Review on Digital Data into DNA synthesis, storage and sequencing

[1]Anushree Raj, [2]Rio D'Souza

[1]Assistant Professor, [2]Professor
[1]M.Sc. Big Data Analytics Department, [2] Computer Science and Engineering Department,
[1]St Agnes College Autonomous Mangalore, Karnataka, India, [2] St Joseph Engineering College Mangalore, Karnataka, India

***Abstract*: Storing data in the traditional way is getting too expensive. Solution for this in future could be the DNA storage techniques. DNA digital data storage is defined as the process of encoding and decoding binary data to and from synthesized DNA strands. DNA molecules are genetic blueprints for living cells and organisms. Researchers have made lot of attempts on the possibilities of recording, storage, and retrieval of information on DNA molecules.**

IndexTerms: Digital data into DNA, DNA synthesis, DNA storage, DNA Sequencing

---

1. INTRODUCTION

Synthetic DNA sequences have long been considered a potential medium for digital data storage [1, 2, 3]. DNA is an attractive possibility because it is extremely dense, with a theoretical limit above 1 EB/mm3 (eight orders of magnitude denser than tape), and long-lasting, with observed half-life of over 500 years in harsh environments [4]. This paper reviews the DNA storage and retrieval process on how a digital data can be synthesized and stored into a DNA molecule. Basic unit is DNA strands (100-200 nucleotides) are capable of storing 50-100 bits of data. Any digital data is initially transformed into DNA sequence which then undergoes the process of synthesis, storage and sequencing, which can later produce the decoded data, to resume its original form. DNA strands are stored in pools and the address of data is stored in the strand itself. The key-value is used to obtain the PCR primer sequence which gets attached to the encoded sequence and final DNA sequence is synthesized to store in the DNA library. The key is used to get the stored sequence which retrieves the molecule associated with the key. The DNA is extracted from the pool and further undergoes amplification to produce the decoded sequence. Finally the sequencer produces the digital data. This paper makes an attempt to discuss the various techniques used and procedure involved in DNA synthesis, storage and sequencing of DNA. DNA sequencing is the process of determining the sequence of nucleotides (As, Ts, Cs, and Gs) in a piece of DNA. A DNA strand, or oligonucleotide, is a linear sequence of these nucleotides. The two ends of a DNA strand, referred to as the 5′ and 3′ ends, are chemically different. DNA sequences are conventionally represented starting with the 5′ nucleotide end. The interactions between different strands are predictable based on sequence..

2. DNA SYNTHESIS

Replication is a process where strands of the DNA molecule separate, and a complement to each strand is synthesized. Thus one strand of DNA can be used as a template for a new strand of DNA. The natural environment of the cell contains the necessary ingredients for this to occur: free floating nucleotides (G, C, A, and Ts that will make up the new strand), an enzyme called DNA polymerase that help the nucleotides find the correct location along the template strand, and other proteins that assist in the process. This process is called "semi-conservative" because the each new DNA molecule consists of one new strand and one original strand that served as the template.

Polymerase Chain Reaction is a method for exponentially amplifying the concentration of selected sequences of DNA within a pool. A PCR reaction requires: substrates, template, primer and enzymes.

Substrates

Four deoxyribonucleotide triphosphates (dNTP's) are required for DNA synthesis. These are dATP, dGTP, dTTP and dCTP. The high energy phosphate bond between the a and b phosphates is cleaved and the deoxynucleotide monophosphate is incorporated into the new DNA strand.

Template

The nucleotide that is to be incorporated into the growing DNA chain is selected by base pairing with the template strand of the DNA. The template is the DNA strand that is copied into a complementary strand of DNA.

Primer

The enzyme that synthesizes DNA, DNA polymerase, can only add nucleotides to an already existing strand or primer of DNA or RNA that is base paired with the template.

Enzymes

An enzyme, DNA polymerase, is required for the covalent joining of the incoming nucleotide to the primer. To actually initiate and sustain, DNA replication requires many other proteins and enzymes which assemble into a large complex called a replisome. It is thought that the DNA is spooled through the replisome and replicated as it passes through.

DNA replication is a semi-conservative process in which a DNA polymerase uses one DNA strand as a template for the synthesis of a second, complementary, DNA strand [5]. Initiation of Ad DNA replication occurs at either DNA end and proceeds by a strand-displacement mechanism giving rise to type I and type I1 molecules [6] A model first proposed by [7] suggested that a free molecule of the TP could act as a primer for the initiation of replication by formation of a covalent linkage with dCMP, the 5 ' -terminal nucleotide, that would provide the 3 ' -OH group needed for elongation by the DNA polymerase. The 5 ' end of each nascent daughter strand is covalently linked to a protein of 80 kD [8, 9], structurally related to the 55-kD protein covalently linked to the 5 ' ends of Ad DNA.

Another viral protein, the DNA-binding protein (DBP), binds ssDNA and dsDNA and stimulates the initiation of replication [10]. An α-helical structure located in the amino-terminal region of protein p6 is involved in binding DNA through its minor groove [11]. A p6 dimer binds 24 bp, bending or kinking the DNA every 12 bp. Protein p6 is a 123-amino-acid protein that stimulates initiation of ϕ29 DNA replication by reducing the Km for dATP and facilitating the transition from initiation to elongation [12]. Stimulation by p6 is due to formation of a nucleoprotein complex that spans 200-300 bp from each DNA end [13, 14]. p6 binding results in a 4.2- fold compaction of the DNA in which one superhelical turn has 63 bp (2.6 protein p6 dimers) with a pitch of 5.1 nm and a diameter of 6.6 nm [15]

Reverse Transcription:

Hepadnaviruses contain a 3-kb circular, dsDNA genome in which the minus (uncoding) strand is complete with unique 5 and 3 termini. The mechanism by which hepadnavirus DNA replicates involves reverse transcription [16].

To synthesize full length minus-strand DNA, the primer and RT must be repositioned at the 3' end of the viral RNA. The steps involved are (i) initiation of minus strand cDNA synthesis (ii) reposition of minus strand (iii) full length minus strand cDNA synthesis of (+) strand DNA synthesis (iv) Synthesis of double stranded viral DNA [17, 18]. Full-length plus-strand DNA is subsequently assembled by the action of DNA ligase. The final product of reverse transcription is a linear viral DNA flanked by two LTRs, which then integrates into the host chromosome in a reaction that is dependent on the viral integrase [19, 20, 21]

3. DNA STORAGE

To store data in DNA, the concept is the same, but the process is different. DNA molecules are long sequences of smaller molecules, called nucleotides – adenine, cytosine, thymine and guanine, usually designated as A, C, T and G. Rather than creating sequences of 0s and 1s, as in electronic media, DNA storage uses sequences of the nucleotides. There are three basic codes for storing data in DNA [22] which are Huffman code, comma code and the alternating code. A DNA storage system consists of a DNA synthesizer that encodes the data to be stored in DNA, a storage container with compartments that store pools of DNA that map to a volume, and a DNA sequencer that reads DNA sequences and converts them back into digital data. Figure 1 [27] shows an overview of the integrated system.

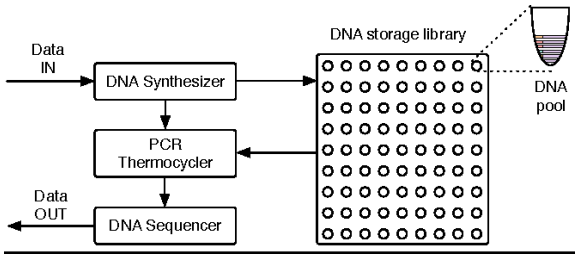

Figure 1. DNA strorage system

There are different types of errors associated with DNA data storage systems which are physical errors and genetic errors. Physical errors occur during synthesis and sequencing of DNA and genetic errors are caused by mutations which occur naturally during evolution and prolongation. Error can be insertion, deletion or substitution of single base in DNA sequences. Substitution of single base can be considered as bit flip errors. Other type of error can be deletion of bunch of DNA nucleotides categorized as burst error. Reading error rates ranges from 1-3 % while writing error has error rates upto 15 %. Error models proposed so far

have focus on physical errors like substitution and deletion of oligonucleotides [23, 24] but no work has been done on insertion error model. There are three basic codes for storing data in DNA [22] which are Huffman code, comma code and the alternating code.
DNA has essential property for long term archival of data compare to digital storage devices. To witness long term storage of DNA and improve the DNA stability, researchers have develop chemical based [25] method to encapsulate DNA into glass sphere and preserve it from environmental damage for long term archival

4. DNA SEQUENCING

The DNA sequencing procedure determines the nucleotide sequence of a terminally labeled DNA molecule by breaking it at adenine, guanine, cytosine, or thymine with chemical agents. Partial cleavage at each base produces a nested set of radioactive fragments extending from the labeled end to each of the positions of that base. Polyacrylamide gel electrophoresis resolves these single-stranded fragments; their sizes reveal in order the points of breakage. The autoradiograph of a gel produced from four different chemical cleavages, each specific for a base shows a pattern of bands from which the sequence can be read directly.
The sequencing requires DNA molecules, either double stranded or single-stranded, that is labeled at one end of one strand with 32p. This can be a 5' or a 3' label. There are then two strategies: either (i) the double-stranded molecule is cut by a second restriction enzyme and the two ends are resolved on a polyacrylamide gel and isolated for sequencing or (ii) the doubly labeled molecule is denatured and the strands are separated on a gel [26],
DNA can be sequenced by a chemical procedure that breaks a terminally labeled DNA molecule partially at each repetition of a base. The lengths of the labeled fragments then identify the positions of that base. When the products of these four reactions are resolved by size, by electrophoresis on a polyacrylamide gel, the DNA sequence can be read from the pattern of radioactive bands [27].

5. DNA SEQUENCING METHODS

5.1. Sanger's method (enzymic method)

It consisted of a catalysed enzymic reaction that polymerizes the DNA fragments complementary into the template DNA of interest. The enzymic method for DNA sequencing has been used for genomic research as the main tool to generate the fragments necessary for sequencing, regardless of the sequencing strategy [28, 29].

5.1.1. Random approach
Genomic DNA is randomly fragmented into smaller pieces, normally ranging from 2 to 3 kb. The fragments, inserted into a vector, are replicated in a bacterial culture. Several positive amplifications are selected, and the DNA is extensively sequenced [30, 31].

5.1.2. Direct approach
An unknown sequence of DNA is inserted into a vector and amplified. This approach is known as primer walking, and its major advantage is the reduced redundancy [32, 33, 34]

5.1.3. Enzyme technology
Polymerase chain reaction (PCR) was discovered and the use of a heat-stable DNA polymerase from Thermus aquaticus (Taq polymerase), increased the ability to perform sequencing reactions (cycle-sequencing) with reduced amounts of DNA template compared to isothermal enzymes became possible [35, 36, 37]

5.1.4. Sample preparation
Sample preparation included the following steps: (i) DNA scission and cloning into a vector (ii) vector amplification to produce a phage-infected culture; and (iii) purification from the cell culture to yield pure single-stranded [38, 39, 40].

5.2. Labels and DNA labelling

5.2.1. Radioisotopes:
It used deoxyadenosine 5'-(α-[^{35}S]thio)triphosphate as the label incorporated into the DNA fragments. This strategy resulted in an increase in band sharpness on autoradiography as well as in the resolution of band separation [41].

5.2.2. Chemiluminescent detection:
In this system, the 5'-end of an oligonucleotide linked to biotin was used as the primer in the sequencing reaction. The advantages are the sequencing reactions were obtained directly from the PCR products; this method does not require cloning of the DNA before sequencing, and it was possible to multiplex several reactions on the same gel and detect one at a time with appropriate enzyme linked primers [42, 43, 44, 45].

5.2.3. Fluorescent dyes:
Different dyes like fluorescein, 4-chloro-7-nitrobenzo-2-1-diazole (NBD), tetramethyl-rhodamine, and Texas Red where used to automate DNA sequence [46, 47]

5.3. Fragment separation and analysis

5.3.1. Electrophoresis:

The separation of labelled DNA fragments is done using by polyacrylamide gel electrophoresis to complete automation of the enzymic DNA sequencing method [48, 49].

5.3.2. Mass spectrometry – an alternative:

Mass spectrometry (MS) has been viewed as the technique to allow the sequencing of hundreds of bases in a few seconds. Matrix-assisted laser desorption ionization–time of flight (MALDI–TOF) MS, and electrospray ionization (ESI) MS are two of the most suitable MS techniques for sequencing DNA using the Sanger method [50, 51].

5.4. Maxam and Gilbert method (chemical method)

In this method, end-labelled DNA fragments are subjected to random cleavage at adenine, cytosine, guanine, or thymine positions using specific chemical agents. The chemical attack is based on three steps: base modification, removal of the modified base from its sugar, and DNA strand breaking at that sugar position[52]. The products of these four reactions are then separated using polyacrylamide gel electrophoresis. The sequence can be easily read from the four parallel lanes in the sequencing gel.

5.5. Pyrosequencing method (using pyrophosphate)

Pyrosequencing is a real-time DNA-sequencing method based on the detection of the PPi released during the DNA polymerization reaction [53, 54], Initially, this approach was used for continuous monitoring of DNA polymerase activity [55].there are two different pyrosequencing approaches: solid-phase sequencing [56] and liquid-phase sequencing [57]

5.6. Single molecule sequencing method (Single strand DNA)

Single-molecule sequencing was initially conceived as a laser-based technique that allows the fast sequencing of DNA fragments of 40 kb or more at a rate of 100–1000 bases per second [58]. This technique is based on the detection of individual fluorescent nucleotides in a flowing sample stream [58, 60].

5.7. Single molecule SMRT(TM) sequencing

SMRT sequencing is based on the sequencing by synthesis approach. The DNA is synthesised in so called zero-mode wave-guides (ZMWs) - small well-like containers with the capturing tools located at the bottom of the well. The sequencing is performed with use of unmodified polymerase and fluorescently labelled nucleotides flowing freely in the solution. The wells are constructed in a way that only the fluorescence occurring by the bottom of the well is detected. The fluorescent label is detached from the nucleotide at its incorporation into the DNA strand, leaving an unmodified DNA strand [61].

5.8. Single molecule real time (RNAP) sequencing

This method is based on RNA polymerase (RNAP), which is attached to a polystyrene bead, with distal end of sequenced DNA is attached to another bead, with both beads being placed in optical traps. RNAP motion during transcription brings the beads in closer and their relative distance changes, which can then be recorded at a single nucleotide resolution [62]. The sequence is deduced based on the four readouts with lowered concentrations of each of the four nucleotide types.

5.9. Large-scale sequencing strategies

Large scale sequencing aims at sequencing very long DNA pieces, such as whole chromosomes. It consists of cutting (with restriction enzymes) or shearing (with mechanical forces) large DNA fragments into shorter DNA fragments. The fragmented DNA is cloned into a DNA vector, and amplified in E.coli. Short DNA fragments purified from individual bacterial colonies are individually sequenced and assembled electronically into one long, contiguous sequence. This method does not require any pre-existing information about the sequence of the DNA and is referred to as de novo sequencing.

The high demand for low-cost sequencing has driven the development of high-throughput sequencing technologies that parallelize the sequencing process, producing thousands or millions of sequences at once. High-throughput sequencing technologies are intended to lower the cost of DNA sequencing beyond what is possible with standard dye-terminator methods High throughput sequencing [63].

5.10. Lynx therapeutics' massively parallel signature sequencing (MPSS)

MPSS is an ultra high throughput sequencing technology. When applied to expression profile, it reveals almost every transcript in the sample and provides its accurate expression level. MPSS was a bead-based method that used a complex approach of adapter ligation followed by adapter decoding, reading the sequence in increments of four nucleotides [64].

5.11. Polony sequencing

It is an inexpensive but highly accurate multiplex sequencing technique that can be used to read millions of immobilized DNA sequences in parallel. It combined an in vitro paired-tag library with emulsion PCR, an automated microscope, and ligation-based sequencing chemistry to sequence an E. coli genome [65].

5.12. Illumina (Solexa) sequencing

Solexa developed a sequencing technology based on dye terminators. In this, DNA molecule are first attached to primers on a slide and amplified, this is known as bridge amplification. Unlike pyrosequencing, the DNA can only be extended one neucleotide at a time. A camera takes images of the fluorescently labelled nucleotides, then the dye along with the terminal 3' blocker is chemically removed from the DNA, allowing the next cycle [66].

5.13. SOLiD sequencing

The technology for sequencing used in ABISolid sequencing is oligonucleotide ligation and detection. In this, a pool of all possible oligonucleotides of fixed length is labelled according to the sequenced position [67]. This sequencing results to the sequences of quantities and lengths comparable to illumine sequencing.

5.14. DNA nanoball sequencing

It is high throughput sequencing technology that is used to determine the entire genomic sequence of an organism. The method uses rolling circle replication to amplify fragments of genomic DNA molecules. This DNA sequencing allows large number of DNA nanoballs to be sequenced per run and at low reagent cost compared to other next generation sequencing platforms [68].
There is different possible DNA sequencing methods used while retrieving the sequence back from the library storage

6. CONCLUSION

Different techniques are in progress to get effective synthesis and sequencing with less error rates. Digital computer files can be quite large – even terabytes in size for large databases. Remarkable work was done by team of student at Chinese Hong Kong University using E.coli as medium to store the data. It has storage capacity of 4502, 000- gigabyte hard disks per gram of bacteria. This technology has many advantages over the magnetic data storage medium mainly information cannot be hacked and can defend against cyber attacks which points to higher data security than computer storage.
Reverse transcriptase has a high error rate when transcribing RNA into DNA since, unlike most other DNA polymerases, it has no proofreading ability. This high error rate allows mutations to accumulate at an accelerated rate relative to proofread forms of replication. Also time taken to store data in DNA is too long. These are the challenges researchers are still working on.

REFERENCES

[1] G. M. Church, Y. Gao, and S. Kosuri. Next-generation digital information storage in DNA. Science, 337(6102):1628, 2012.
[2] C. T. Clelland, V. Risca, and C. Bancroft. Hiding messages in DNA microdots. Nature, 399:533–534, 1999
[3] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust, B. Sipos, and E. Birney. Towards practical, high-capacity, low maintenance information storage in synthesized DNA. Nature, 494:77–80, 2013.
[4] A review of DNA sequencing techniques. Lilian T. C. Francia[1], Emanuel Carrilho[2] and Tarso B. L. Kist[3], Quarterly Reviews of Biophysics 35, 2 (2002), pp. 169–200. " 2002 Cambridge University Press DOI: 10.1017/S0033583502003797 Printed in the United Kingdom.
[5] Kornberg, A. and T. Baker. 1992. DNA replication. W.H. Freeman, New York.
[6] Kelly, R.E., M.L. DeRose, B.W. Draper, and G.M. Wahl. 1995. Identification of an origin of bidirectional replication within the coding region of the ubiquitously ex- pressed CAD gene. Mol. Cell. Biol. 15: 4136-4148.
[7] Romero, A., R. Lbpez, R. Lurz, and P. Garcia. 1990. Temperate bacteriophages of Strep- tococus pneumoniae that contain protein covalently linked to the 5 ' ends of their DNA.J. Virol. 64: 5149-5155.
[8] Challberg, M.D., S.V. Desiderio, and T.J. Kelly. 1980. Adenovirus DNA replication in vitro: Characterization of a protein covalently linked to nascent DNA strands. Proc. Natl. Acad. Sci. 77: 5105-5109.
[9] Lichy, J. H., M. S. Horwitz, and J. Hurwitz. 1981. Formation of a covalent complex be- tween the 80,000-dalton adenovirus terminal protein and 5 ‘-dCMP in vitro. Proc. Natl. Acad. Sci. 78: 2678-2682.
[10] Salas, M., J. MCndez, J.A. Esteban, M. Serrano, C. Gutierrez, J.M. Hermoso, A. Bravo, M.S. Soengas, J.M. Lbzaro, M.A. Blasco, R. Freire, A. Bernad, J.M. Sogo, and L. Blanco. 1993. Terminal protein priming of DNA replication: Bacteriophage ϕ29 as a Priming DNA Synthes model system. In Virus strategies, molecular biology and pathogenesis (ed. W. Doer- fler and P. Bohm), pp. 3-19. Verlag Chemie. Weinheim, Germany
[11] Freire, R., M. Salas, and J.M. Hermoso. 1994. A new protein domain for binding to DNA through the minor groove. EMBO J. 13: 4353-4360.

[12] Blanco, L., A. Bernad, and M. Salas. 1988. Transition from initiation to elongation in the protein-primed ϕ29 DNA replication. Salt-dependent stimulation by the viral protein p6. J. Virol. 62: 41 67-41 72.
[13] Prieto, I., M. Serrano, J.M. Lizaro, M. Salas, and J.M. Hermoso. 1988. Interaction of the bacteriophage ϕ29 protein p6 with double-stranded DNA. Proc. Natl. Acad. Sci. 85 314- 318
[14] Serrano, M., J. GutiCrrez, I. Prieto, J.M. Hermoso, and M. Salas. 1989. Signals at the bacteriophage ϕ29 DNA replication origins required for protein p6. binding and activity. EMBO J. 8: 1879-1885.
[15] Soengas, M.S., C. Gutitrrez, and M. Salas. 1995. Helix-destabilizing activity of $29 single-stranded DNA binding protein: Effect on the elongation rate during strand displacement DNA replication. J. Mol. Biol. 253: 517-529.
[16] J. N. Zadeh, B. R. Wolfe, and N. A. Pierce. Nucleic acid sequence design via efficient ensemble defect optimization. Journal of Computational Chemistry, 32(3):439–452, 2011.
[17] Weiss, R., N. Teich, H. Varmus, and J. Coffin, eds. 1985. RNA tumor viruses, 2nd edi- tion, parts 1 and 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
[18] Coffin, J.M. 1990. Retroviridae and their replication. In Virology, 2nd edition (ed. B.N. Fields et al.), pp. 1437-1500. Raven Press, New York.
[19] Brown, B., B. Bowerman, H.E. Varmus, and J.M. Bishop. 1987. Correct integration of retroviral DNA in vitro. Cell 49: 347-356.
[20] Craigie, R., T. Fujiwara, and F. Bushman. 1990. The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. Cell 62: 829-837.
[21] Katz, R.A., G. Merkel, J. Kulkosky, J. Leis, and A.M. Skalka. 1990. The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro. Cell 63: 87-95
[22] Secret Signatures Inside Genomic DNA Masanori Arita, Yoshiaki Ohashi, Biotechnology progress, Volume20, Issue5 2004 Pages 1605-1607
[23] David Haughton, Félix Balado: Repetition Coding as an Effective Error Correction Code for Information Encoded in DNA. BIBE 2011: 253-260
[24] Codes for DNA Sequence Profiles Han Mao Kiah, Gregory J. Puleo, and Olgica Milenkovic, Senior Member, IEEE.
[25] Robust chemical preservation of digital information on DNA in silica with error-correcting codes. Grass RN[1], Heckel R, Puddu M, Paunescu D, Stark WJ. Angew Chem Int Ed Engl. 2015 Feb 16;54(8):2552-5. doi: 10.1002/anie.201411378. Epub 2015 Feb 4.
[26] Hayward, G. S. (1972) Virology 49,342-344.
[27] A new method for sequencing DNA, Allan M, Maxam and Walter Gilbert Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138 Contributed by Walter Gilbert, December 9,1976
[28] Sanger F & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. molec. Biol. 94, 441–448.
[29] Griffen, H. G. & Griffen, A. M. (1993). DNA sequencing – recent innovations and future trends. Appl. Biochem. Biotechnol. 38, 147–159.
[30] Lander, E.S, Linton, L. M. Birren, B., et al. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860–921.
[31] Venter J. C., Adam M. D., Myers, E. W., et al. (2001). The sequence of the Human Genome. Science 291, 1304–1351.
[32] Studier, F. W. (1989). A strategy for high-volume sequencing of cosmid DNAs – random and directed priming with a library of oligonucleotides. Proc. natn. Acad. Sci. USA 86, 6917–6921.
[33] Martin-Gallardo, A., McCombite, W. R. & Gocayne, J. D. (1992). Automated DNA sequencing and analysis of 106 kilobases from human. Nat. Genet. 1, 34–39.
[34] Voss H. Weiman S. (1993). Automated low-redundancy large-scale DNAsequencing by primer walking. BioTechniques 15, 714–721.
[35] Tabor, S. & Richarson, C. C. (1987). DNA-sequence analysis with a modified bacteriophage-T7 DNApolymerase. Proc. natn. Acad. Sci. USA 84, 4767–4771.
[36] Tabor, Huber. E. & Richardson, C. C. (1987). Escherichia coli thioredoxin confers processivity on the DNA-polymerase activity of the gene-5 protein of bacteriophage-T7. J. biol. Chem. 262, 16212–16223.
[37] Mullis, K. & Fallona, F. A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Meth. Enzym. 155, 335–350
[38] Martin, W. J. & Davies, W. R. (1986). Automated DNA sequencing: progress and prospects. Biotechnology 4, 890–895
[39] Ahemad, A. (1984). Use of transposon-promoted deletions in DNA sequence analysis. J. molec. Biol. 178, 941–948.
[40] Deininger, P. L. (1983). Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. Analyt. Biochem. 129, 216–223.
[41] Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983). Buffer gradient gels and $&S label as an aid to rapid DNA sequence determination. Proc. natn. Acad. Sci. USA 80, 3963–3965.
[42] Beck, S. & Alderton, R. P. (1993). A strategy for amplification, purification, and selection of M13 templates for large-scale DNA sequencing. Analyt. Biochem. 212, 498–505.
[43] Gillivet, P. M. (1990). Chemiluminescent multipex DNA sequencing. Nature 348, 657–658
[44] Olsen, C. E. M., Martin, C. S. & Bronstein, I. (1993). Chemiluminescent DNA sequencing with multiplex labeling. BioTechniques 15, 480–485
[45] Cherry, J. L., Young, H. H., Disera, L. J. (1994). Enzyme-linked fluorescent detection for automated multiplex DNAsequencing. Genomics 20, 68–74.

[46] Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P. (1986). Fluorescence detection in automated DNA sequence analysis. Nature 321, 674–679
[47] Adams, M. D., Field, C. & Venter, J. C. (1996). Automatic DNA Sequencing and Analysis. San Diego: Academic Press
[48] Garroff, H. & Ansorge, W. (1981). Improvements of DNA sequencing gels. Analyt. Biochem. 115, 450–457. (1990). Chemiluminescent multiplex
[49] Kostechka, A. J., Merchbanks, M. L. & Smith, L. M. (1992). High speed automated DNA sequencing in ultrathin slab gels. Biotechnology 10, 78–81
[50] Karas, M. & Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10000 daltons. Analyt. Chem. 60, 2301– 2303.
[51] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. (1990). Electrospray ionizationprinciples and practice. Mass Spectrom. Rev. 9, 37–70
[52] Maxam, A. M. & Gilbert, W. (1977). A new method for sequencing DNA. Proc. natn. Acad. Sci. USA 74, 560–564
[53] Nyren, P. & Lundin, A. (1985). Enzymatic method for continuos monitoring of inorganic pyrophosphate synthesis. Analyt. Biochem. 151, 504–509
[54] Hyman, E. D. (1988). A new method of sequencing DNA. Analyt. Biochem. 174, 423–436
[55] Nyren, P. (1987). Enzymatic method for continuous monitoring of DNA polymerase activity. Analyt. Biochem. 167, 235–238.
[56] Ronaghi, M., Karamohemad, S., Petersson, B., Uhlen, M. & Nyren, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. Analyt. Biochem. 242, 84–89.
[57] Ronaghi, M., Uhlen, M. & Nyren, P. (1998a). A sequencing method based on real-time pyrophosphate. Science 281, 363–365
[58] Jett, J. H., Keller, R. A., Martin, J. C., Marron, B. L., Moyzis, R. K., Ratlife, R. L. (1989). Highspeed DNA sequencing – an approach based upon fluorescence detection of single molecules. J. biomolec. struct. Dyn. 7, 301–309.
[59] Shera, E. B., Seitzinger, N. K., Davis, L. M., Keller, R. A. & Soper, S. A. (1990). Detection of single fluorescent molecules. Chem. Phys. Lett. 174, 553–557
[60] Harding, J. D. & Keller, R. A. (1992). Singlemolecule detection as an approach to rapid DNA sequencing. Trends Biotechnol. 10, 55–57.
[61] Metzker, M.L. (2010). Sequencing technologies - the next generation. Nature Reviews 11(1): 31-46.
[62] Marguerat, S. & Bahler, J. (2010). RNA-seq: from technology to biology. Cellular & Molecular Life Sciences 67(4): 569–579
[63] Gore, M.A. et al (2009b). Large-scale discovery of gene-enriched SNPs. The Plant Genome 2(2): 121-133.
[64] Frommer, M. et al (1992). A genomic sequencing protocol that yields a positive display of 5methylcytosine residues in individual DNA strands. Proc. Natl. Acad. Sci. USA 89(5): 1827–1831
[65] Krober, M. et al (2009). Phylogenetic characterization of a biogas plant microbial community integrating clone library 16S-rDNA sequences and metagenome sequence data obtained by 454-pyrosequencing. Journal of Biotechnology 142(1): 38-49.
[66] Mardis, E.R. (2010). A decade's perspective on DNA sequencing technology. Nature 470(7333): 198-203
[67] Morozova, O. & Marra, M.A. (2008). Applications of next-generation sequencing technologies in functional genomics. Genomics 92(5): 255-264.
[68] Schöb, H. & Grossniklaus, U. (2006). The first high-resolution DNA "methylome". Cell 126(6): 1025-1028.