Takeda Award 2001 Achievement Fact

Techno-Entrepreneurial Achievements for Individual/Humanity Well-Being

Technical Achievement:

The development of a large-scale genome sequencing system by establishing " the whole genome shotgun strategy "that utilizes modularized data acquisition system and high-throughput DNA sequencers

The prize is awarded jointly to Michael W. Hunkapiller (Applied Biosystems) and J. Craig Venter (Celera Genomics).

Michael W. Hunkapiller is honored for his contribution to the development of the automated high-throughput DNA sequencers and the promotion of the foundation of Celera.

J. Craig Venter is honored for the foundation of Celera and the development of " the whole genome shotgun strategy."

(Awardees are in alphabetical order.)

〈Executive Summary〉

The Takeda Foundation has decided to award the 2001 Takeda Award in the Individual and Humanity Well-being Field to Michael W. Hunkapiller and J. Craig Venter for their contributions to the development of a large-scale genome sequencing system by establishing" the whole genome shotgun strategy "that utilizes modularized data acquisition system and high-throughput DNA sequencers

Hunkapiller recognized the importance of automated and high throughput analytical instruments in the coming era of large-scale genome sequencing, and produced a fully automated high throughput DNA sequencer, the PRISM3700. He combined the technologies of high throughput multi-capillary electrophoresis with fluorescent dye chemistry and an automated sample exchange system, and the sheath flow detection method to

create the PRISM3700, which has a 10 fold higher productivity than conventional slab-gel DNA sequencers. Later, the PRISM3700 became a powerful tool for human genome sequencing and was used by both the Human Genome Project (HGP) and Celera. Genomics.

Venter recognized the labor and time saving possibilities of the whole genome shotgun method, in which the whole genome DNA of an organism is randomly broken into millions of fragments and sequenced, and the resulting short sequences are aligned through overlapping regions using a computer algorithm, although the major players of genome analysis relied upon sequencing fragments of genome DNA after their locations are confirmed in the original DNA. He developed an assembly algorithm that can handle large volumes of sequencing data produced by shotgun sequencing of large-scale genomes such as the human genome.

In 1998, with a way open to the production of vast volumes of sequencing data by the emergence of the fully automated high throughput DNA sequencer, the PRISM3700, and the development of the whole genome shotgun sequencing algorithm, Hunkapiller and Venter developed a modular sequencing system that enabled the systematic and concentrated production and treatment of genome information with the operation of 300 PRISM3700 sequencers in one place, which is a novel engineering intellect and knowledge in the biological measurement. They founded a private company, Celera Genomics, and succeeded in sequencing the human genome in a short period.

Hunkapiller and Venter also demonstrated that a commercial enterprise can carry out multi-centered basic research such as the sequencing of the human genome in a faster and more economical fashion than publicly funded research facilities and succeeded in attracting private capital to their work. Their achievement should be highly evaluated as techno-entrepreneurship.

The Celera¡s effort stimulated and accelerated the publicly funded Human Genome Project and as a result, has opened the way to industrial applications of sequence information and contributed greatly to the development of genomic science. The achievement by Hunkapiller and Venter is having a great impact on medicine, agriculture, pharmaceuticals, bioinformatics, and is expected to make a far-reaching and profound contributions to individual and humanity well-being.

Technical Achievement

The development of a large-scale genome sequencing system by establishing" the whole genome shotgun strategy " that utilizes modularized data acquisition system and high-throughput DNA sequencers

Achievement and Creativity

1. Development of high-throughput DNA sequencers

1.1 Sequencing of genomes

Michael W. Hunkapiller produced a fully automated high throughput multi-capillary DNA sequencer, which has a 10 fold higher productivity than conventional slab-gel DNA sequencers.

The genetic information in the DNA is stored as a sequence of bases that encodes a protein and DNA sequencing is the determination of the exact order of the base pairs in a segment of DNA. A genome is a complete set of all genes and other DNA that does not code for proteins in the chromosomes. A genome is not only a complete blueprint of a living cell but also a history of evolution experienced by the organism before, and the sequencing of the genome can lead to both the elucidation of gene function and an understanding of evolution. Hunkapiller recognized the importance of automated and high throughput analytical instruments for large scale genome sequencing, predicting that an era of the large scale genome sequencing was coming.

1.2 Development of first DNA sequencing technology

In 1977, the first techniques for DNA sequencing were reported independently by A.M. Maxam and W. Gilbert[1], and F. Sanger and co-workers[2]. Later they received the Nobel Prize for the development of DNA sequencing techniques. The Maxam and Gilbert method involves the use of base selective degradation reactions to produce radio-isotope labeled fragments, which are then submitted to slab-gel electrophoresis to determine the sequence. In the Sanger method, fragments to be sequenced are reacted with radio-labeled short DNA chains (primers) and deoxynucleotides mixed with chain-terminating dideoxynucleotides to produce radio-labeled fragments. Since the reaction with one kind of chain-terminating dideoxynucleotide gives a sequence order concerning one kind of deoxynucleotide, four different reactions with four different kinds of chain-terminating dideoxynucleotides are necessary to determine the complete sequence order of a given fragment. The radio-labeled fragments are submitted to slab-gel electrophoresis and bands are detected by autoradiography (Figure 1. (a)). The sequence is interpreted from

the pattern of alternative bands in the lanes corresponding to the terminal base of the fragment. These methods provided a powerful tool for DNA sequencing, but are very tedious and labor-intensive.

1.3 Development of fluorescent detection method

An advance in sequencing technology occurred in 1986 and 1987 when L. E. Hood and coworkers at Caltech[3], W. Ansorge and coworkers at EMBL[4], and J. M. Prober and coworkers at DuPont[5] independently reported laser-induced fluorescence detection methods for DNA bases. Their methods involve the use of fluorescent labeled primers to produce fluorescent labeled fragments. Since the fluorescent methods do not involve the use of hazardous radio-isotope labeled primers and time-consuming autoradiographical treatment, they replaced the radioisotope labeling method. Hunkapiller was a post-doctoral researcher in Hood 's laboratory at Caltech, where he participated in the development of a fluorescent detection method using four dyes with four different absorption spectra (Figure 1 (b)). In this method, bases are monitored by detecting fluorescent light near the end of the slab-gel while the gel is running. This real time monitoring led to the development of automated high-throughput DNA sequencers. Later, Hunkapiller joined Applied Biosystems, Inc. (ABI), and the Caltech method has been successfully used in the detection systems of the DNA sequencers developed by ABI[6]. The productivity of DNA sequencers increased several hundred folds after the development of the fluorescence detection methods for DNA bases.

1.4 Development of high throughput multi-capillary DNA sequencers

However, even after the development of fluorescent detection methods, the productivity of DNA sequencers was not sufficient to decode the human genome in a limited time. Another advance was necessary to achieve higher productivity. This occurred when capillary electrophoresis was introduced for DNA sequencing[7-9]. The separation speed is limited in slab gel electrophoresis because of the heat produced when the high electric field is applied to the gel. Since heat elimination is very rapid from the large surface area of a capillary, a higher electric field can be applied to capillary electrophoresis, thus speeding up the separation process. By using a capillary gel, the separation speed is increased about 10 fold over conventional slab-gel systems (Figure 1. (c)).

Treating multiple samples at the same time is essential for high throughput. This can be achieved by employing multi-capillary systems (Figure 1(d)), but a serious problem arises from the detection of DNA bases in multi-capillary systems. The detection of fluorescence

from DNA bases is complicated by the scattering of light from the porous matrix and capillary walls. R. Mathies developed a method that uses a confocal fluorescence scanner to avoid light scattering[10]. Using this confocal detection method, Molecular Dynamics commercialized the first multi-capillary DNA sequencer, the MegaBACE1000, in 1997.

Another solution to this problem was provided independently by Hideki Kambara of Hitachi Corporation and J. C. Dovichi of the University of Alberta. Essentially, their method involves the use of a sheath flow, which is a hydrodynamically focused stream from the outlet of the capillaries, and the direct detection of the fluorescence from the sheath flow in the sheath flow cuvette (Figure.2) [11,12]. Since this method does not use capillaries in the fluorescence detection part, the detection does not suffer from light scattering from the capillary walls, and the sensitivity is greater than that of confocal detection. This method also allows the introduction of laser from the side wall of the cuvette, which simplifies the laser introduction apparatus.

Hunkapiller was in overall charge of the development of DNA sequencers at ABI. While struggling with the development of multi capillary electrophoresis, he combined the ABI proprietary technologies of high throughput multi-capillary electrophoresis with fluorescent dye chemistry and an automated sample exchange system, and the sheath flow detection method, acquiring a license from Hitachi and the University of Alberta, and put the automated high-throughput multi capillary DNA sequencer, PRISM3700, on the market. The PRISM3700 uses 96 capillaries and its overall productivity is 10 times greater than the most productive slab-gel DNA sequencers.

2. The shotgun strategies

2.1 Whole genome shotgun sequencing

J.Craig Venter recognized the labor and time saving possibilities of the whole genome shotgun method, and developed an assembly algorithm that can handle large volumes of sequencing data produced by shotgun sequencing of large scale genomes such as the human genome.

The whole genome shotgun method is a strategy that skips the laborious and time-consuming mapping process of the hierarchical shotgun approach by reassembling random fragments taken from the genome as a whole using computer algorithms (Figure 3.). The strategy seems very simple and straightforward, but when it is applied to the sequencing of large genomes such as the human genome, it is necessary to produce millions of sequence fragments and reassemble them in a limited time. The strategy

requires high-throughput DNA sequencers that enable the sequencing of hundreds of thousands of bases per day, and a computer algorithm that permits the assembly of millions of random fragments. Venter founded the Institute for Genome Research (TIGR), where an algorithm was developed for the assembly of millions of cDNA fragments synthesized with the human messenger RNA[13]. Using this algorithm, Venter and his coworkers at TIGR completed the sequencing of the *Haemophilus influenza* genome, which contains 1.8 million bases[14]. Although sequencing of a megabase-scale genome was possible, most genomic research scientists thought it would not work for the human genome, which contains millions of repetitive DNA sequences.

2.2 Hierarchical shotgun sequencing

The sequencing of the human genome can lead to an understanding of human evolution, underlying molecular mechanisms of disease, and the interplay between the environment and heredity in defining the human condition. The sequencing of the human genome was started as an international project, the Human Genome Project (HGP), and in the early phase of the sequencing of the human genome by HGP, efforts were concentrated on making genetic, physical and sequence maps, and, later, a hierarchical shotgun approach (Figure 3.) to sequencing the human genome was adopted.

In the hierarchical shotgun approach, the entire genome is shredded into large fragments of several hundred thousands bases, and these large fragments are then positioned on the genome chromosomes by looking for distinctive marker sequences, whose locations have already been pinpointed (mapping process). The large fragments are then shattered into tiny fragments and each fragment is sequenced, and computer algorithms that recognize matching sequence information from overlapping fragments are used to reconstruct the complete sequence of the large fragments. The mapping process turned out to be laborious and time-consuming and the sequencing speed was not fast enough due to the limited productivity of the DNA sequencers available in those days. It was obvious that both methodological and technological innovations were necessary to complete the human genome sequencing on schedule.

2.3 The development of the whole genome shotgun strategy and sequencing of the human genome

In January 1998, the emergence of the fully automated high throughput DNA sequencer, the PRISM3700, and the development of the whole genome shotgun sequencing

algorithm, changed the situation that confronted human genome sequencing. Right before launching the sequencer on the market, Hunkapiller invited Venter to see the new sequencer[15] and made a bold proposal to found a private company whose business would involve sequencing the human genome using the newly developed DNA sequencers, building up a database for the human genome, and selling it to customers. Hunkapiller and Venter discussed the feasibility of the plan and evaluated the strategies for completing the sequencing of the human genome.

They set up a strategy to apply the whole genome shotgun method to sequencing the human genome. The strategy (the whole genome shotgun strategy) involves the use of a modular sequencing system from sample preparation to assembly of sequencing data[13] (Figure 4.). Four principal modules were designed to operate independently: (i) library transformation, plating, and colony picking; (ii) DNA template preparation; (iii) dideoxy sequencing reaction set-up and purification; and (vi) sequence determination with the ABI PRISM3700 DNA sequencer. Also an automated trace-processing pipeline was developed to process each sequence file. In order to assure the quality of the processes, procedural controls were established to maintain the validity of the sequencing results as the sequencing reactions proceeded through the process.

In May 1998, they founded a private company, Celera Genomics, with three hundred automated high-throughput PRISM3700 DNA sequencers, and large scale computers to sequence the human genome. They declared that they would complete the sequencing of the human genome in three years using the whole genome shotgun method[16].

Before challenging the sequencing of the human genome, they tested the whole genome strategy by sequencing the Drosophila genome which contains 120 million bases, and completed the sequencing in only 4 months[17].

Armed with these developments at Celera Genomics and referring to the publicly available data, Hunkapiller, Venter, and coworkers completed the draft sequence of the human genome in 9 months and proved the validity of their strategy[13]. Without Hunkapiller and Venter, the acceleration of human genome sequencing would not have been occurred.

3. Repercussion effects

The achievement of Venter and Hunkapiller has wide and profound repercussion effects. The foundation of Celera and the attempt to sequence the human genome stimulated the

HGP and accelerated their efforts[18]. They completed the draft sequence of the human genome in June, 2000, and shortened the original schedule for the completion of the sequencing of the human genome from 2005 to 2003. The acceleration of genome sequencing by Hunkapiller and Venter has pushed the genomic research forward and post-sequencing research has already started.

The sequencing of the human genome by Celera Genomics has produced the very important engineering intellect and knowledge that the sequencing of large genomes, such as the human genome, can be conducted rapidly by the application of the whole genome shotgun method using systematic and concentrated production and treatment of large scale information in one place. The effectiveness and swiftness of the Celera¡s method demonstrated the importance of the collaboration between biologists and information scientists, and the power of large scale computers in genetics. Their methods can be applied to many areas of bioscience and technology and their achievement opens new horizons in the combination of information technology and bioscience, bioinformatics.

Their achievement also demonstrated that a commercial enterprise sometimes can conduct a large-scale basic research such as the sequencing of the human genome in a faster and more economical fashion than publicly funded research and this fact has encouraged the private sector to invest in other genomic companies, and the number of genome-related venture companies has doubled in two years[19].

The successful use of high-throughput DNA sequencers for human genome sequencing has made scientists realize their usefulness in the sequencing of genomes of other animals and plants, which have enormous potential in both medicine and agriculture. The commercialization of high-throughput DNA sequencers has also contributed to the growth of the biotech instruments industry, which manufactures not only DNA sequencers, but also DNA and peptide synthesizers, peptide sequencers, mass spectrometers, and DNA microchips.

Thus, the achievements by Hunkapiller and Venter in genome sequencing is having a great impact on both scientific and industrial sectors and is expected to be of great benefit to individual and humanity well-being.

Technical Achievement

The development of a large-scale genome sequencing system by establishing" the whole genome shotgun strategy " that utilizes modularized data acquisition system and high-throughput DNA sequencers

4. CURRICULUM VITAE

Michael W. Hunkapiller

B.S. in chemistry from Oklahoma Baptist University
Ph.D. in chemical biology from the California Institute of Technology.
1983 Joined the Research and Development Department of Applied Biosystems
1993 Applied Biosystems was acquired by Perkin-Elmer Corporation
1997 Vice President of Perkin-Elmer Corporation
1998 President of Applied Biosystems

J. Craig Venter

B.A. in biochemistry from the University of California, San Diego
Ph.D in physiology and pharmacology from the University of California, San Diego.
1984 Joined the National Institute of Neurological Disorders and Stroke
1992 Founded the Institute for Genomic Research
1998 Founded Celera Genomics

REFERENCES

1) Maxam, A.M., and Gilbert, W., " A new method for sequencing DNA, "*Proc. Natl. Acad. Sci. U.S.A.* 74, 560-564 (1977).
2) Sanger, F., Nicklen, S., Coulson, A.R., " DNA sequencing with chain-terminating inhibitors, " *Proc. Acad. Natl. Sci. U.S.A.* 74, 5463-5467 (1977).
3) Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H., and Hood, L.E., " Fluor escence detection in automated DNA sequence analysis, " *Nature* 321, 674-679 (1986).
4) Ansorge, W., Soroat, B.S., Stegemann, J., and Schwager, C., " A non-radioactive automated method for DNA sequence de termination, " J. Biochem. Biophys. Methods 13, 315-323 (1986).
5) Prober, J.M., Trainer, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jensen, M.A., Baumeister, K.," A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides, " *Science* 238, 336-341 (1987).
6) Smith, L.M., Hood, L.E., Hunkapiller, M.W., and Hunkapiller, T.J., " Automated DNA sequencing technique, " USP 5,171,534 (1992).
7) Drossman, H., Luckey, J.A., Kostichka, A.J., D 'Cunha, J., and Smith, L.M., " High-speed separation of DNA sequencing rea ctions by capillary electrophoresis, " *Anal. Chem.* 62, 900-903 (1990)
8) Guttman, A., Cohen, A.S., Heiger, D.N., Karger, B.L., " Analytical and micropreparative ultrahigh resolution of oligonucleo tides by polyacrylamide gel high-performance capillary electrophoresis, " *Anal. Chem.* 62, 137-141 (1990).
9) Swerdlow, H., and Gestaland, R., " Capillary gel elelctrophoresis for rapid, high resolution DNA sequencing, " *Nuc. Acids Res.* 18, 1415-1419 (1990).
10) Huang, X., Quesada, M.A., and Mathies, R.A., " Capillary Array Electrophoresis Using Laser-induced Confocal Fluores cence Detection, " *Anal. Chem.* 64, 967-972, (1992).
11) Kambara, H., and Takahashi, S., " Multiple-sheath flow capillary array DNA analyzer, " *Nature* 361, 565-566, (1993).
12) Dovichi, N.J., and Zhang, J.Z., " Multiple capillary biochemical analyzer, " USP 5,439,578 (1995).
13) Venter, J.C., et al., " The sequence of the human genome, " *Science* 291, 1304-1351, (2001).
14) Fleischmann, R.D., et al., " Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd, " *Science* 269, 496-512 (1995).
15) Davies, K., in " Cracking the genome, " Free Press, New York (2001).
16) Venter, J.C., Adams, M.D., Sutton, G.G., Kerlavage, A.R., Smith, H.O., and Hunkapiller, M., " Shotgun sequencing of the human genome, " *Science* 280, 1540-1542, (1998).
17) Myers, E.W., in " A Whole-Genome Assembly of Drosophila, " *Science* 287, 2196-2204, (2000).
18) International Human Genome Sequencing Consortium, " Initial sequencing and analysis of the human genome, " *Nature* 409, 860-921 (2001).
19)" Convergence: the Biotechnology Industry Report, " *Ernst & Young LLP, Millennium edition.* p16, 2000.

Technical Achievement

The development of a large-scale genome sequencing system by establishing "the whole genome shotgun strategy" that utilizes modularized data acquisition system and high-throughput DNA sequencers

Figure 1. Evolution of DNA sequencers: from radio-isotope labeled slab-gel to fluorescent labeled capillary electrophoresis

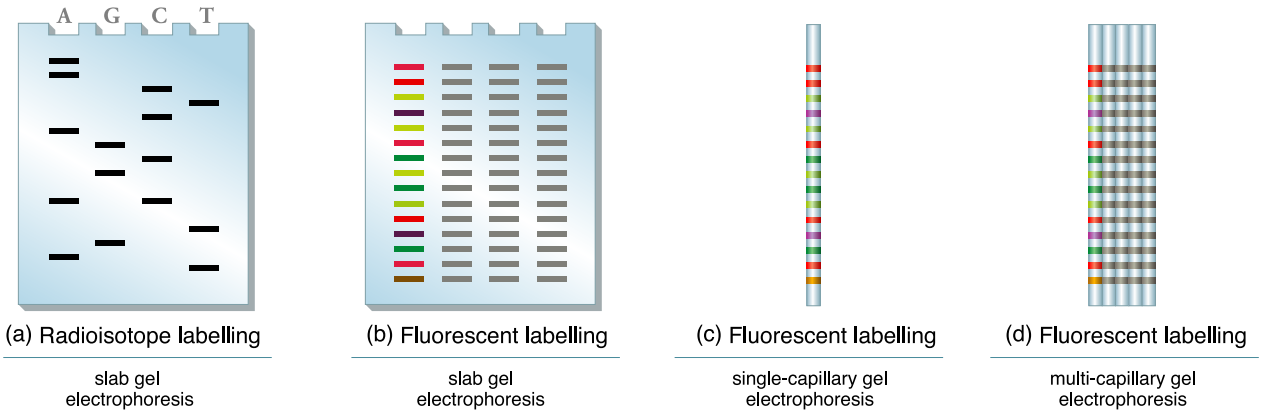


Figure 2. Schematic diagram of the sheath flow detection apparatus

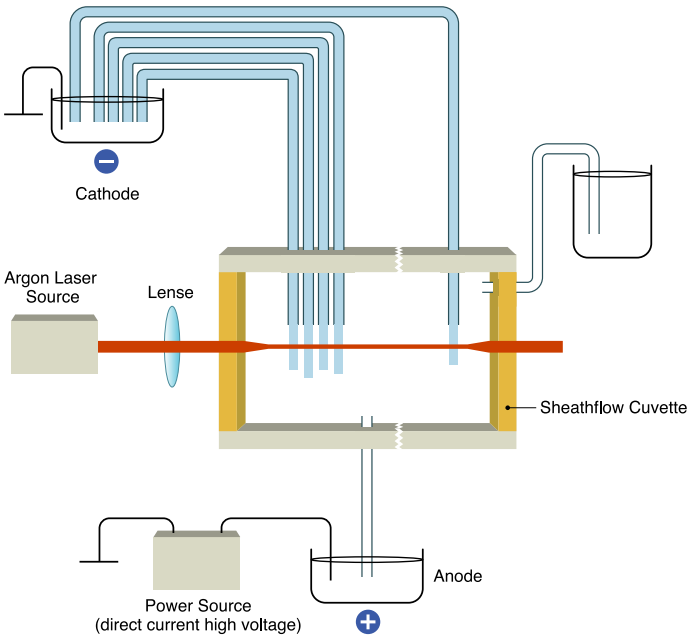
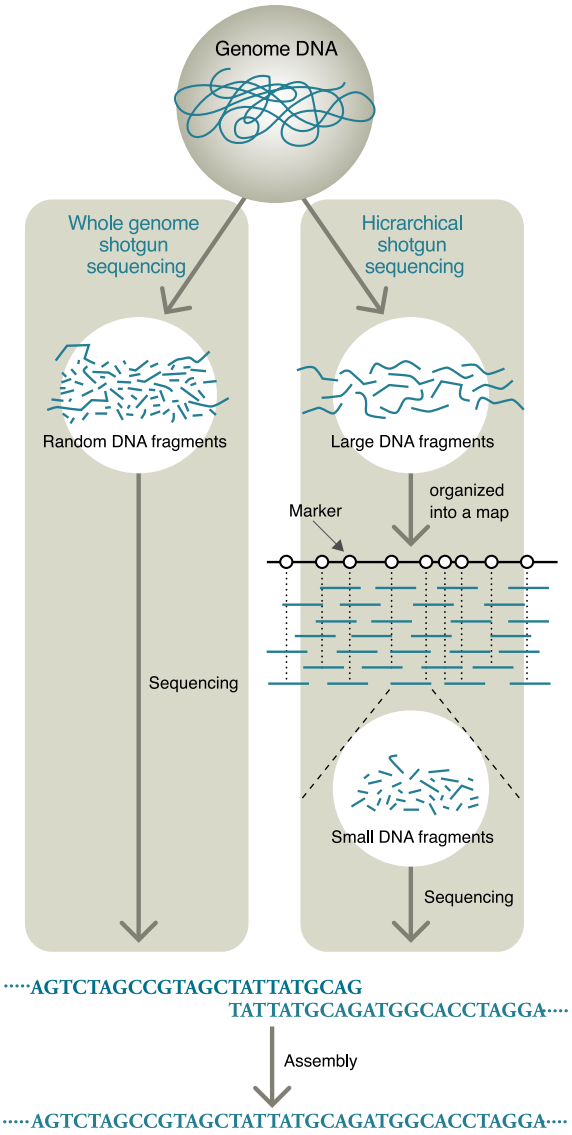


Figure 3. Schematic diagram of genome sequencing



Technical Achievement

The development of a large-scale genome sequencing system by establishing "the whole genome shotgun strategy" that utilizes modularized data acquisition system and high-throughput DNA sequencers

Figure 4. The whole genome shotgun processes (Venter, J.C., *et al.*, *Science* **291**, 1304-1351(2001))