مجلة الدراسات التربوية والعلمية - كلية التربية - الجامعة العراقية
العدد الخامس والعشرون - المجلد الثاني عشر - علوم الحياة - آب 2025 م
doi.org/10.52866/esj.2023.12.25.13

DNA sequencing technologies: A review article

Mohammad Abdul Rahmman Mohammad AL-Maeni , Shaymaa Fouad Rasheed Al-Khazraji
Asst. prof. Asst. prof.
Biology Department, College of Science, University of Baghdad, Baghdad, Iraq
Corresponding Email: mhogene@yahoo.com, Tel: 07725578767
ORCID: https://orcid.org/0000-0001-6662-8827

Abstract :

The sequencing technology is a powerful tool in molecular biology and has tremendous applications in many area of life sciences. Detection of genome variation, SNPs, type of mutation and achievement of genomes comparison are some applications of sequencing technology in field of molecular biology. This review aimed to explain sequencing technology from its fancy as first generation sequencing to the most developed sequencing technology that represented by third generation sequencing technology, in addition, to compare the features of different sequencing techniques.

KeyWords: Maxam and Gilbert , Sanger , Illumina, Pacbio and nanopre sequencing .

تقنيات تسلسل الحمض النووي: مقالة

محمد عبد الرحمن محمد المعيني , شيماء فؤاد رشيد الخزرجي
مساعد. البروفيسور مساعد. البروفيسور
قسم الأحياء، كلية العلوم، جامعة بغداد، بغداد، العراق

مستخلص:

تُعدّ تقنية التسلسل أداةً فعّالة في علم الأحياء الجزيئي، ولها تطبيقاتٌ واسعةٌ في العديد من مجالات علوم الحياة. ومن تطبيقاتها في مجال علم الأحياء الجزيئي، الكشف عن تباين الجينوم، وتعدد أشكال النوكليوتيدات المفردة (SNPs)، ونوع الطفرات، ومقارنة الجينومات. تهدف هذه المراجعة إلى شرح تقنية التسلسل، بدءًا من بداياتها كتقنية الجيل الأول، وصولًا إلى أكثر تقنيات التسلسل تطورًا، والمتمثلة في تقنية الجيل الثالث، بالإضافة إلى مقارنة خصائص تقنيات التسلسل المختلفة.

Introduction

There are a lot of sequencing technologies that have been discovered with different principles, starting with first generation sequencing and ending with third generation sequencing. Sequencing technology started as chemical method using Maxam and Walter Gilbert method, however due complexity of this process and its less scalable through comparison, its replaced by Sanger method which relies on principle of sequencing by synthesis [1]. Sanger method is replaced by the next generation sequencing technique due to high cost, low throughput and time consuming method and the first NGS technique was Roche 454 pyrosequencing in 2005. Then Illumina/ Solexa was started in 2006 in sequencing with novel idea with 30 million reads and 150-250 length of reads. After that, Ion Torrent was introduced in 2010 with higher speed compared with other techniques and lower cost with 400bps length of read[2]. In 2010, third generation technique was emerged as a new technique by Helicos Bioscience that does not rely on PCR amplification (by using thermal cycler) and no error prone with PCR and with length of read more than 5 kbps. In 2011, PacBio released with less error and larger length of reads that is represented by Sequel System with different types. More recently, in 2014 is released with larger length of reads, higher speed and low cost[3].

First generation sequencing: Maxam and Gilbert sequencing

Principle of sequencing

The first sequencing technique raised between 1976–1977 relying on principle of modification and cleaving at specific site of DNA, this treatment generates breaks at different positions which are G, A+G, C, C+T and carried out in four reactions. Chemical modification comprises depurinate of (A+G) using formic acid and methylation of guanines (G) using dimethyl sulfate while hydrolyzing (C+T) using hydrazine and addition of sodium chloride leads to hydrolyze (C) only without thymine. The modified bases then were subsequently be cleaved in four different reaction G, A+G, C, C+T using piperidine. Prior to modification and cleavage treatment, the

5′ end of fragmented DNA is labeled with radioactive isotope ^{32}P therefore different labeled fragments are generated in different reaction of G, A+G, C, C+T. These fragments were passed on gel electrophoresis (acrylamide gel) in four line of single gel and then visualized though displaying the gel with X-ray film. The bands on gel were read from bottom to the top to detect the sequence from four reactions. On gel, To detect C base, the band should appear at C and C+T lines while to detect T base, the band should only be appeared at C+T line. On same way we detect the base as it is G or A (Figure 1)[4] .

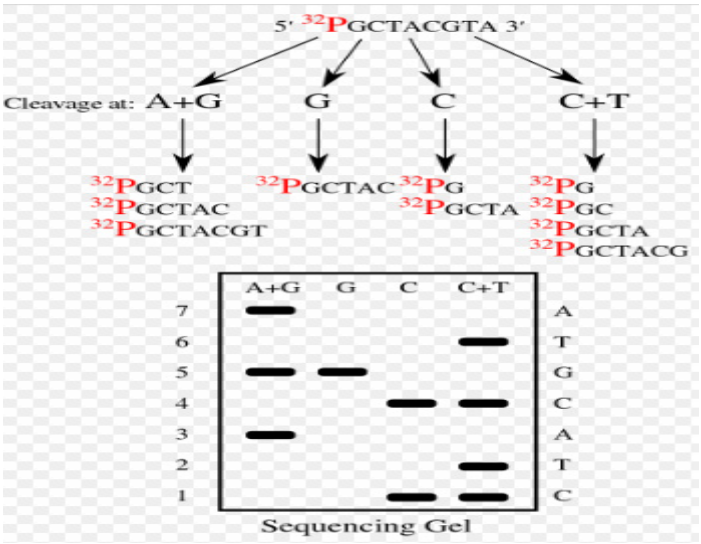

Figure 1: Sequencing Technique using Maxam and Gilbert method (adopted from Wikipedia)

Modification and cleavage reaction

Ideally, the cleavage carried out at two main reactions, the first reaction cleavage at G and A while the second reaction cleavage at C and T. Modification on first and second reactions is achieved to cleavage at G and C only respectively. What generally happens is that before cleavage the base gets modified and then the modified base is removed from its sugar, after that the bond that carry the sugar in backbone of DNA is destroyed. To cleave G only, G gets methylated by addition methyl group and the base G removed by heating while its sugar is removed from backbone of DNA by heating

with alkali. **dilute acid** is added in case to cleave at both A and G. To cleave both C and T, hydrazine used for remove the bases while piperidine used for cleaving backbone. addition of sodium chloride leads to cleave (C) only without thymine [5].

Labelling the stand at 5′ end of fragmented DNA leads to produce two fragments at each site of cleavage, one is labelled while another was unlabeled. Example of cleavage at G site will produce label fragment stared at 5′ end and ending with base that comes before the site of a G. In this case labeled fragment will only be appeared on gel after exposing the gel to x-ray film (Figure 2) [5].

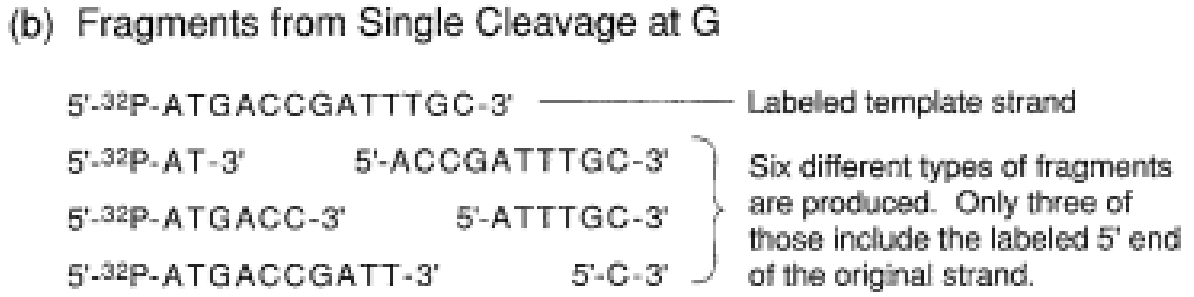

Figure 2: Schematic representation of labeled and unlabeled fragment cleaved at G site (adopted from Heather and Chain, 2016)

There are many disadvantages of using this technique, these are as follows

1 – This technique requires using of many chemicals which are hazardous for human

2 - This technique is so expensive and complex

3 - The length of fragment that can be sequenced using this technique is not more than 500 bps[6]

First generation sequencing: Sanger sequencing

The simple idea behind Sanger sequencing is to amplify the target DNA and terminate the amplification reaction within the template at each base that is carried out in four different reactions. each reaction specifies with a particular base A, G, C, T therefore the product is different fragments ended at specific base at four different reaction

relying on type of base that the reaction will terminate on. The termination of the reaction is achieved due to the addition of ddNTP that lack OH at the 3 prime of sugar of nucleotide therefore DNA polymerase is unable to form the phosphodiester bond between ddNTP nucleotide and the next nucleotide to be added to the growing chain Figure 3[7].

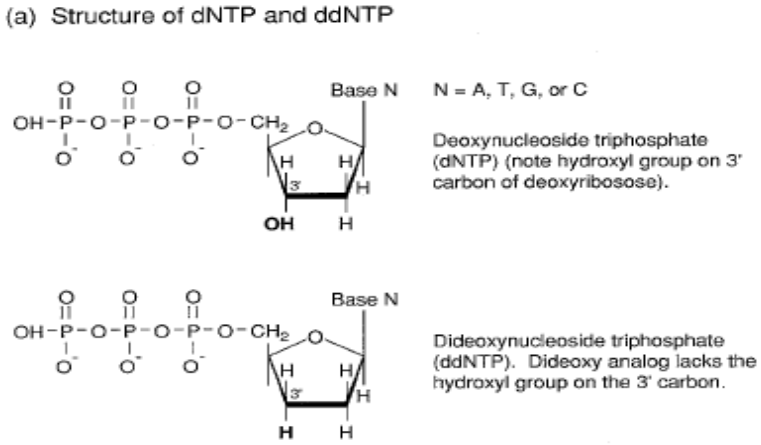

Figure 3: The difference in the structure of dNTP and ddNTP (adopted from Sourav Pan, 2024)

In brief, the primers are added along with DNA polymerase to the template to be sequenced, the addition of dNTP by DNA polymerase will continuous the amplification of the template, however, the addition of ddNTP instead of dNTP to the growing chain lead to terminate the reaction. Suppose you have in the template T base so, the DNA polymerase added ddATP instead of dATP and this lead to terminate the reaction at reaction carried out in the first tube therefore we will have the first band and because we have four different ddNTP nucleotides we carried the reaction in four different tubes, each tube specifies with specific ddNTP. Therefore, the components that added in each tube are DNA polymerase, dNTP and ddNTP and many copies of template and primer. The produced fragments in four different tubes were passed through four line of single gel and the sequence is read from bottom of gel to the top after exposed the gel to x-ray film in autoradiography technique.

The idea of Sanger sequencing is improved lately so the reaction is carried out in single tube but with different fluoresce labeled ddNTP, each color correlated with specific ddNTP. As DNA polymerase added specific fluoresce labeled ddNTP that is complementary to the base in template strand and terminate the reaction, the formed band will migrate through capillary gel electrophoresis (single line gel) and pass through window thereby detector and laser located at the end of the capillary gel. The laser shine the band formed due to the presence of fluoresce labeled ddNTP and detector detect the color on fluoresce labeled ddNTP and recorded as a peak with specific color so each peak with specific color corresponds to specific ddNTP therefore the sequence is recorded as chromatogram file Figure 5.

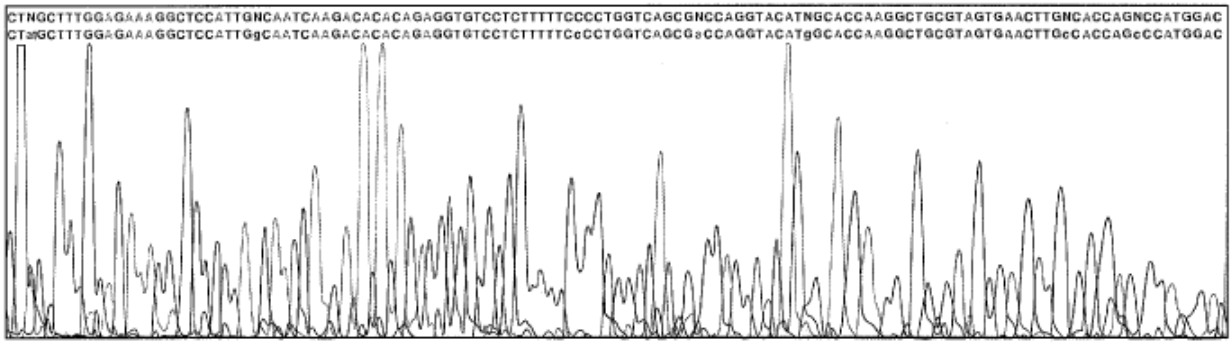

Figure 5: Showing chromatogram file that is produced through Sanger sequencing

Before sequencing of template, we need primer for known sequence therefore we can clone the template with Ml 3 cloning vector so the flanking regions of the template that is inserted within cloning vector M13 will be known therefore we can easily design primer Figure 6 [8, 9].

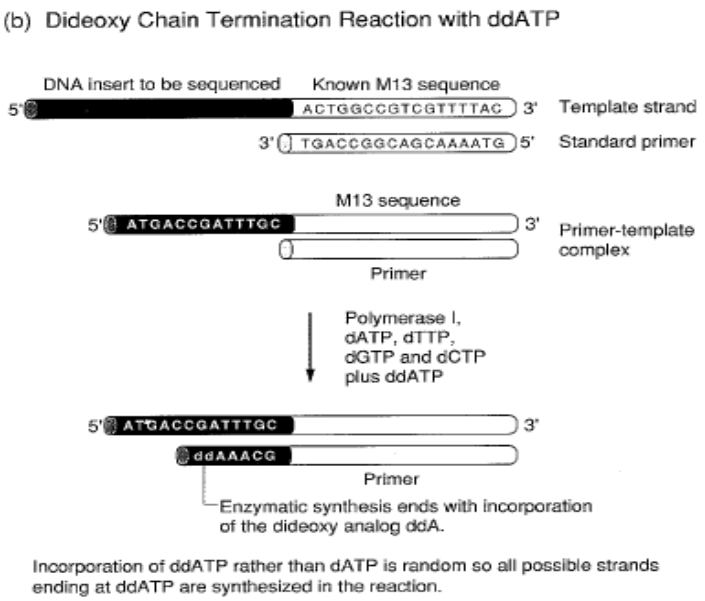

Figure 6: M13 vector used to clone the template and select primer from flanking regions of M13 vector for sequencing

Sanger sequencing can only detect the sequence for fragment up to 1 kbps and the reason for that Sanger sequencing relying on separation of different bands through gel and resolution of separation is not accurate for the fragment less than 25-40 bps and more than 900-1000 bps table 1[10].

Technology	Number of lanes	Injection volume (nL)	Analysis time	Average read length	Throughput (including analysis; Mb/h)	Gel pouring	Lane tracking
Slab gel	96	500–1000	6–8 hours	700 bp	0.0672	Yes	Yes
Pacific Biosciences SMRT (2019)			10–20 hours	10–30 kb	1,300		
Oxford Nanopore MinIon (2019)			3 days	13–20 kb[18]	700		
Microchip	96	0.1–0.5	6–30 minutes	430 bp	0.660	No	No
Ion Torrent Ion 530 (2019)			2.5–4 hours	200–600 bp	110–920		
Illumina/Solexa (2008)			2–3 days	30–100 bp	20		
Illumina NovaSeq (2019)			1–2 days	2x50–2x150 bp	22,000–67,000		
Illumina MiSeq (2019)			1–3 days	2x75–2x300 bp	170–250		
Capillary array electrophoresis	96	1–5	1–3 hours	700 bp	0.166	No	No
BGI MGISEQ-T7 (2019)			1 day	2x150 bp	250,000		
ABI/SOLiD (2008)			8 days	35 bp	5–15		
454/Roche FLX (2008)		< 0.001	4 hours	200–300 bp	20–30		

Table 1: Average length of read for different instruments used in sequencing technology

Second generation sequencing: Pyrosequencing

Second generation sequencing started at Pyrosequencing technique with 100-fold higher throughput than Sanger sequencing with reducing in cost with 25%. The principle relies on sequencing by synthesis, so the template strand is copied by DNA polymerase through adding complementary nucleotide and this leads to release pyrophosphate for each one of nucleotide added. Pyrophosphate contribute in enzymatic reaction with final product represented by light signals. Then camera captured the light and recorded the added nucleotide. The nucleotide A, T, C, and G added one by one and monitoring and recorded the light signal for the complementary nucleotide to the template strand while excess nucleotides were washed through addition apyrase enzyme which degrade unbound nucleotide with template. This process is repeated nucleotide added one by one and for each time one nucleotide added if it is complementary to template strand pyrophosphate will release otherwise this nucleotide will be washed by the action of apyrase enzyme. Enzymatic reaction for converting pyrophosphate to light started through converting pyrophosphate to ATP by the action of sulfurylase, then luciferase used with ATP to convert luciferin to light Figure 7. One of the most famous application for Pyrosequencing technique is to detect epigenetic through DNA methylation[11].

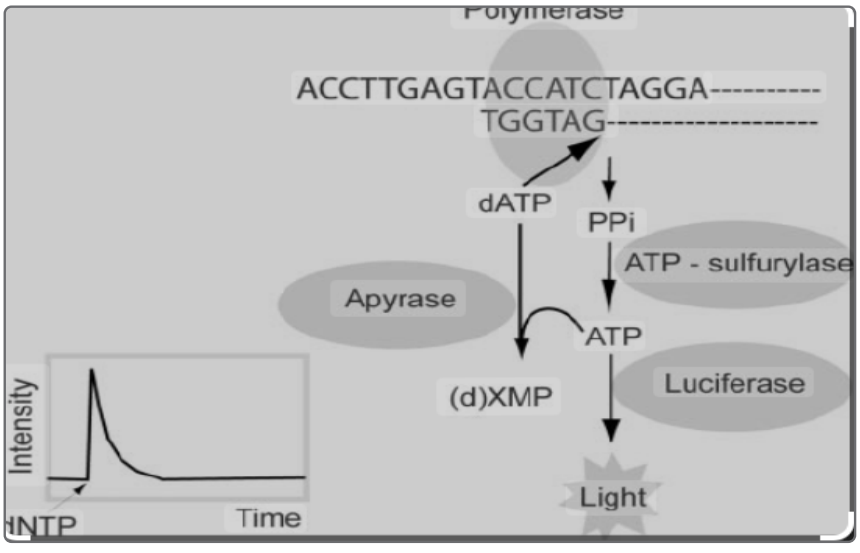

Figure 7: Enzymatic reaction for Pyrosequencing sequencing technology (adopted from Ronald Davis, 2015)

Principle of Pyrosequencing technique

Whole genome DNA is sheared into many fragments and ligated with adaptors for amplification resulted fragments into many copies, then the amplified fragment dissociated into single stands DNA and ligated into beads thereby each fragment binds with one bead. Emulsion solution is created (water with oil emulsion) containing one fragment bound with bead with all enzymes and substance required for this process, this emulsion called microreactor. Then each ssDNA of each fragment in each microreactor amplified into millions of copies and sequenced as mentioned previously Figure 8 [12].

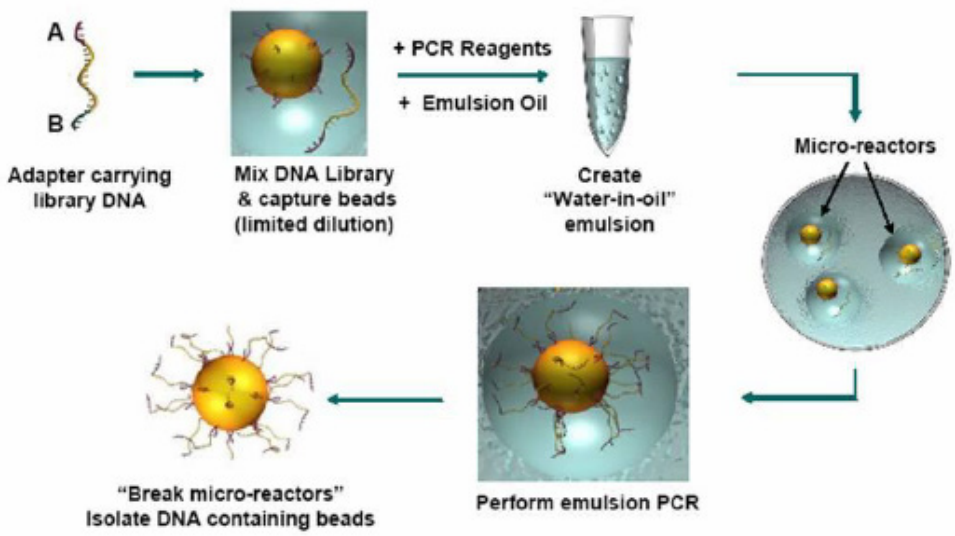

Figure 8: steps of creation emulsion PCR

Second generation sequencing: Ion Torrent

The main idea that build Ion Torrent sequencing technology is similar to the Pyrosequencing technology in sequencing by synthesis, however Ion Torrent relies on electrochemical method for detection the sequencing through using huge amount of sensors. Through the action of DNA polymerase in synthesis the complementary strand for the template and incorporate the complementary nucleotide to the template hydrogen ion will be released from 3' OH group through formation phosphodiester bond between incoming nucle-

otide and nucleotide already existed in the strand and this lead to change in pH. The changing of pH can be detected using sensors and transformed into electrical signal that finally indicated on incorporation the complementary nucleotide to the template. Absence of electrical signal means nucleotide is not complementary to the template therefore DNA polymerase unable to add nucleotide to the template. As in Pyrosequencing technology, the nucleotide added one by one at time for the four different nucleotides A, G, C, T, starting with A nucleotide and it will be added by DNA polymerase enzyme if it is complementary to the template, hydrogen ion will be released and this lead to changing the pH that can be detected using sensors and transformed into electrical signal, otherwise A nucleotide will be washed and second nucleotide will be added which is G nucleotide. The process is repeated till the sequence get detected. Before sequencing, the genomic DNA is fragmented into different fragments and adaptors used to link these fragments into beads by which each bead carries one DNA fragment that will be amplified into thousands of copies before sequencing. Each bead carries one DNA fragment with multiple copies settle in well of chamber with semiconductor chip. Rising in H+ ion in well of chamber create electrical current nearby gate of the transistor that can be transmitted into signal for each particular base added A, G, C and T[13].

There is some limitation in this sequencing technology that represented by:

1 - It is difficult to detect the homopolymeric repeat as multiple H+ ions will be released in one cycle so it is difficult to detect the homopolymeric repeat as it is 8 or 9 repeats of e.g. AAAAAAAA

2 – The length of read is quite small which is up to 400 bps and it is less-throughput sequencing compared with other sequencing technology[14] .

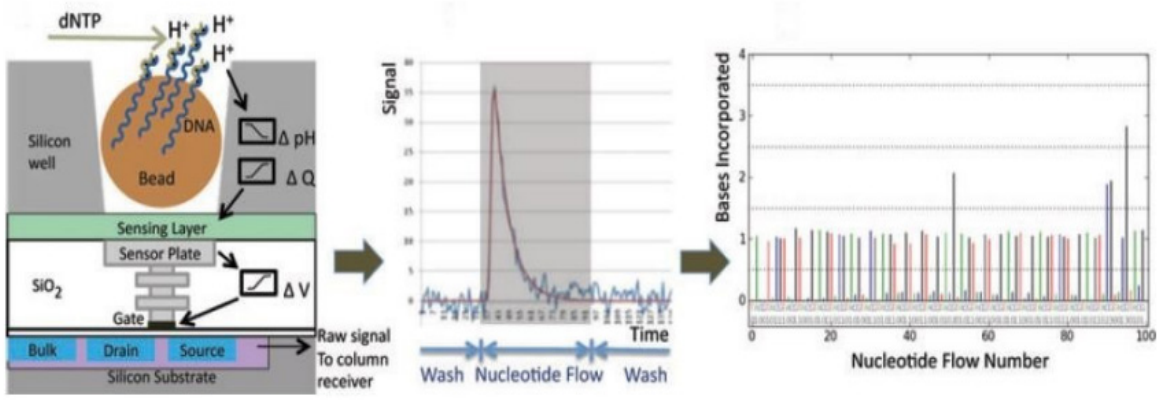

Figure 9: Steps of Ion Torrent sequencing

Second generation sequencing: Illumina sequencing

This sequencing technology started at randomly fragmented genomic DNA and ligated into two different adaptors. The adaptors used to link the fragment into flow cell channel and bridge amplification starts immediately after unlabeled nucleotides and enzyme were added into flow cell channel. Double strand bridge resulted from bridge amplification that will be denatured to leave single strands template forming cluster with millions of copies for each fragment ligated to flow cell channel. The sequencing start at addition four label reversible terminator nucleotides, DNA polymerase and primer into cluster of templates attached on flow cell channel. The image of emitted fluorescence was recorded after laser excitation for first nucleotide added which is complementary to first base in template. Then, terminator and fluorescence was removed from first attached base nucleotide and the second label reversible terminator nucleotide added which is complementary to second base in template and again image of emitted fluorescence was recorded after laser excitation. This process is repeated until the sequence of all attached templates on flow cell channel will be recorded simultaneously. In this technique, the genome is sheared into fragments and only 250 bps can be read from fragment due to reading of larger fragment leading to contaminate

the next image. The two adaptors were used to achieve the sequence from both directions forward and reverse direction. bridge amplification is carried out to increase depth of coverage and accuracy of sequencing, finally all overlaps read will be aligned to construct the full genome sequence

Illumina sequencing is better than Sanger sequencing by Speed, Cost, Sample size, Accuracy as follows;

1 – The amount of template needed for sequencing in Sanger sequencing is very large while in Illumina sequencing we need single copy of template.

2 - Illumina sequencing is faster than Sanger sequencing by which chemical reaction combined with signal detection in Illumina sequencing while they are two separated process in Sanger sequencing

3 - 300Gb of DNA can be sequenced in single run in Illumina sequencing while in Sanger sequencing maximum ~1kb sequenced in single run

4 - It is possible to do more repeat in Illumina sequencing which means greater coverage and high accuracy sequencing[15, 16] .

Third-Generation Sequencing: PacBio sequencing

PacBio defined as single molecule real time (SMRT) which is first sequencing technology that used nanosensor. The idea of sequencing builds on sequencing by synthesis for DNA fragment with 50 kbps. The previous sequencing technologies that rely on sequencing by synthesis used DNA polymerase that move through template to build complementary strand and incorporate each base and emit light that can be capture by camera and detect the type of each incorporated base. This sequencing technology relies on capacity of binding DNA polymerase with template therefore DNA polymerase cannot replicate large segment of DNA because it loses their binding capacity. Therefore, new idea comes with PacBio sequencing technology through fixing DNA polymerase in well of silicon chip through immobilization process while the template moves through DNA polymerase to synthesis the complementary strand. Synthesis the complementary strand is carried out in thousands of wells of silicon chip in area called zeromode wave-

guides (ZMWs). The laser and detector located at the lowest part of ZMW sensor and it is responsible for recoding the signal that comes from incorporation complementary nucleotide to the template. The detection of emitted light is carried out in thousands of well in ZMW nanosensor simultaneously. In this technology the same stand is sequenced multiple time because template is connected with circular double strand adaptor which is called SMRT-bell allowing traversing the circular DNA in opposite direction Figure 11. Comparison between TGS of PacBio and naopore sequencing technology is illustrated in table 2 [3, 17]Proteomics & Bioinformatics. 2015;13(5.

Third-Generation Sequencing: Nanopore sequencing

The idea of nanopore sequencing is build relies on stimulation electrical current due to disorder in channels that formed from proteins with nanopore size due to the passage of single strand DNA through these channels. These channels formed from nanosensors proteins and entering single stand DNA through pore of these channels achieved by helping of motor protein that directed the DNA through pore of these channels. Alteration of electrical current can be distinguishing for each base, however, the electrical current that detected by nanopore proteins accomplished five nucleotides therefore the nanopore systems detects 1000 different signals for each different combination of five bases that is called K-mer. Converting the electrical current into bases is knowing as bascalling. This technology is better than NGS by which the length of read is about 4 Mb and time consuming to prepare library is very short about 10 minutes. Each flow cells contains around 512 channels which can used to sequence multiple fragments of DNA simultaneously with accuracy reached to >98.3% in R9 chemistry while the accuracy is reached to >99% for R10 chemistry. The sequencing can be carried out for both forward and reverse strands of DNA through using linked adaptor [18].

Conclusion

The use of sequencing technology relies on authors that is some authors prefer to use NGS over TGS due high

depth of coverage which means high repeats that indicate to high accurate sequencing, however, other authors prefer to use TGS over NGS because the length of read is very large compared with length read produced by NGS therefore its less error in assembly and produce high accuracy sequence. In addition, PCR used to amplify the template to thousands of copies and this may lead to produce error thorough amplification, however in TGS the sequencing achived without template amplification. Therefore, we come to explain different sequencing technologies and its advantages and disadvantages.

References

[1]"Heather JM, Chain B. The Sequence of sequencers: the History of Sequencing DNA. Genomics. 2016;107(1):1-8. doi: https://doi.org/10.1016/j.ygeno.2015.11.003."

[2]"Gupta N, Verma VK. Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality. Microorganisms for Sustainability. 2019;17:313-341. doi: https://doi.org/10.1007/978-981-13-8844-6_15."

[3]"Ardui S, Ameur A, Vermeesch JR, Hestand MS. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucleic Acids Research. 2018;46(5):2159-2168. doi: https://doi.org/10.1093/nar/gky066."

[4]"Kang Y, Kang CS, Kim and C. History of Nucleotide Sequencing Technologies: Advances in Exploring Nucleotide Sequences from Mendel to the 21st Century. ww-whst-jorg.:549-558. doi: https://doi.org/10.7235/HORT.20190055."

[5]J. M. Heather, and B. Chain, "The sequence of sequencers: The his-

tory of sequencing DNA," *Genomics,* vol. 107, no. 1, pp. 1-8, Jan, 2016.doi: 10.1016/j.ygeno.2015.11.003.

[6]"Wong KC, Zhang J, Yan S, et al. DNA Sequencing Technologies. ACM Computing Surveys. 2019;52(5):1-30. doi: https://doi.org/10.1145/3340286."

[7]"Chen YJ, Huang X. DNA sequencing by denaturation: Principle and thermodynamic simulations. Analytical Biochemistry. 2009;384(1):170-179. doi:https://doi.org/10.1016/j.ab.2008.09.048."

[8]"Chen CY. DNA polymerases drive DNA sequencing-by-synthesis technologies: both past and present. Frontiers in Microbiology. 2014;5. doi: https://doi.org/10.3389/fmicb.2014.00305."

[9]"Guo J, Yu L, Turro NJ, Ju J. An Integrated System for DNA Sequencing by Synthesis Using Novel Nucleotide Analogues. Accounts of chemical research. 2010;43(4):551-563. doi: https://doi.org/10.1021/ar900255c."

[10] "Dovichi NJ. DNA sequencing by capillary electrophoresis. Electrophoresis. 1997;18(12-13):2393-2399. doi: https://doi.org/10.1002/elps.1150181229."

[11] "Moorthie S, Mattocks CJ, Wright CF. Review of massively parallel DNA sequencing technologies. The HUGO Journal. 2011;5(1-4):1-12. doi: https://doi.org/10.1007/s11568-011-9156-3."

[12] "Siqueira JF, Fouad AF, Rôças IN. Pyrosequencing as a tool for better understanding of human microbiomes. Journal of Oral Microbiology. 2012;4(1):10743. doi: https://doi.org/10.3402/jom.v4i0.10743."

[13] "Garrido-Cardenas J, Garcia-Maroto F, Alvarez-Bermejo J, Manzano-Agugliaro F. DNA Sequencing Sensors: An Overview. Sensors. 2017;17(3):588. doi: https://doi.org/10.3390/s17030588."

[14] "Merriman B, R&D Team IT, Rothberg JM. Progress in Ion Torrent semiconductor chip based sequencing. ELECTROPHORESIS. 2012;33(23):3397-3417. doi: https://doi.org/10.1002/elps.201200424."

[15] "Anderson MW, Schrijver I. Next Generation DNA Sequencing and the Future of Genomic Medicine. Genes. 2010;1(1):38-69. doi: https://doi.org/10.3390/genes1010038."

[16] "Hu T, Chitnis N, Monos

D, Dinh A. Next-generation sequencing technologies: An overview. Human Immunology. 2021;82(11). doi: https://doi.org/10.1016/j.humimm.2021.02.012."

[17] "Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genomics, Proteomics & Bioinformatics. 2015;13(5):278-289. doi: https://doi.org/10.1016/j.gpb.2015.08.002."

[18] "Bayley H. Nanopore Sequencing: From Imagination to Reality. Clinical Chemistry. 2015;61(1):25-31. doi: https://doi.org/10.1373/clinchem.2014.223016."