

End to End Genome Sequencing : Telomere to Telomere Consortium

August 5, 2021

On May 27, a preprint titled “The complete sequence of the human genome” was posted in the online repository bioRxiv . In this preprint, scientists from the Telomere-to-Telomere (T2T) Consortium, an international collaboration of around 30 institutions, reported the most complete sequencing of the human genome until now.

In news: End-to-end genome sequencing

Placing it in syllabus: Science & Technology

Dimensions

- What is Genome Sequencing?
- T2T Consortium Project and the Technology used (Sequencing Technology)
- History of Genome Sequencing Projects
- Importance of the project
- Challenges
- Potential uses of Genome Sequencing in general

Content:

What is Genome Sequencing?

- The genome, or genetic material, of an organism (bacteria, virus, potato, human) is made up of DNA. This resides in the nucleus of every cell of the organism.
- The DNA consists of a double-stranded molecule, each of which is built up by **four bases – adenine (A), cytosine (C), guanine (G) and thymine (T).**

- Every base on one strand pairs with a complementary base on the other strand (A pairs only with T, and C only with G).
- Each organism has a unique DNA sequence which is composed of bases. Each genome contains all of the information needed to build and maintain that organism.
- In humans, a copy of the entire genome—more than 3 billion DNA base pairs—is contained in all cells that have a nucleus.
- **Genome Sequencing** is a **laboratory procedure** that **determines the order of DNA nucleotides**, or bases (the order of As, Cs, Gs, and Ts) that make up an organism's DNA
- If you know the sequence of the bases in an organism, you have identified its unique DNA fingerprint, or pattern.

T2T Consortium Project and the Technology used:

- The **Telomere-to-Telomere (T2T) consortium** is led by researchers at the National Institutes of Health and the University of California, Santa Cruz
- It is an **international collaboration** of around 30 institutions
- The T2T is an **open consortium and all are welcome to join the effort** to generate the first truly complete assembly of a human genome
- It focuses on the **first gapless assembly of a human genome**, finishing each chromosome from one end to the other
- The Telomere-to-Telomere (T2T) consortium announced our v1.0 assembly that includes more than 150 Mbp of novel sequence compared to GRCh38, achieves near-perfect sequence accuracy, and unlocks the most complex regions of the genome to functional study.
- **protein-coding sequences** or protein-coding genes are DNA sequences that get transcribed on ribonucleic acid (RNA)

as an intermediate step.

- These in turn make the proteins responsible for various functions such as keeping the body healthy or determining the colour of the eye – proteins carry out the instructions encoded in the genes.
- The DNA used did not belong to any person.
- According to a report in *Nature*, it was a cell line derived from a tissue known as a **complete hydatidiform mole**.
- This is the tissue that forms when a sperm inseminates an egg that has no nucleus. Hence, this tissue has the chromosomes of just the father.

History of Genome Sequencing Projects:

- The Human Genome Project that began in 1990 gave the first results of the complete human genome sequence in 2003.
- For the first time, we were able to read the blueprint of human life.
- However, though it was announced as the complete human genome, about 15% of it was incomplete.
- Due to limitations of technology, scientists were not able to piece together some repetitive parts of the human genome.
- Solving some of the problems, an updated “complete” version was released in 2013, which still missed out on 8% of the genome.

Importance of the project:

One step Closer to Whole Human Genome Mapping:

- The researchers have nearly completed the job, adding 200 million base pairs and 115 new protein-coding genes to the list.
- They have, in the process, discovered over a hundred new genes that code for proteins. The total size of the

genome they have sequenced is close to 3.05 billion base pairs.

- This adds 200 million base pairs to the last draft of the human genome that was published in 2013.
- With this, we are a step closer to mapping the whole human genome.

Standard of Reference:

- One of the most important uses of this release will be that it forms a standard for comparison in future sequencing attempts
- This sequence of the human genome will be a gold standard of reference for future attempts.

Unprecedented Level of Accuracy:

- The level of accuracy is unprecedented.
- Earlier, Researchers were trying to piece together strands of DNA that were a few hundred base pairs long.
- The technology used by the Telomere-to-Telomere Consortium used sequencing technology that could scan 20,000 base pairs at one go. This is a significant technological feat.

Challenges:

Missing Y- Chromosome Data:

- The present release has no information about the Y chromosome.
- All chromosomes in an arbitrary cell's nucleus are found in pairs – we have 23 pairs of chromosomes in each cell.
- However, the sex cells such as sperm and egg cells contain only one of each pair of chromosomes (haploid cells).
- So, while egg cells always carry a copy of the X chromosome, sperms can carry either an X chromosome or a

Y chromosome.

- The cell line that the researchers studied had an X chromosome only and no Y chromosome. Therefore, information about the Y chromosome is missing in this release.

Chances of Errors:

- It is also not 100% complete. The researchers say that about 0.3% of the genome may have errors.

Potential uses of Genome Sequencing in general:

Predictive medicine:

- The primary purpose of sequencing one's genome is to obtain information of medical value for future care.
- genome sequencing has the potential to increase the ability to act preemptively prior to disease development or commence treatment for a disease that has not yet been diagnosed.

Drug Efficacy Studies (pharmacogenomics):

- Another advantage of genome sequencing is that information regarding drug efficacy or adverse effects of drug use can be obtained.
- The relationship between drugs and the genome is called pharmacogenomics.

Discovering Gene Mutations:

- genome sequence does not change unless influenced by environmental factors, for example, in the development of many cancers
- Genome sequencing can help discovery of rare or novel variants.
- This point can be illustrated by considering BRCA gene mutations, which confer a high risk of breast and ovarian cancer development in women.

- Early discovery of these mutations provides options for prophylactic action.

Mould your thought: What is genome sequencing? With reference to the recent efforts, discuss the importance of Human Genome Sequencing.

Approach to the answer:

- Introduction
- Define Genome Sequencing
- Discuss the T2T consortium results
- Discuss the potential uses of genome sequencing
- Conclusion