# CHAPTER 1

# INTRODUCTION

Genomics is an interdisciplinary field of biology focusing on the structure, function, evolution, mapping, and editing of genomes. A genome is an organism's complete set of DNA, including all of its genes as well as its hierarchical, three-dimensional structural configuration (Satzinger, 2008). In contrast to genetics, which refers to the study of individual genes and their roles in inheritance, genomics aims at the collective characterization and quantification of all of an organism's genes, their interrelations and influence on the organism (WHO, 2004). By providing researchers with detailed information about the genetic makeup of organisms, it has opened up new avenues of research and has the potential to lead to many important advances in fields such as medicine, agriculture, and evolutionary biology.

The history of genome sequencing has been started 50 years before. Eventually, many attempts were made after the discovery of the structure of DNA by Watson, Crick and Franklin in 1953. In 1965, Robert Holley first time sequenced the tRNA (Holley, 1968). But Walter Fiers successfully sequences the DNA of a complete gene. This gene can encode the protein coat of the bacteriophage MS2. He utilised the RNAses to digest the virus RNA and isolate oligonucleotides, and then separate them via electrophoresis/chromatography (Declercq *et al*., 2019; Jou *et al*., 1972). Then, Frederick Sanger and his team brought a major breakthrough by developing first method of DNA sequencing in the first half of 80 decades in the last century. In this method, they used a chain termination approach to determine the sequence of a DNA strand that is named after the discoverer called Sanger sequencing method. This method was used to sequenced the bacteriophage ΦX174 (Sanger, *et al*., 1977) followed by the sequencing of other microbes, plants, and animal's genome including human. The significant success was coming in 2003 after the declaration of first draft of human genome sequence (HGS). Since then, several countries have made strides in genome sequencing, including the United States, China, and the United Kingdom. On the contrary, with the advent of new technologies, genome sequencing has become faster, cheaper, and more accurate, enabling scientists to study the genetic basis of diseases, evolution, and biodiversity.

Bangladesh, however, has been slow to adopt this technology, largely due to resource constraints and limited infrastructure. Despite these challenges, the country has made significant progress in recent years, with the collaborations of many international research

institutions. Bangladesh makes a millstone in genomic studies by unveiling the first draft jute sequence. In 2010, the Prime minister Sheikh Hasina has disclosed in the parliament that Bangladeshi researchers have successfully done genome sequencing of jute which will contribute to improving jute fibre (bdnews24.com, 2010). She also added a group of Bangladeshi researchers of Dhaka University's Biochemistry and Biotechnology departments, led by an expatriate, Dr Maqsudul Alam, made the accomplishment.

In recent years, there has been an increase in attention towards genome studies in Bangladesh, a country with a rich diversity of flora and fauna. Bangladeshi researchers have already been successful to decode the whole-genome sequencing of the national fish Ilish, national fruit jackfruit, black Bengal goat, water buffalo, rohu carp and several microorganisms such as *Macrophomina phaseolina* fungus, SARS-Cov-2, chikungunya virus etc. In addition, Bangladeshi scientists also started the human genome sequencing project in 2019 (Khan *et al*., 2021). But there are no databases or platforms, nor even review papers from where Bangladeshi researchers or students will acknowledge how many organisms have been genomes sequenced. For instance, currently, there are 1,922,378 projects found on whole genome sequences (WGS) in the national center for biotechnology information (NCBI) (Source-NCBI, 2023). After unveiling the sequencing of how many technologies were derived and the economic impact study of these technologies in our GDP also absences. So, the objectives of the study are-

- To evaluate the genome sequencing methods in different generations,
- To assess the genomic studies of organisms in Bangladesh and
- Future prospects of Bangladesh in the genome industry.

# CHAPTER 2

## MATERIALS AND METHODS

This seminar paper is a review article. So, all information collected in this paper is from secondary sources. During the preparation of the review paper, I followed various tools including search engines such as Google, and Bing, relevant journals sites GoogleScholar, Research Gate, PubMed and Sci-hub. After gathering all the information available, I agreed, organized, and explained in my way and finally prepared for this meeting to meet the objectives. In addition, my major professor and course instructors also provided many important suggestions for preparing this seminar paper.

## CHAPTER 3

## REVIEW OF FINDINGS

### Genome and its feature

The term genome was created in 1920 by Hans Winkler professor of botany at the University of Hamburg, Germany (Winkler, 1920). In the fields of molecular biology and genetics, a genome is all the genetic information of an organism (Roth, 2019). A genome feature is a characteristic or element of a genome, which is the complete set of genetic material (DNA) of an organism. Genome features can include genes, regulatory elements, non-coding regions, repetitive sequences, transposable elements, and structural variations such as deletions, insertions, and inversions.

Genes are the most well-known genome feature, and they encode the information for the synthesis of proteins or RNA molecules. Regulatory elements control gene expression by turning them on or off, and they can be located near or far from the gene itself. The proportion of the genome occupied by coding sequences varies widely. A larger genome does not necessarily contain more genes, and the proportion of non-repetitive DNA decreases along with increasing genome size in complex eukaryotes.

Non-coding regions are sections of DNA that do not encode protein-coding genes, but they may have other important functions such as controlling chromosome structure or serving as sites for DNA replication initiation. For instance, Non-coding sequences make up 98% of the human genome.

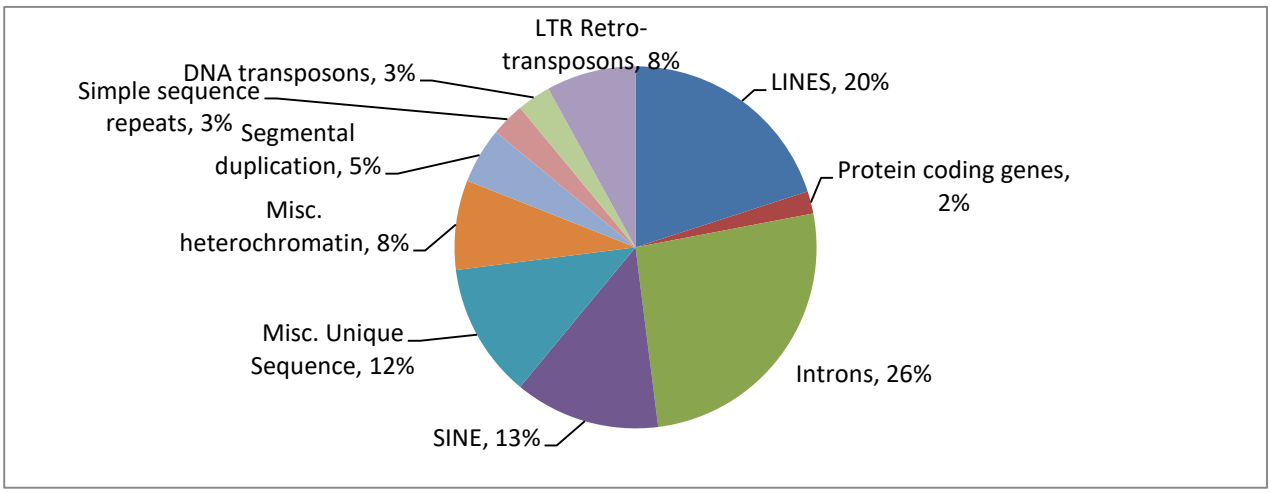


**Figure 1.** Components of human genome (Sources- Goldmann, 2019).

Repetitive sequences are stretches of DNA that are repeated throughout the genome, and they can vary in size and complexity. Transposable elements are DNA sequences that can move from one location to another within the genome, sometimes causing mutations or rearrangements. Structural variations are changes in the size or arrangement of DNA segments, and they can affect gene expression or protein function. Overall, genome features play important roles in shaping the characteristics and traits of organisms, and they are studied extensively in fields such as genomics, genetics, and molecular biology.

**Methods of genome sequencing**

Many methods develop for sequence the genome over the last 50 years. Due to advancement of the technology and automation of modern sequencers, genome sequencing methods can be divided into three generations. The first complete genome project named The Human Genome Project (HGP) was based on the Sanger sequencing method and it took around 13 years and an astronomical three billion USD to complete it. However, DNA sequencing came a long way since. Using the next generation sequencing (NGS) technologies of today, the entire HGP could be done in less than two weeks and at a cost of just around € 1,000 (Behjati & Tarpey, 2013). So, it's a huge advancement of genome sequencing technology.

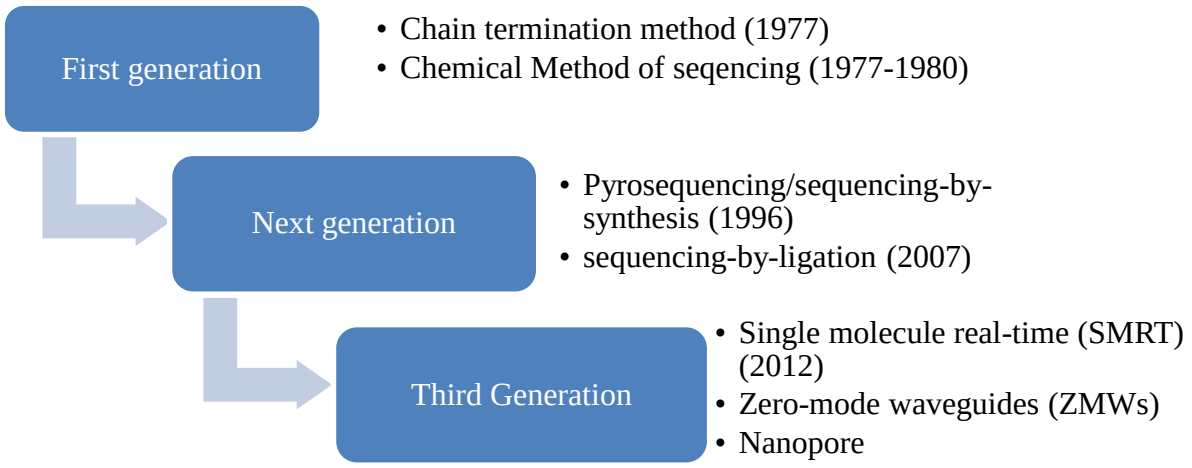


**Figure 2.** Genome Sequencing Techniques in different generation.

**The first generation sequencing techniques**

Frederick Sanger is the pioneer of genome sequencing technique. He developed a DNA sequencing method which he called the chain termination method or dideoxy technique (Sanger & Nicklen, 1977). This method used radiolabeled partially digested fragments and dominated

the sequencing world for three decades. Sanger's ground breaking work in genomics earned him two Nobel Prizes, one in 1980 and another in 1958, and he is still considered a giant in the field. In 1977, Sanger used his method to sequence the complete genome of the bacteriophage PhiX174, which became the most popular DNA positive control in labs worldwide.

In the same year, Maxam and Gilbert introduced a chemical modification-based technique for DNA sequencing that did not rely on DNA polymerase (Gužvić, 2013; Heather and Chain, 2016). Despite being time-consuming and tiring, the research community recognized the potential of Sanger sequencing and worked on automating the process. In 1984, Fritz Pohl established the GATC1500 sequencing technology platform, which did not rely on radioactive labeling. Leroy Hood and Michael Hunkapiller of Applied Biosystems, Inc. (ABI) automated the Sanger sequencing process in 1987 with two significant improvements. They used fluorescent dyes to label DNA fragments instead of radioactive molecules and made data acquisition and analysis possible on the computer. The resulting instrument, the ABI 370, marked a major step forward for DNA sequencing and paved the way for the next generation of sequencing technologies (Gužvić, 2013; Hood *et al*., 1987).

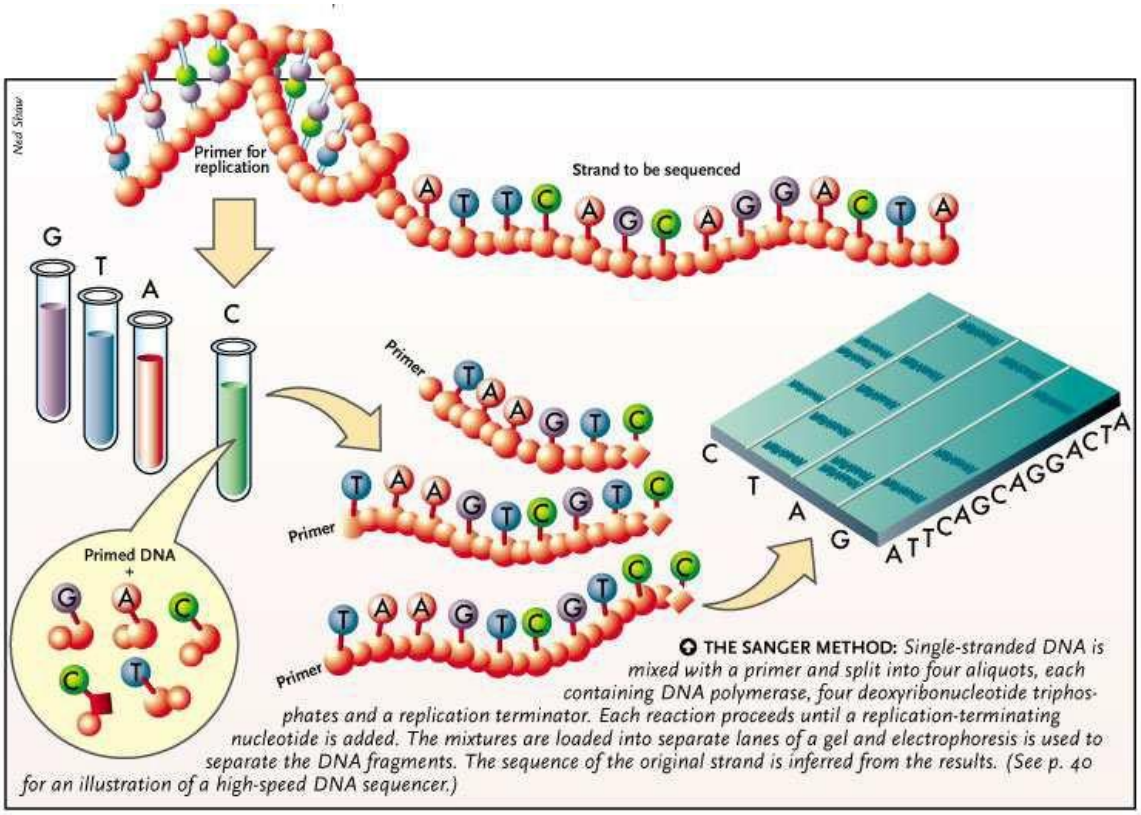


**Figure3.** A representation of the Sanger chain termination sequencing method extracted. (Source: Winnick, 2004).

**Next generation sequencing (NGS) techniques**

In 1996, a new DNA sequencing technique named pyrosequencing was introduced by Mostafa Ronaghi, Mathias Uhlen, and PålNyŕen, marking the emergence of the second generation of DNA sequencing. This high-throughput, automated technology measures luminescence generated during sequencing, based on pyrophosphate synthesis (sequencing-by-synthesis technology). Other biotechnology companies, with their own technologies, emerged later. In 1998, Shankar Balasubramanian and David Klenerman founded Solexa and developed a new sequencing-by-synthesis method using fluorescent dyes.

The key feature of next-generation sequencing (NGS) is the parallelization of a large number of reactions through automation and miniaturization. The 454 system, the first NGS platform to come to market, was implemented in an automated system in 2005 by Jonathan Rothberg and colleagues. Illumina, which acquired Solexa in 2007, became the leader of NGS platform market, offering the most widely used NGS technology in the world. Other notable platforms based on different technologies include SOLiD system's "sequencing-by-ligation" in 2007 and Ion Torrent by Life Technologies in 2011, which uses "sequencing-by-synthesis" technology detecting hydrogen ions when new DNA is synthesized.

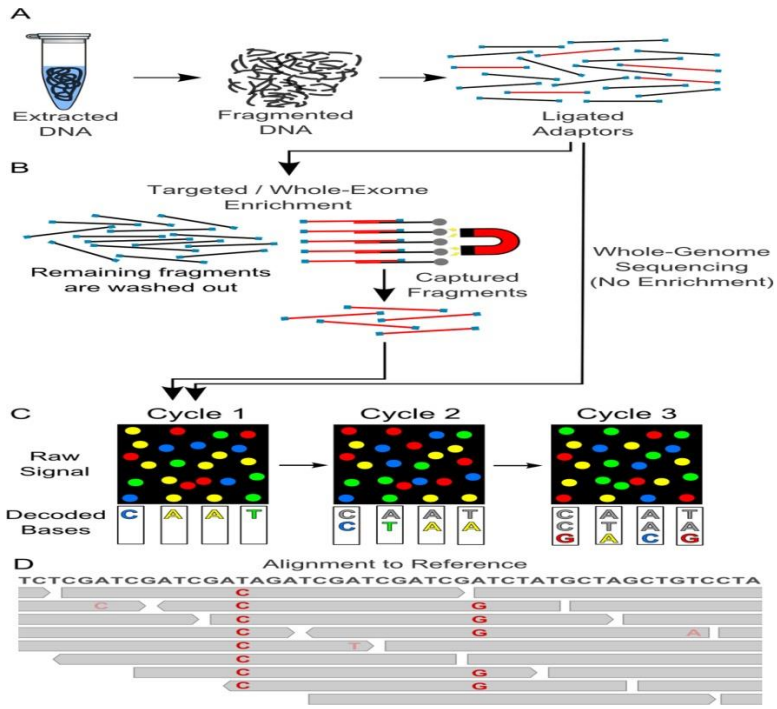


**Figure 4.** Steps in next-generation sequencing. (A) Extracted DNA is randomly broken into <1000 bp fragments. Known adaptor sequences are ligated to fragments. (B) All fragments are sequenced in WGS, whereas in whole-exome and targeted sequencing only a subset of the

original fragment pool is sequenced. (C) An example of a NGS platform (Illumina). It relies on spatial separation of fragments on a slide and clonal amplification by PCR to generate fragment clusters. Four fluorescently labelled nucleotides are added to the slide and compete to be incorporated to the growing chains. In each cycle, the clusters are excited by laser and the emitted fluorescence (colored circles) is recorded by an image-capturing device. As the position of each individual cluster remains fixed, the sequencer creates a 'time lapse' with the recorded images from all cycles, with each cluster generating a read. (D) Individual reads (gray rectangles) aligned to the reference genome. The coverage for each genomic position is the number of reads that overlap at that position. The first base in the figure (T) has 5× coverage; the last base (A) is covered nine times. Bases that match the reference sequence have been omitted. Examples of homozygous and heterozygous single-nucleotide variants are shown (left and right, respectively). Examples of sequencing and/or mapping errors are shown as faded bases (Sources- Schnekenberg *et al.*, 2014).

**The third generation sequencing techniques**

It is true that the distinction between second and third generation sequencing technologies can be somewhat blurry, and there is ongoing debate about how to define these terms. Some argue that SMRT sequencing technologies like PacBio's should be considered third generation due to their ability to directly observe the incorporation of nucleotides in real time (Heather and Chain, 2016). Other features that have been associated with third generation sequencing technologies include long read lengths and single-molecule sequencing, which allow for the detection of more complex genomic features like structural variations and epigenetic modifications.

In addition to PacBio's ZMW-based sequencing, there are other third generation sequencing technologies that use different approaches. Oxford Nanopore's nanopore-based sequencing is another notable example. This technology uses biological nanopores to detect changes in electrical conductivity as DNA strands pass through, allowing for the direct observation of nucleotide sequence (Hayden, 2012; Lu *et al* ., 2016). The company has also developed portable handheld devices like the MinION, which can be used in remote or field settings.

It's worth noting that Illumina's NovaSeq platforms, while not typically considered third generation sequencing technologies, do represent a significant advance in sequencing power and throughput compared to earlier Illumina systems. These platforms can generate outputs of up to 3000 Gb per run, allowing for more efficient sequencing of large genomes or multiple

samples in parallel (Illumina, 2017). Nabsys' HD-Mapping technology, which uses sequence-specific tags and nano-detectors to create genome maps, is another interesting approach that could have applications in both research and clinical settings.

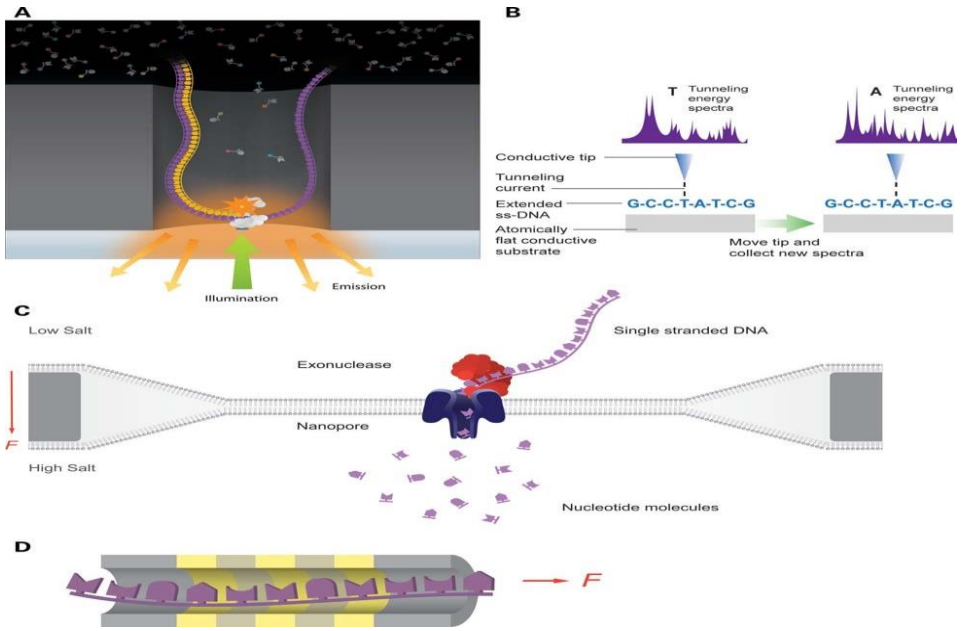


Figure 5. How third-generation DNA-sequencing technologies work. Third-generation DNA-sequencing technologies are distinguished by direct inspection of single molecules with methods that do not require wash steps during DNA synthesis. (A) Pacific Biosciences technology for direct observation of DNA synthesis on single DNA molecules in real time. A DNA polymerase is confined in a zero-mode waveguide and base additions measured with florescence detection of gamma-labeled phosphonucleotides. (B) Several companies seek to sequence DNA by direct inspection using electron microscopy similar to the Reveo tech-nologypicturedhere, in which an ssDNA molecule is first stretched and then examined by STM. (C) Oxford Nanopore technology for measuring translocation of nucleotides cleaved from a DNA molecule across a pore, driven by the force of differential ion concentrations across the membrane. (D) IBM's DNA transistor technology reads individual bases of ssDNA molecules as they pass through an arrow aperture based on the unique electronic signature of each individual nucleotide. Goldbands represent metal and gray bands dielectriclayers of the transistor (Source- Schadt *et al.*, 2011).


**Figure 6.** Genome sequence platforms of different generation.

**Differences of different generation genome sequencing**

| Characteristics | First generation | Second generation | Third generation |
|---|---|---|---|
| Fundamental technology | Size-separation of specifically end-labeled DNA fragments, produced by SBS or degradation | Wash-and-scan SBS | SBS, by degradation, or direct physical inspection of the DNA molecule |
| Resolution | Averaged across many copies of the DNA molecule being sequenced | Averaged across many copies of the DNA molecule being sequenced | Single-molecule resolution |
| Current raw read accuracy | High | High | Moderate |
| Current throughput | Low | High | Moderate |
| Current cost | High cost per base | Low cost per base | Low-to-moderate cost per base Low cost per run |
| RNA-sequencing method | cDNA sequencing | cDNA sequencing | Direct RNA sequencing and cDNA |
| Time from start of sequencing reaction to result | Hours | Days | Hours |
| Sample preparation | Moderately complex, PCR amplification | Simple preparation | Moderately complex, PCR amplification |
| Data analysis | Routine | Complex because of large data volumes and because short reads complicate assembly and alignment algorithms | Data analysis |
| Primary results | Base calls with quality values | Base calls with quality values | Base calls with quality values |

**Table 1.** Differences of different generation genome sequencing (Source- Schadt *et al.*, 2010)

## Genomic Studies in Bangladesh

Bangladesh is a South Asian country with a population of over 160 million people. The country is home to a diverse range of flora and fauna, including many endangered species. In recent years, there has been a growing interest in sequencing the genomes of various organisms found in Bangladesh. Bangladeshi scientists have successfully decoded the genome sequence of jute (2010) and Ilish fish (2018). Even the scientists of Bangladesh Council of Scientific and Industrial Research (BCSIR) started the whole genome sequence of humans in 2019 under the funded of People's Republic of Bangladesh. So, several genome sequencing projects have been successfully completed and researchers trying to unveil the coding of genome including plants, animals, and microbes.

## Plant Genomics in Bangladesh

Agriculture is a critical sector of the Bangladeshi economy, and the country is home to a rich diversity of crop species, including rice, wheat, maize, and vegetables. Genomic studies of crop plants in Bangladesh have focused on understanding the genetic basis of agronomic traits, disease resistance, and abiotic stress tolerance. These studies have provided valuable insights into the genetic diversity and adaptation of crop plants in Bangladesh and have the potential to contribute to the development of improved varieties and sustainable agricultural practices. However, according to the latest available information, there has been some progress in plant genome sequencing in Bangladesh. Here are a few examples of plant genome sequencing projects in Bangladesh:

| Species Name | Whole/Partial Genome Sequencing | Platforms | Size of the genome | Ax. No of the genes | GC% | NCBI Gene Bank accession no | References |
|---|---|---|---|---|---|---|---|
| Jute | | | | | | | |
| *C. olitorius* | WGS | WGS approach on the 454 platform | ~448 Mb | 37,031 | 34.10 | AWUE00000000 | Islam *et al*., 2017 |
| *C. capsularis* | WGS | WGS approach on the 454 platform | ~404 Mb | 30,096 | 34.84 | AWWV00000000 | Islam *et al*., 2017 |
| Jackfruit | WGS | Illumina NextSeq 550 | 1.04 Gb | 41,088 | 34.10 | CNGB CNP0000486 | Islam *et al*., 2017 |

**Table 2.** Genome sequencing of plants in Bangladesh

## Jute genome sequencing project

Unravelling the draft genome sequencing of jute is a historic achievement for Bangladesh. The project was then named Swapna Jaatra (meaning a journey to achieve a cherished goal) Jute is known as the "Golden fiber" of Bangladesh for hundreds of years. The annual global production of jute generates a farm value of ~US$2.3 billion (FAO, 2014). So, research on jute fiber development, physiology, genomics, and evolution is much needed to increase our knowledge and ability for improving jute yield and fiber yield. So, an endeavour to decode the jute genome was made by the Government of Bangladesh in late 2009. Two groups were made one is Biology teams led by Professor Haseena Khan of the Molecular Biology Laboratory in the Department of Biochemistry and Molecular Biology at the University of Dhaka and another one is computing team led by Mr. Mahboob Zaman of DataSoft Systems Bangladesh Limited, respectively. Professor Maqsudul Alam who worked in University of Hawaii at Manoa, led and coordinated both teams.

This team did the whole-genome shotgun (WGS) sequencing of two commercially cultivated species of jute, *Corchorus olitorius* and *Corchorus capsularis* with the Roche/454 platform (Islam *et al*., 2017) and assembled the genomes using CABOG. They estimated the genome sizes for *C. olitorius* and *C. capsularis* to be ~448 and ~404 Mb, respectively (Table 2 & 3). The assemblies cover 91.6% and 82.2% of the estimated genome sizes for *C. olitorius* and *C. capsularis*, respectively (Islam *et al*., 2017). From the table 3, it was observed that 37,031 *C. olitorius* and 30,096 *C. capsularis* protein-coding genes were predicted by using a combination of de novo, homology and transcriptome-based approaches. More than 50% of the *C. olitorius* and *C. capsularis* genomes were composed of repetitive elements, which is similar to cotton (~57%) and double that of cacao (~24%). The predicted GC content (%) was 34.10 and 34.84 *C. olitorius* and *C. capsularis*, respectively (Table 2 & 3).

The NCBI GenBank accession no. of this project is AWUE00000000 for *C. olitorius* and AWWV00000000 for *C. capsularis*. The genomic and transcriptomic raw data was also deposited in the NCBI Sequence Read Archive (SRA) under SRP049494 and SRP053213 for *C. olitorius* and *C. capsularis*, respectively. The genome sequences provide a valuable resource to advance our understanding of fibre biogenesis in jute, thus serving as the foundation of genetic improvement for productivity and fibre quality.

| Genome features | *C. olitorius* | *C. capsularis* |
|---|---|---|
| Estimated genome size (Mb)* | 447.95 | 404.09 |
| Assembled genome size (Mb) | 445.05 | 338.13 |
| Number of scaffolds (≥500 bp) | 22,944 | 6,125 |
| Number of N50 scaffolds | 31 | 14 |
| N50 scaffold length (Mb) | 3.30 | 4.13 |
| Longest scaffold (Mb) | 45.45 | 28.54 |
| GC content (%) | 34.10 | 34.84 |
| Transposable elements (%) | 53.72 | 56.17 |
| Predicted protein-coding genes | 37,031 | 30,096 |
| Gene density† | 0.90 | 0.91 |
| miRNA | 1,010 | 666 |
| tRNA | 488 | 203 |
| rRNA | 80 | 110 |

(† Gene density expressed in number of genes per 10 kb and based on total contig length (410.19 Mb and 331.96 Mb for *C. olitorius* and *C. capsularis*, respectively)

**Table 3.** Assembly and annotation features of the *C. olitorius* and *C. capsularis* genomes (Source- Islam *et al*., 2017).

**Jackfruit genome sequencing**

Jackfruit (popularly known as 'Kanthal') genome sequence is another milestone in plant genomics studies of Bangladesh because it is our national fruit. Bangladesh is one of the largest producers of jackfruit and accounts for about 21% of total fruit production of the country, second only to Mango as the principal fruit crop. During 2019-20, Bangladesh produced 1.1 million tons of jackfruit covering 16,592 hectares area (Statistics, 2020). Its demand is increasing gradually due to its low price, high nutritious value, diversified uses and potential for commercial cultivation (Sidhu, 2012). So, an initiative was taken by Bangladeshi researchers to generate a draft whole-genome sequence (WGS) of BARI Kanthal-3 to obtain molecular insights including genes associated with year-round fruiting trait of this important unique variety. This project was funded by the 'BSMRAU Physical Facility and Research Capacity Strengthening Project' under the Ministry of Education of the People's Republic of Bangladesh.

The whole genome sequencing was done by using the Illumina NextSeq 550 desktop sequencer (Islam *et al*., 2022). The genome size of BARI Kanthal-3 was estimated to be 1.04 Gb with a heterozygosity rate of 1.62% based on K-mer analysis of the short-read data (Table 2). The GC

content of BARI Kanthal-3 was 34.10% which is comparable to the GC content of a seasonal A. heterophyllus from Indian and Chinese origins that were recorded at 32.9% and 34.9%, respectively (Islam *et al* ,. 2022). They annotated 41,088 protein-coding genes in BARI Kanthal-3 assembly using the Braker2 gene prediction pipeline (Islam *et al.*, 2022). The WGS projects had been deposited at NCBI under GenBank accession numbers of CNGB CNP0000486, and, NCBI Bioproject PRJNA565858, respectively.

One of the limitations of this study is lack of information on the role of intergenic SNPs in BARI Kanthal-3 and the impact on gene function (i.e., changes in protein-coding genes, differential gene expression, and the specific functions of protein-coding genes). To further understand the underlying molecular mechanisms of unique traits of this new variety, a high quality long read genome assembly of BARI-Kanthal-3 and comparative transcriptome analysis with seasonal jackfruit are needed.

**Animal Genomics in Bangladesh**

Bangladesh has a diverse range of animal species, including livestock, poultry, and wild animals, many of which are important for food security and economic development. Genomic studies of animal species in Bangladesh have focused on understanding the genetic basis of traits such as meat quality, milk yield, and disease resistance. However, according to the publicly available information, there have been some initiatives in Bangladesh to sequence animal genomes. Here are a few examples:

| Species Name | Whole/Partial Genome Sequencing | Platforms | Size of the genome | Ax. No of the genes | GC% | NCBI Gene Bank accession no | References |
|---|---|---|---|---|---|---|---|
| Hilsa | WGS | Illumina and PacBio sequencing platforms | 816 MB | 31,254 | 43.61 | QYSC01000001 - QYSC01124209 | Das *et al.*, 2018 |
| Rohu carp | WGS | GridION sequencer & Illumina HiSeq X-Ten | 945.5 Mb | 31274 | | JACTAM000000000 | Arick *et al.*, 2023 |
| Black Bengal Goat (Local Male, Chattogram) | WGS | Illumina HiSeq 2500 sequencing platform | 3.04 Gb | 26,458 | 41.77 | SMSF01000001 – SMSF01003972 | Siddiki *et al.*, 2019 |
| Buffalo | Whole genome shotgun project | Illumina HiSeq2000 | 2,770,477,792/2.77 Gb | 24613 | | NPZD00000000 | Mintoo *et al.*, 2019 |

**Table 4.** Genome sequencing of animals in Bangladesh

**Hilsa fish sequencing project**

Hilsa (*Tenualosa ilisha*) fish also known as ilish in Bangladesh is very popular for its taste and the texture of its flesh. Ilish is our national fish and it belongs to the shad in Clupeidae family. It contributes to 11% of total fish production and 1% to the national GDP, 3% of the total export earnings and about 2.5 million people in Bangladesh are directly dependent on Hilsa in providing for their families (Md *et al*., 2016 & DoF, 2017). So, due to its national importance, it was sequenced by Illumina HiSeq 4000 and Pacific Bioscience Sequel (Das *et al*., 2018). It was submitted to NCBI from the molecular biology lab, Department of Biochemistry and Molecular Biology, University of Dhaka under the supervision of Haseena Khan at 2018-09-10. NCBI GeneBank Accession numbers is QYSC01000001– QYSC01124209.

Avizit Das and Oly Ahmed collected Fresh *Tenualosa ilisha* samples from the river Padma at Rajshahi on 2017/09/20 and instantly preserved on dry ice. White and red muscles of the fish were used for DNA extraction. The assembled genome size of *Tenualosa ilisha* was found 816 Mb (Mega base pair) and approximately 82% of the genome has been assembled (Das *et al.,* 2018). According to Das *et al*., 2018, the benchmarking universal single-copy orthologs (BUSCO) analysis revealing 95% completeness as well as significantly lower number of scaffolds and considerably better N50 indicates the genome to be of high quality. From the table 4, we observed 31,254 genes were predicted and GC content was determined to be 43.61%. This research group observed that the Hilsa genome was found to be comparable to the Atlantic herring (807 Mb genome and 28,335 genes) and to the genome of the common carp (1.8 Gb and 52,000 genes). The limitation of this project was noted that the number of the regions unassembled in the genome was 4605 and the total number of bases positioned in this gap observed 2,268,925 bp.

**Rohu carp sequencing project**

Rohu carp was sequenced and funded by the USAID Aquaculture for Income and Nutrition project (Grant ID EEM-G-00-04-00013-00). It was important because farm-raised rohu comprises 3.7% of the finfishes produced annually and represents the 11th most commonly farmed finfish (FAO, 2020). In the 2019-2020 fiscal year, the annual aquaculture production of rohu in Bangladesh was 386.3 thousand tonnes, the second-highest among all aquaculture species in the country (DoF, 2020).

Blood samples from five individual rohu (referred to as Rohu-1 through Rohu-5) were collected from a fish farm located in the District of Rangpur, Bangladesh for whole genome sequencing.

This samples were sequenced by GridION sequencer (Oxford Nanopore Technologies, Oxford, UK) and Rohu-1 genomic DNA was also sequenced on an Illumina HiSeq X-Ten (2x150 bp) (Arick *et al*., 2023). The final assembled genome size was 945.5 Mbp, representing 97.9% of the estimated genome size and 31,274 genes was annotated (Table 4). These findings help the researcher to utilize the new rohu genome in modernizing some aspects of rohu genetic improvement programs. The assembled genome sequence and annotations are available at GenBank under accessions no JACTAM000000000.

**Black Bengal Goat (*Capra hircus*) genome sequencing**

Black Bengal goat (BBG) belongs to the Bovidae family and is found throughout Bangladesh. It is estimated that more than 90% of the goat population in Bangladesh comprised the Black Bengal, the remainder being Jamunapari and their crosses (Husain, 1993). BBG have some outstanding features including higher prolificacy, fertility, resistance against common diseases, adaptability to the adverse environmental condition, early maturity, seasonality and superiority in the litter size. Moreover, it plays a vital role in the economy of Bangladesh by contributing 1.66% of the GDP (DLS 2017) (Siddiki *et al*., 2019).

A local healthy male Black Bengal goat was selected from Chattogram, Bangladesh. Then the genomic DNA sample collected for sequencing. Illumina HiSeq 2500 sequencing platform was used to complete BBG sequencing (Siddiki *et al*., 2019). The final assembled genome size of BBG was observed 3.04 Gb with 724.80 Mb (Megabase pair) gaps and GC content was 41.77% (Table 4). It showed 82.5% completeness which was assessed by BUSCO version 3.0.2. 26,458 and genes were annotated with Maker version 3.0 pipeline. The whole genome sequence data had been submitted in the NCBI Gene Bank under the Accession numbers SMSF01000001–SMSF01003972. The number of unassembled regions was 3943 and the total number of bases placed in this gap was 724,808,570 bp was the limitation of this genome sequencing. This study was supported by funding from Chattogram Veterinary and Animal Sciences University (CVASU).

**River water buffalo genome sequencing**

There are two types of domestic water buffalo, the River buffalo (*Bubalus bubalis*, 2n=50) and the Swamp Buffalo (*Bubalus carabanesis*, 2n=48) (Michelizzi *et al*., 2010). They provide more than 5% world's milk supply. Their milk has higher fat, lactose, protein, and higher minerals content than the milk of the cow. Buffalo milk is used to make butter, butter oil, high quality

cheeses, and various other higher quality dairy products (Buffalo, 2000). Their meat is very tender and palatable and their hides have economic importance as the raw materials of high quality leather products. In many parts of the world, especially in Southeast Asia countries, water buffalo provides 20%–30% of farm power, and their dung is used as fertilizer and fuel in many highly populated countries (Yindee *et al*., 2010).

The river water buffalo genomic DNA was extracted from eight years old plain land reverian type male water buffalo's blood from Bangladesh (Mintoo *et al*., 2019). Illumina HiSeq2000 paired end sequencing was performed with PE101 for short insert libraries, and PE50 for long insert libraries. A total of 2.77 Gbp of assembled sequences were obtained and 24,613 genes were predicted (Table 4). This whole genome shotgun project was deposited at DDBJ/ ENA/GenBank under the accession NPZD00000000.

**Microorganisms Genomics in Bangladesh**

Bangladesh has made significant progress in the field of genomics research in recent years, including the sequencing of the genomes of various microorganisms. The following are some of the notable genome sequencing efforts in Bangladesh.

**Virus**

The coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It belongs to the genus *Betacoronavirus*, family *Coronaviridae* and Bangladesh also experienced a steady rise in infections. The first confirmed case of SARS-CoV-2was observed on 8March 2020; as of 08April 2023, there were 2,038,052 confirmed COVID-19 cases (Source- dghs.gov.bd). To understand the genomic characteristics and evolutionary relationship, some public and private research organizations in Bangladesh continued their research on genome sequencing of SARS-CoV-2. Sanger and next generation sequencing techniques were used to identify the genome of SARS-CoV-2. From the table 5 the genome size of SARSCoV-2 of different individuals ranges between 29724-29892 bp. The GC content (%) ranges was 38.0 to 42.4. These data were deposited in global initiative on sharing all influenza data (GISAID) and NCBI.

Furthermore, the complete genome sequences of a buffalo coronavirus (BufCoV HKU26) detected from the faecal samples of two domestic water buffaloes (*Bubalus bubalis*) in Bangladesh (Lau *et al.*, 2016). The genomes of BufCoV HKU26 strains B1-24F and B1-28F were 31 021 and 30 975 in lengths, with GC content of 40% (Table 5). They possessed 98–

99% nucleotide identities to the genomes of BCoVs, supporting the classification of BufCoV HKU26 as a member of the species Betacoronavirus 1. The genome sequences of BufCoV HKU26 had been lodged at GenBank under accession numbers KU558922 and KU558923. Bangladesh also witnessed the outbreak of chikungunya virus (CHIKV) (genus *Alphavirus*, family *Togaviridae*). In July 2017, CHIKV samples were collected from a febrile male who visited from Bangladesh to Brisbane, Australia. Then the complete genome sequences were done by NextSeq 500 machine (Pyke *et al.,* 2020). From the table 5, it results showed 11,812 nt with 50.1% GC content and gene bank accession no is MF773566.

| Species Name | Whole/Partial Genome Sequencing | Platforms | Size of the genome | GC% | NCBI Gene Bank accession no | References |
|---|---|---|---|---|---|---|
| SARS-CoV-2 isolate Female patient | coding-complete genome sequences | Sanger sequencing | 29,724 | | MT476385.1 | Moniruzzaman *et al.,* 2020 |
| SARS-CoV-2 isolate 3 individuals | WGS | Illumina NextSeq 550 | 29758-29892 | 39 | MT539159, MT539158 and MT539160 | Akter *et al.,* 2020 |
| SARS-CoV-2 sublineage B.1.617.2 strains 15 individual sample | coding-complete genome sequences | amplicon sequencing approach | 29769-29772 | 39.9-42.4 | MZ377102 to MZ377116 | Afrad *et al.,* 2021 |
| SARS-CoV-2 isolate From Noakhali 55 individuals | WGS | Illumina® DRAGEN | >29000 | | MN908947.3 | Hossain *et al.,* 2021 |
| SARS-CoV-2 P.1 variant (20J/501Y.V3) | coding-complete genome sequences | MiniSeq sequencing | 29,789 | 38 | MZ020420 | Sarkar *et al.,* 2021 |
| SARS-CoV-2 Delta variant (B.1.617.2) strain | near-complete genome sequence | MiniSeq sequencing | 29,860 bp | 38.0 | MZ437368 | Banu *et al.,* 2021 |
| BufCoV HKU26 strains B1-24F | WGS | ABI Prism 3700 DNA Analyzer | 31 021 | 40.00 | KU558922 and KU558923 | Lau *et al.,* 2016 |
| BufCoV HKU26 strains B1-28F | WGS | ABI Prism 3700 DNA Analyzer | 30 975 | 40.00 | | Lau *et al.,* 2016 |
| CHIKV | WGS | NextSeq 500 machine(paired-end sequencing) | 11,812 | 50.01 | MF773566 | Pyke *et al.,* 2020 |

**Table 5.** Genome sequencing of viruses in Bangladesh

**Bacteria**

There have been some reports of whole genome sequencing of bacterial strains using next generation sequencing such as Illumina sequencing technology. The first genome sequence of *Pasteurella multocida* BAUTB2 isolated from a buffalo that died from hemorrhagic septicemia in Rajshahi, Bangladesh (Sarker *et al.,* 2018). Using Illumina HiSeq technology, the BAUTB2 genome length was determined to be 2,439,149 bp with 40.8% GC content (Table 6). The heavy metal tolerant strain, *Bacillus anthracis* FHq, isolated from the tannery effluents of Savar, Bangladesh. Whole-genome sequencing analysis revealed that the genome of the strain was around 5.2 Mbp long, and the G + C content was 35.4% (Table 6). The objective of the study was to identify significant genes that are included in the resistance of heavy metals (Haque *et al*., 2022).

| Species Name | Whole/Partial Genome Sequencing | Platforms | Size of the genome | Ax. No of the genes | GC% | NCBI Gene Bank accession no | References |
|---|---|---|---|---|---|---|---|
| *Pasteurella multocida* BAUTB2 | WGS | Illumina HiSeq technology | 2,439,149 bp | **2,307** | 40.8 | QLYS00000000 | Sarker *et al.,* 2018 |
| Bacillus anthracis FHq strain | WGS | Illumina Miniseq technology | 5.2 Mbp/ 5238597 bp | 6045 | 35.4 | PRJNA668995 (NCBI Bioproject) | Haque *et al.,* 2022 |
| *Citrobacter freundii SRS1* | WGS | Illumina Miniseq technology | 5.4 Mbp | 5472 | 51.7 | PRJNA643771 (NCBI Bioproject) | Uddin *et al.,* 2022 |
| *Chromobacteriuma mazonense* BA- SUSDA_45 strain's | WGS | Illumina MiSeq platform | 4,295,151 bp | 4181 | 62.3 | JAFCYX0 0 0 0 0 0 0 0 0.1 | Rahman *et al.,* 2023 |

**Table 6.** Genome sequencing of bacteria in Bangladesh

The whole-genome sequencing of another gram-negative bacteria *Citrobacter freundii SRS1* revealed that the estimated genome size was to be 5.4 Mbp long, and the G+C content was 51.7% (Uddin *et al*., 2022). This study aim was to identify heavy metal-resistant genes, including arsenic resistant genes in *C. freundii SRS1* genome. Another article reported the draft genomic sequence *Chromobacterium amazonense* BA- SUSDA_45 strains' (Rahman *et al*., 2023). The bacterium was isolated from cypermethrin pesticide contaminated soil and then the sequencing was carried out. The genome contains 53 scaffolds with a total length of 4,295,151 bp having 62.30% GC content. Annotation using Prokaryotic Genome Annotation Pipeline (PGAP) reveals 4181 genes among which 4096 were coding sequences, 76 tRNAs, 3 rRNAs, 4 noncoding RNAs. The strain is capable in degrading pesticide (cypermethrin) supplemented

into the media as the sole carbon source for their growth, development and metabolism. But the mechanism is yet to be revealed how this bacterium use and utilize pesticide for their survival. So, genome sequencing of this strain can helps the researchers to give detail insights of the particular pathways that involve in these processes.

**Fungus (*Macrophomina phaseolina*) genome sequencing**

In 2012, the same group of scientists who decoded the genome of jute also sequenced the *Macrophomina phaseolina*, a Botryosphaeriaceae fungus, which is responsible for causing seedling blight, root rot, and charcoal rot of more than 500 crop and non-crop species throughout the world (Wyllie, 1988). The sequencing took place at the laboratory of Bangladesh Jute Research Institute and was done as part of The Basic and applied Research on Jute project (Daily Sun, 2013). *M. phaseolina* strain MS6 was isolated from an infected jute plant at Bangladesh Jute Research Institute (BJRI), Dhaka. Whole-genome shotgun sequencing of the *M. phaseolina* MS6 strain was performed using the 454 and Illumina sequencing platforms funded by the government of Bangladesh (Islam *et al*., 2012). It was done to gain insight of *M. phaseolina* into the molecular basis of pathogenesis. The resulting assembly was 49.29 Mb of which 98.53% is nongapped sequence. They predicted 14,249 protein-coding genes and 9,934 were validated by the transcriptome. Whole genome analysis showed that *M. phaseolina* is distinct from those of other known phytopathogenic fungi.

| Species Name | Whole/Partial Genome Sequencing | Platform | Size of the genome | Ax. No of the genes | NCBI Gene Bank accession no | References |
|---|---|---|---|---|---|---|
| *Macrophomina phaseolina* | WGS | 454 and Illumina platforms | 49.295MB | 14,249 | AHHD00000000 | Islam *et al.*, 2012 |

**Table 7.** Genome sequencing of fungus in Bangladesh

The genome is distinct from other fungi by having the highest number of cytochrome P450, glycosidase, and secondary metabolite backbone genes. They predict that this might be one of the main strategies of *M. phaseolina* to overcome the host plant defense response by using various secondary metabolites. This Whole Genome Shotgun project was deposited at GenBank under the accession no AHHD00000000 (Table 7).

These are just a few examples of the genome sequencing projects that have been conducted in Bangladesh, and there are likely many more ongoing or completed studies that have not been reported publicly.

**Application of genomic knowledge in practical life**

The applications of genome sequencing are wide-ranging and diverse. In medicine, genome sequencing is used for the diagnosis and treatment of genetic diseases. It allows doctors to identify the specific genetic mutations responsible for a patient's disease and to develop personalized treatments based on their genetic profile. Genome sequencing is also used in cancer research, where it is used to identify mutations that contribute to the development and progression of cancer. In biotechnology, genome sequencing is used to develop new drugs and therapies. It allows researchers to identify potential drug targets and to design drugs that are more effective and less toxic. Genome sequencing is also used in agriculture, where it is used to develop crops that are more resistant to pests and diseases.

**Future Prospects of Bangladesh to genome industry**

The genomics market is increasing continuously and Bangladesh has a huge opportunity to enter this market. According to Fortune Business insights, the global genomics market is projected to grow from $27.81 billion in 2021 to $94.65 billion in 2028 at CAGR of 19.4% in forecast period, 2021-2028.

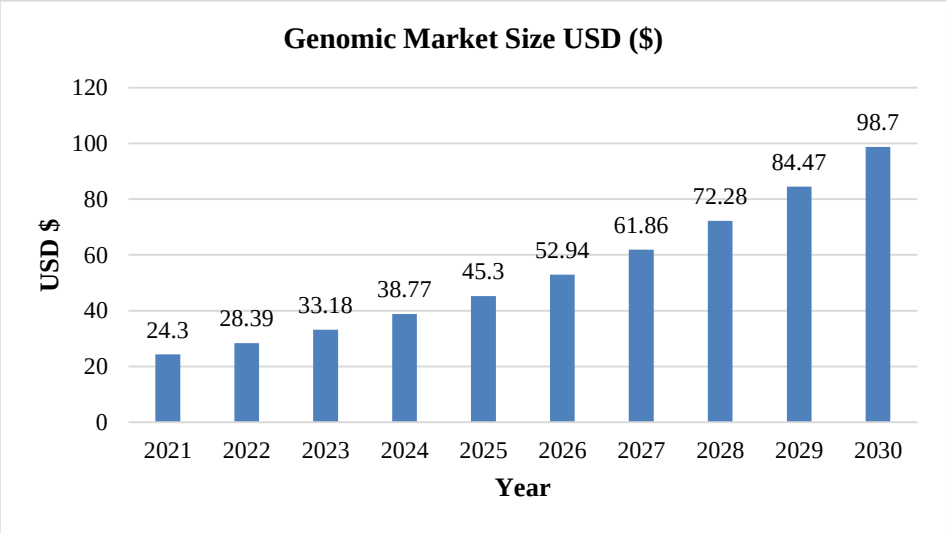


**Figure 7.** Prediction of the growth of genome market in world.

(Source- https://www.precedenceresearch.com/genomics-market)

In recent years, advances in technology have drastically reduced the cost of genome sequencing, as scientists can now reference previous sequences to help them put new ones together more quickly. This has spurred a new wave of companies that are offering genome sequencing services to patients. Today, it is possible to sequence a human genome for as little as $1,000, but some companies that offer this service are looking to bring that number down to as low as $100, furthering the potential for widespread adoption (Fortune Business insight,

2022). Currently, Governments of various nations are supporting and investing in public and private research institutions. Their focus is to emphasize efforts to resolve the complexity of the human genome, define the genomic basis of human health and disease, and ensure that genomics is used safely to enhance patient care and benefit society through government, public, and private institutions. In recent years, governments in various countries have made significant investments in the field of genomics for the development of new technologies. According to Fortune business insights, some examples of developments are given below on this front.

1. In January 2020, the Department of Biotechnology (DBT) initiated the "Genome India Project" (GIP) – the project aims to collect 10,000 genetic samples from citizens across India to build a reference genome. Some of the areas of focus of this project are precision health, rare genetic disorders, mutation spectrum of genetic and complex disease in the Indian population, genetic epidemiology of multifactorial lifestyle diseases, and translational research.
2. In October 2019, The National Center for Advancing Translational Sciences (NCATS) funded USD 7 million to HudsonAlpha Institute for Biotechnology in Huntsville, Alabama for the research purpose to investigate the application of cutting-edge DNA sequencing technologies that could potentially enhance diagnosis and treatment of many common and rare diseases.
3. In September 2019, a whole-genome sequencing project of USD 224 million was created, forming a partnership of pharmaceutical firms and health experts, which will examine and sequence the genetic code of 500,000 volunteers at the UK Biobank, based in Stockport, UK. The project aims to improve health through genetic research and improve the prevention, diagnosis, and treatment of a wide range of serious and life-threatening illnesses, including cancer, heart diseases, diabetes, arthritis, and dementia.
4. In February 2019, the minister of Science and Sport (Canada) announced funding of USD 22.7 million in addition to funding of USD 33.4 million by provincial governments to support 36 research projects through Genome Canada. The projects include various sectors such as health, agriculture, anural resources, and the environment.

Moreover, favourable regulations from government bodies, a robust number of pipeline projects, and emerging application areas such as aquaculture are opening a new door for market growth. So, Bangladesh can grab this opportunity to enter into the global genome market. Actually, Genome sequencing is not a rocket science. If any lab has the potentiality to conduct PCR tests, can learn genome sequencing. But, researchers need to be consistent and patient, use their resources optimally and analyze the data intelligently. Despite these

promising developments, genome sequencing in Bangladesh still faces significant challenges. Limited funding, lack of trained personnel, and infrastructure deficiencies are among the key issues that must be addressed to further advance this field. Another challenge is the lack of proper regulation and oversight in the use of genetic information. There is a need for policies and guidelines to ensure that genomic data is used ethically and responsibly, particularly in the fields of medicine and agriculture. To address these challenges, the government of Bangladesh and international research institutions must work together to build research capacity, invest in infrastructure, and establish ethical guidelines for genomic research. In addition, there must be greater public awareness and education about the benefits and risks of genome sequencing to encourage participation in research and ensure that the benefits are widely shared. Although the good news is that several universities in Bangladesh have started offering courses and degree programs in bioinformatics and genomics, which will help to address the shortage of trained personnel.

# CHAPTER 4

# CONCLUSION

In conclusion, the first generation genome sequencing technique was time-consuming and tiring, while NGS utilizes automation and miniaturization to complete large number of reactions more efficiently. In contrast, the third generation sequencing techniques have the ability to directly observe the incorporation of nucleotides in real time. On the contrary, it is becoming much easier and cost-effective due to the advancement of technologies. For instance, the first human genome project required 13 years and 3 billion dollars whereas now $1000 is required to get result within a couple of weeks. Bangladesh has been already sequenced many nationally important plants, animals and microorganisms including jute which was announced in 2010. The same research group unveiled the code of *M. phaseolina* and it was the first whole genome sequence in Bangladesh that was published and submitted to NCBI in 2012. In the following years, the national fish Ilish, national fruit jackfruit, black Bengal goat and so on were also sequenced. During COVID-19 pandemic scientists in Bangladesh also sequenced the coding nucleotide of SARSCoV-2 isolates from different individuals for identifying genomic characteristics and evolutionary relationships. Moreover, genome sequencing of many important viruses like chikungunya and some bacteria isolates was also revealed to identify the specific gene. The progress made so far has set the stage for further research and development in the field of genomics in Bangladesh and has the potential to contribute significantly to the country's scientific and economic growth. However, several challenges still need to be addressed, particularly in terms of funding, infrastructure, and regulation. In addition, genomics market will be tripled in next 7-8 years which is a great opportunity for Bangladeshi entrepreneurs. The continued development of new technologies, genome sequencing is likely to become faster, more accurate, and more accessible, enabling further advances in medicine, agriculture, forensics, and ecology.

## CHAPTER 5

## REFERENCES


Afrad, M. H., Khan, M. H., Rahman, S. I. A., Bin Manjur, O. H., Hossain, M., Alam, A. N., ... &Qadri, F. (2021). Genome sequences of 15 SARS-CoV-2 sublineage B. 1.617. 2 strains in Bangladesh. *Microbiology Resource Announcements*, *10*(28), e00560-21.

Akter, S., Banu, T. A., Goswami, B., Osman, E., Uzzaman, M. S., Habib, M. A., ... & Khan, M. S. (2020). Coding-complete genome sequences of three SARS-CoV-2 strains from Bangladesh. *Microbiology resource announcements*, *9*(39), e00764-20.

Arick, M. A., Grover, C. E., Hsu, C. Y., Magbanua, Z., Pechanova, O., Miller, E. R., ... & Peterson, D. G. (2023). A high-quality chromosome-level genome assembly of rohu carp, Labeorohita, and its utilization in SNP-based exploration of gene flow and sex determination. *G3: Genes, Genomes, Genetics*, *13*(3), jkad009.

"Bangladeshi scientists decode genome of jute fungus". *Daily Sun*. Archived from the original on 7 July 2013.

Banu, T. A., Sarkar, M. M. H., Akter, S., Goswami, B., Jahan, I., Osman, E., ... & Khan, M. S. (2021). Genome sequencing of the SARS-CoV-2 Delta (B. 1.617. 2) variant of concern detected in Bangladesh. *Microbiology Resource Announcements*, *10*(48), e00849-21.Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing?. *Archives of Disease in Childhood-Education and Practice*, *98*(6), 236-238.

Buffalo, F. W. (2000). An asset undervalued. United Nations Food and Agriculture Organization. website http://www.aphca.org/publica- tions/files/w_buffalo.pdf.

Das, A., Ianakiev, P., Baten, A., Nehleen, R., Ehsan, T., Ahmed, O., ... & Khan, H. (2018). Genome of *Tenualosa ilisha* from the river Padma, Bangladesh. *BMC Research Notes*, *11*(1), 1-3.

Declercq, W., Vandenabeele, P., & Saelens, X. (2019). Walter Fiers (1931–2019). *Cell*, *179*(6), 1241-1243.

DGHS (2023), National Statistics, COVID-19 Dyanamic dashboard for Bangladesh. Website-dghs.gov.bd, Accessed 8 april 2023.

DoF. 2017. Yearbook of Fisheries Statistics of Bangladesh, 2016-17 . Fisheries Resources Survey System (FRSS), Department of Fisheries. Bangladesh: Ministry of Fisheries and Livestock.

DoF. 2020. Yearbook of Fisheries Statistics of Bangladesh, 2019-20. Fisheries Resources Survey System (FRSS), Department of Fisheries. Bangladesh: Ministry of Fisheries and Livestock.

FAO. 2020. The State of World Fisheries and Aquaculture 2020: Sustainability in action. Rome, Italy: FAO (The State of World Fisheries and Aquaculture (SOFIA)).

Fortune Business Insight, Website- https://www.fortunebusinessinsights.com/industry-reports/genomics-market-100941.

"Genome sequencing of local jute disclosed". *bdnews24.com*. Retrived 16 June 2010.

Goldmann, J. M. (2019). *Characterization of de novo Mutations in the Human Germline* (Doctoral dissertation, [Sl]:[Sn]).Guzvic, M. (2013). The History of DNA Sequencing/ISTORIJAT SEKVENCIRANJA DNK. *Journal of Medical Biochemistry*, *32*(4), 301.

Haque, F., Jabeen, I., Keya, C. A., &Shuvo, S. R. (2022). Whole-genome sequencing and comparative analysis of heavy metals tolerant Bacillus anthracisFHq strain isolated from tannery effluents in Bangladesh. *AIMS microbiology*, *8*(2), 227.

Hayden, E. C. (2012). Nanopore genome sequencer makes its debut. *Nature*, *10*.

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics 107*(1), 1-8.

Holley, R. (1968) Alanine transfer RNA, Nobel lecture. [online] Available from: from https://www.nobelprize.org/uploads/2018/06/holley-lecture.pdf

Hood, L. E., Hunkapiller, M. W., & Smith, L. M. (1987) Automated DNA sequencing and analysis of the human genome. *Genomics 1*(3), 201-12.

Hossain, M., Huq, T. S., Rahman, A., Islam, M. A., Tabassum, S. N., Hasan, K. N., ... & Reza, H. M. (2021). Novel mutations identified from whole-genome sequencing of SARS-CoV-2 isolated from Noakhali, Bangladesh.

Husain, S.S.(1993). A study on the productive performance and genetic potentials of Black Bengal goats. *A Ph.D. Thesis*, Bangladesh Agricultural University, Mymensingh.

Illumina, Inc (2017) Illumina Introduces the NovaSeq Series—a New Architecture Designed to Usher in the $100 Genome. [online].

Ilyas, M. (2017). Next-generation sequencing in diagnostic pathology. *Pathobiology*, *84*(6), 292-305.

Islam, M. S., Haque, M. S., Islam, M. M., Emdad, E. M., Halim, A., Hossen, Q. M. M., ... &Alam, M. (2012). Tools to kill: genome of one of the most destructive plant pathogenic fungi Macrophominaphaseolina. *BMC genomics*, *13*, 1-16.

Islam, M. S., Saito, J. A., Emdad, E. M., Ahmed, B., Islam, M. M., Halim, A., ... &Alam, M. (2017). Comparative genomics of two jute species and insight into fibre biogenesis. *Nature plants*, *3*(2), 1-7.

Islam, T., Afroz, N., Koh, C., Hoque, M. N., Rahman, M., Gupta, D. R., ... & Sharpe, A. G. (2022). Whole-genome sequencing of a year-round fruiting jackfruit (*Artocarpus heterophyllus Lam.*) reveals high levels of single nucleotide variation. *Frontiers in Plant Science*, *13*, 5167.

Jou, W. M., Haegeman, G., Ysebaert, M., & Fiers, W. (1972). Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature*, *237*(5350), 82-88.

Kass, D.H., Batzer, U., &Batzer, M.A. (2003). Genome Organization : Human Advanced Structures and Processes Genetics and Molecular Biology human genome # genome organization # DNA sequence complexity # gene families # chromosomes.

Khan, S., Akter, S., Goswami, B., Habib, A., Banu, T. A., Barton, C., ... & Hossain, M. (2021). Whole genome mapping and identification of single nucleotide polymorphisms of four Bangladeshi individuals and their functional significance. *BMC Research Notes*, *14*, 1-6.

Lau, S. K. P., Tsang, A. K. L., Ahmed, S. S., Alam, M. M., Ahmed, Z., Wong, P. C., ... & Woo, P. C. Y. (2016). First genome sequences of buffalo coronavirus from water buffaloes in Bangladesh. *New microbes and new infections*, *11*, 54-56.

Lewin B (2004). *Genes VIII* (8th ed.). Upper Saddle River, NJ: Pearson/Prentice Hall. ISBN 978-0-13-143981-8.

Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION sequencing and genome assembly. *Genomics, proteomics & bioinformatics*, *14*(5), 265-279.

Md, S. J., Uddin, A. M. M. B., Md, P. S., Tanmay, M. H., & Rahman, F. (2016). Livelihood status of hilsa (Tenualosa ilisha) fishermen of greater Noakhali regions of Bangladesh. *Fish Aquac J*, *7*(168), 2.

Michelizzi, V. N., Dodson, M. V., Pan, Z., Amaral, M. E. J., Michal, J. J., McLean, D. J., … Jiang, Z. (2010). Water buffalo genome science comes of age. *International Journal of Biological Sciences*, *6*, 333.

Mintoo, A. A., Zhang, H., Chen, C., Moniruzzaman, M., Deng, T., Anam, M., ... & Liang, X. (2019). Draft genome of the river water buffalo. *Ecology and Evolution*, *9*(6), 3378-3388.

Moniruzzaman, M., Hossain, M. U., Islam, M. N., Rahman, M. H., Ahmed, I., Rahman, T. A., ... &Salimullah, M. (2020). Coding-complete genome sequence of SARS-CoV-2 isolate from Bangladesh by sanger sequencing. *Microbiology Resource Announcements*, *9*(28), e00626-20.

NCBI (2023). 'NCBI sequence set browser'. Website-https://www.ncbi.nlm.nih.gov/Traces/wgs/. Retrived on 10 april, 2023.

Precedence research, Genome Market (Online), https://www.precedenceresearch.com/genomics-market

Pyke, A. T., McMahon, J., Burtonclay, P., Nair, N., & De Jong, A. (2020). Genome Sequences of chikungunya virus strains from Bangladesh and Thailand. *Microbiology Resource Announcements*, *9*(2), e01452-19.

Rahman, M. A., Sujon, K. M., Habib, M. T., Hossain, M. F., Ferdaus, K. M. K. B., Hoque, K. M. F., ... & Reza, M. A. (2023). Whole genome sequencing data of Chromobacteriumamazonense BASUSDA_45 isolated from soil in Bangladesh capable of degrading pesticide. *Data in Brief*, *46*.

Roth, S. C. (2019). What is genomic medicine?. *Journal of the Medical Library Association: JMLA*, *107*(3), 442.

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., ... & Smith, M. (1977). Nucleotide sequence of bacteriophage φX174 DNA. *Nature*, *265*(5596), 687-695.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, *74*(12), 5463-5467.

Sarker, M. S. A., Rahman, M. T., Mahmud, M. M., Tagliamonte, M. S., Chowdhury, S. M. Z. H., Islam, M. R., ... &Nazir, K. N. H. (2018). First Genome Sequence of Pasteurella multocida Type B Strain BAUTB2, a Major Pathogen Responsible for Mortality of Bovines in Bangladesh. *Microbiology Resource Announcements*, *7*(9), e00901-18.

Sarkar, M. M. H., Rabbi, M. F. A., Akter, S., Banu, T. A., Goswami, B., Jahan, I., ... & Khan, M. S. (2021). Genome sequence of a SARS-CoV-2 P. 1 variant of concern (20J/501Y. V3) from Bangladesh. *Microbiology Resource Announcements*, *10*(27), e00524-21.

Satzinger, H. (2008). Theodor and Marcella Boveri: chromosomes and cytoplasm in heredity and development. *Nature Reviews Genetics*, *9*(3), 231-238.

Schadt, E. E., Turner, S., & Kasarskis, A. (2010). A window into third-generation sequencing. *Human molecular genetics*, *19*(R2), R227-R240.

Schnekenberg, R. P., &Németh, A. H. (2014). Next-generation sequencing in childhood disorders. *Archives of disease in childhood*, *99*(3), 284-290.

Siddiki, A. Z., Baten, A., Billah, M., Alam, M. A. U., Shawrob, K. S. M., Saha, S., ... & Islam, K. (2019). The genome of the Black Bengal goat (*Capra hircus*). *BMC research notes*, *12*(1), 1-3.

Sidhu, A. S. (2012). Jackfruit improvement in the Asia-pacific region: A status report (APAARI) FAO Regional Office for Asia & the Pacific (FAO RAP) Maliwan Mansion, 39, PhraAtit Road Bangkok 10200.

Statistical Bulletin-2014 (Food and Agriculture Organization of the United Nations, 2014).

Statistics., Y. O. A. (2020). Bangladesh Bureau of statistics (BBS), statistics and informatics division (SID), ministry of planning, government of the people's republic of Bangladesh (Government of the People's Republic of Bangladesh).

Uddin, M. J., Haque, F., Jabeen, I., &Shuvo, S. R. (2022). Characterization and whole-genome sequencing of an extreme arsenic-tolerant Citrobacterfreundii SRS1 strain isolated from Savar area in Bangladesh. *Canadian Journal of Microbiology*, *69*(1), 44-52.

"WHO definitions of genetics and genomics". World Health Organization. Archived from the original on June 30, 2004.

*Winkler HL (1920). Verbreitung und Ursache der Parthenogenesis imPflanzen- und Tierreiche. Jena: Verlag Fischer.*

Winnick, E. (2004). DNA sequencing industry sets its sights on the future. *The scientist18*(18),44.

Wyllie, T. D. (1988). Charcoal rot of soybeans: current status.

Yindee, M., Vlamings, B., Wajjwalku, W., Techakumphu, M., Lohachit, C., Sirivaidyapong, S., … Colenbrander, B. (2010). Y-chromo- somal variation confirms independent domestications of swamp and river buffalo. *Animal Genetics, 41*, 433–435.