discrimination from errors. «The quality scores for the certainty of the base call which are provided for the SOLiD sequencer», says Christopher Bauser, Head of Bioinformatics with GATC, «can be further improved using information about the reference genome.» Bauser is regularly testing software provided by sequencer manufacturers and developing inhouse projects. AB offers access to algorithms upon request to interested bioinformaticians.

**Helping Customers Deriving Meaningful Knowledge**

In Europe GATC Biotech, Konstanz (DE) is one of the few included in AB's Service Provider program. The company is already operating the new AB SOLiD™ System sequencer. GATC was chosen for its longstanding experience with automated sequencing. Biophysicist Fritz Pohl, former student of nobel laureate Manfred Eigen discovered Z-DNA and invented the first ever sequencing robot in 1980. Since then, his sons Peter, Thomas, and Fritz Pohl jr. have taken his heritage many steps



First DNA-sequencer invented by Fritz Pohl 1980s



Next-generation sequencer by ABI 2007

further. Today, GATC has «all leading sequencing technologies in one lab» – AB's 3730XL, Illumina's Genetic Analyzer, Roche's GS FLX, and SOLiD™ System.

Another reason is the excellent customer base GATC retains. Indeed, the company is reputed to offer the customer «the technology that fits their projects best.» CTO-Thomas Pohl and his team draw on extensive experience from sequencing projects large and small, routine and specialised. At the workshop they came up with a first comparison of the three major next-generation sequencers (see table).

While each of these technologies retains its proper strengths, the accuracy

level of the SOLiD™ System is astounding, said Pohl at the meeting, and added «this is what GATC's customers expect especially those in drug discovery studying genetic variations and interested in how individual genotypes respond to pharmacological treatments.» Customers from other application fields also benefit from GATC's expertise with different sequencing technologies. One of them is Christian Bachem, coordinator of the global potato genome sequencing project (www.potatogenome.net) at Universiteit Wageningen (NL). The Dutch coordinators of the consortium chose GATC as one of the sequencing providers after a thorough assessment

of numerous service providers across the globe. «They had an extremely good price – quality relationship» commented Bachem on the choice. As concerns the SOLiD™ System-instrument, potential benefits for his project are in that «mate-pairs will provide us with information to finish BAC sequencing

TABLE			
	GS FLX	Genetic Analyzer	SOLiD System
read length	Ø 250 bp	36 bp	35 bp
mate pairs	2 x 25 bp	2 x 25 bp	2 x 25 bp
	insert sizes up to 2 kb (2 x 100 bp soon)	insert sizes up to 600 bp not yet released	wide range of insert sizes (from 200 bp up to 10 kb)
# of reads / run	400.000	40.000.000	85.000.000 (mate pair)
accuracy	raw 99.7%	raw 98.5%	raw 99.94%
samples / run	16 compartments (1 plate)	8 (one flowcell; 8 channels)	16 chambers (2 slides)
applications	de novo sequencing, cDNA libraries/ESTs, amplicon sequencing/long range PCR fragments, human samples, BAC pools, fosmid pools, metagenomes/biofilm, transcriptomes, LAM-PCR	resequencing, ChIP, small RNAs, SAGE & CAGE, de novo sequencing	resequencing, ChIP, small RNAs, SAGE & CAGE, de novo sequencing

Comparison of Next-Gen Sequencers



«GATC is an excellent partner for sequencing. They have all major platforms in-house. That is for us of interest too.» says Daniel Summerer, Head of Application/EOC-technologies with FEBIT (left: Thomas Pohl, CTO GATC).

that we are doing now BAC by BAC. Not only will it be able to orientate individual contigs within BACs, but also will it facilitate gene model building for our annotation pipeline.» With up to three Gigabases capacity, Bachem says, «this sequence data will probably become a resource for the whole project for a very long time – even if you end up with just five per cent of that much data that is immediately useful for our particular project.» □

#### Reference

- De la Vega FM, Sorenson J, Hyland F, McKernan K, Kim W, Finch SJ, Gordon D (2007): Applications of Next-Generation Sequencing in Genetic Epidemiology. Presentation American Society of Human Genetics, San Diego, Oct. 24.
- McLaughlin SF, Peckham HE, Zhang ZH, Ranade S, Kosnopo C, Costa G, Malek J,

Sorenson JM (2007): Whole-genome Resequencing with Short reads: Accurate Mutation Discovery with Mate-Pairs and Quality Values. Presentation American Society of Human Genetics, San Diego, Oct. 26.

- Hyland FCL, Peckham H, Malek J, Cuppen E, Sorenson J, McKernan K, De la Vega FM (2007): Next-generation Sequencing of 1000 Samples To Detect Rare Variants - Opportunities and Constraints. Presentation American Society of Human Genetics, San Diego, Oct. 25.

#### Contact:

GATC Biotech AG  
 Elke Decker  
 Jakob-Stadler-Platz 7  
 D-78467 Konstanz  
 Tel.: 0049 7531 81 600  
 Fax: 0049 7531 81 6081  
 e-mail: e.decker@gatc-biotech.com  
<http://www.gatc-biotech.com>