



2001: A Base Odyssey

The era of genomics and massive parallel sequencing

Matthias Zepper, PhD

February 24, 2025

NGI Stockholm

<https://ngisweden.scilifelab.se>

2001: Draft assemblies of the human genome are published



Figure 1: The private company Celera [Venter2001] and the International Human Genome Sequencing Consortium [Lander2001] both publish a draft sequence of the euchromatic portion of the human genome.

The overture to the genomic era



A remake of the opening scene by SumoSebi, CC-BY-SA on Wikimedia Commons

Stanley Kubrick's *2001- A Space Odyssey* premieres 2 April 1968

1968: Nobel prize for the interpretation of the genetic code

Nobel Prize in Physiology or Medicine 1968



Photo from the Nobel Foundation archive.

Robert W. Holley

Prize share: 1/3



Photo from the Nobel Foundation archive.

Har Gobind Khorana

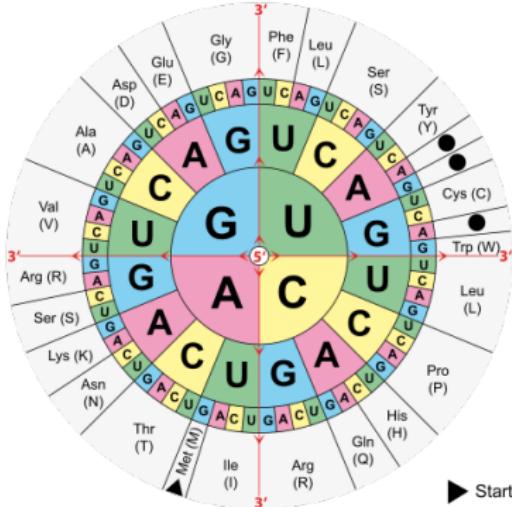
Prize share: 1/3



Photo from the Nobel Foundation archive.

Marshall W. Nirenberg

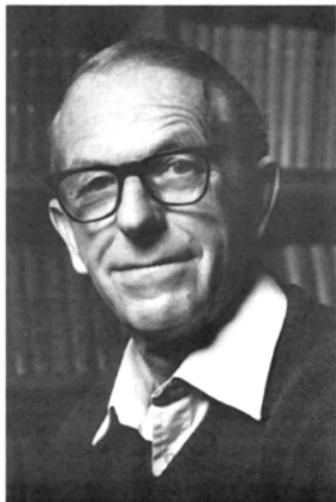
Prize share: 1/3



- The genetic code is (almost) universal^[1]
- It was resolved entirely using synthetic sequences.

[1] <http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/index.cgi?chapter=tgencodes>

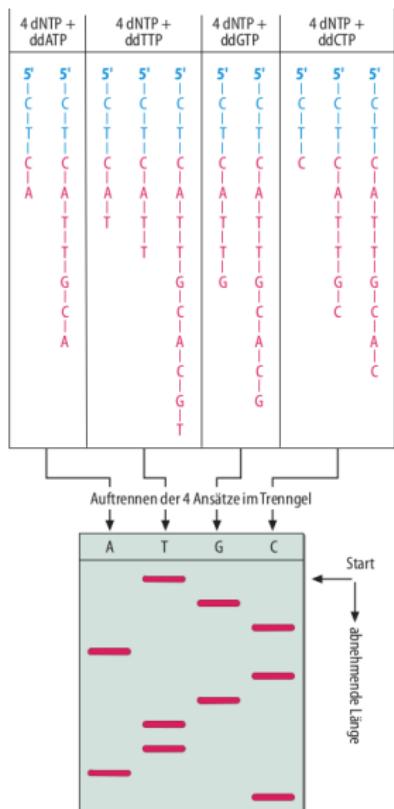
Encoded information of naturally occurring DNA unknown



- Peptides could be sequenced since the 1950s (Sanger method, Edman degradation).
- Sequencing of DNA was one of the most urgent, unresolved problems in the early 1970s.
- Frederick Sanger (Nobel laureate for sequencing Insulin 1958) started working with DNA.

A handwritten signature in cursive script, which appears to read "F. Sanger".

1977: Chain-termination sequencing by Frederick Sanger



- DNA fragments could be separated by size.
- Sanger's method creates sequence-derived length patterns.
- It relies on radioactive labeling and in-vitro amplification of DNA.

DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage ϕ X174)

F. SANGER, S. NICKLEN, AND A. R. COULSON

Medical Research Council Laboratory of Molecular Biology, Cambridge

Figure 2: [Sanger 1977]

1980: Nobel prize for DNA sequencing

Nobel Prize in Chemistry 1980



Photo from the Nobel Foundation archive.

Paul Berg

Prize share: 1/2



Photo from the Nobel Foundation archive.

Walter Gilbert

Prize share: 1/4



Photo from the Nobel Foundation archive.

Frederick Sanger

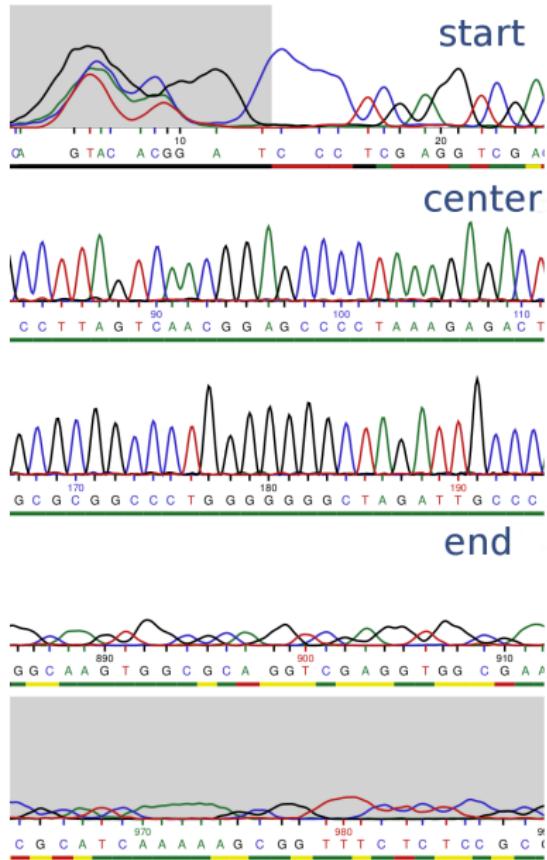
Prize share: 1/4

- Ample DNA input needed
PCR was introduced in 1989
- Four reactions per sequence
- Read length \sim 200bp



<https://www.nobelprize.org/prizes/chemistry/1980/summary/>

Advanced Sanger sequencing for the Human Genome Project



- Fluorescent chain terminators.
- Capillary electrophoresis for size separation of amplicons.
- Parallelized and automated.
- Sequencing technology of the Human Genome Project (1990-2004).

Next-generation sequencing

New high-throughput methods were developed

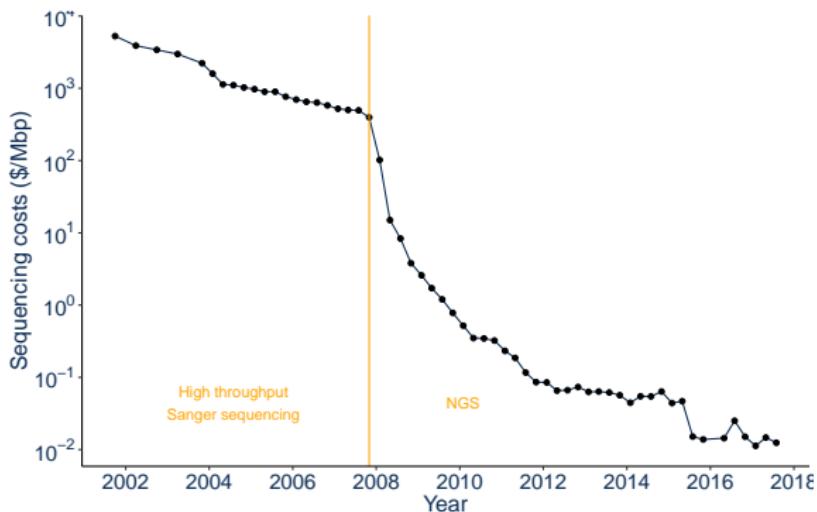


Figure 3: Sequencing costs per one million bases of raw sequence

1990-2004: Human Genome Project sequencing: US \$500 million

2025: Sequencing of a human genome: ~ US \$200-1000

National Human Genome Research Institute (NHGRI)

<https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>

Around 2010: Sanger sequencing was outcompeted by NGS



ABI 3730xl DNA Sequencer
(Sanger Multiplex, 2013)

- ~6912 reads of 400bp
- ~2,76 Mbp / day



Illumina HiSeq 2500
(NGS / MPS, 2013)

- ~600 Million reads of 100bp
- ~60.000 Mbp / day

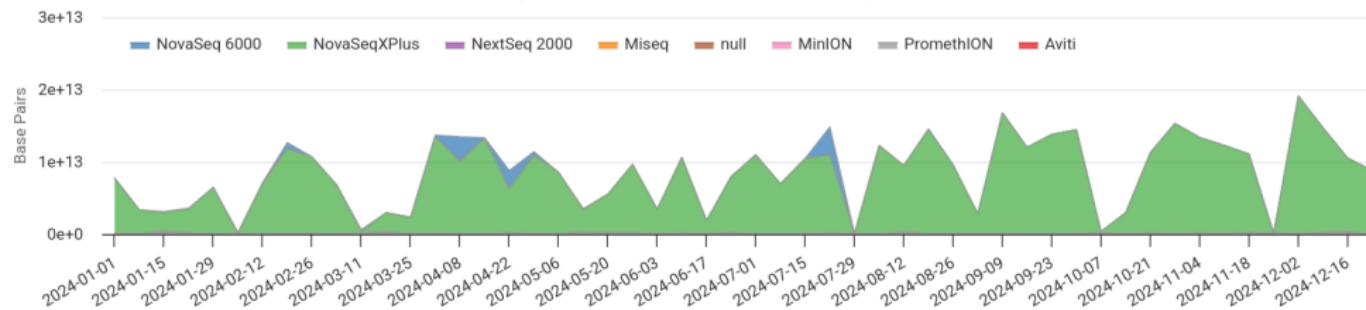
(depending on settings and sequencing chemistry used)

DNA sequencing facilities support life science research



Sequencing Throughput

Average for 52 weeks: 1234 Gbp per day
(1 Human genome equivalent every 3.77 minutes)



- In 2024, NGI Stockholm sequenced on average 1200 Gbp/day

<https://ngisweden.scilifelab.se/resources/ngi-stockholm-status/>

National Genomics Infrastructure Sweden

References

References i