IEEE

BERKSHIRE SECTION

High School STEM Research Challenge – 2024

Next-Generation Gene Sequencing Technology for Medical and Anthropological Usage

Written for:

The Institute of Electrical and Electronics Engineers, Inc.

Berkshire Section
STEM Research Contest

By
Ariel Caine
1305 Swamp Road
Richmond MA 01254
March 4, 2024

Monument Mountain Regional High School
Grade 12
Dr. Elsa Herraez Hernandez

Next-Generation Gene Sequencing Technology for Medical and Anthropological Usage

In 2022, Swedish geneticist Svante Pääbo won the Nobel prize in physiology for his work in the sequencing of the Neanderthal genome. Much like the older Human Genome project, this was a long-running investigation spanning almost two decades, with the first draft published in 2010. Interestingly, there were no new technologies developed for this project specifically- the scientists made creative use of already existing technologies, specifically a combination of Sanger sequencing and massively parallel pyrosequencing, an early type of next-generation sequencing (NGS). (Noonan et al. 2008) This paper will provide an overview of how NGS works, a comparison with Sanger sequencing, and a discussion on its uses and future in the medical and anthropological fields.

Although first developed in 2000, NGS is relatively new compared to other sequencing forms. Its main predecessor is Sanger sequencing, developed in 1975. Sanger sequencing has three main steps: chain-termination PCR, gel electrophoresis, and gel analysis. Chain termination PCR is similar to standard PCR, but includes the addition of a type of modified nucleotide called dideoxyribonucleotides (ddNTPs). Because of the lack of -OH groups in ddNTPS as opposed to standard dNTPs, bonds are not formed and as such copies are cut off at the points where ddNTPs are added. This means that copies of the DNA sequences are terminated at different lengths depending on ddNTP placement. After copying via chain-termination PCR, copies are separated via gel electrophoresis. Smaller segments experience less resistance when moving through the gel, so travel further towards the positive electrode, allowing DNA segments to be organized by size. The gel can be read manually or with an automated laser, with terminal ddNTPs going in order from shortest to longest strands. By reading the order of the ddNTP types, the sequence of

the entire strand can be pieced together. This is why although Sanger sequencing is similar to standard PCR, PCR attempts to sequence a whole strand at once, while Sanger sequencing does it in multiple portions to create a more accurate and reliable picture of the overall sequence. (Sigma-Aldrich)

NGS follows similar steps to Sanger sequencing, but with a few distinct differences. These steps are: fragmentation, library preparation, sequencing, and analysis. Much like in Sanger sequencing, NGS involves breaking the DNA into millions of tiny pieces (this can be done in multiple different ways) and often using (standard) PCR to amplify the pieces to be prepared more easily. During library preparation, markers are attached to the DNA so the sample of origin can be easily identified, as well as attaching sequencing adaptors to be used in sequencing. This sequencing is done through massive parallel sequencing (allowing the millions of pieces to be sequenced together) using an NGS sequencing machine. The bioinformatic analysis is done by a program, which can piece together the millions of sequences to create one cohesive sequence. The program can also compare the new sequence to human sequence databases to look for variation and mutations, which is the factor of interest to anthropologists and those in the medical field. (Qin, 2019) Sanger sequencing and NGS are very similar, but NGS is significantly more efficient and can be used more easily in comparisons between both other experimental sequences and sequence libraries. This is why it poses such a benefit to genetic studies in many fields.

The case of Neanderthal DNA sequencing is the highest-profile example of NGS use in evolutionary biology studies. This is because it reflects a larger ability to sequence ancient DNA, which has historically been extremely difficult. The advent of NGS allowed for a 480x increase in the size of ancient DNA sample available, which is its main benefit in the study of ancient

DNA (aDNA). One fo the main challenges posed by the sequencing of aDNA is that it is often damaged, or only short lengths can be recovered and successfully sequences. One NGS read produces significantly more data than other strategies, and it performs fairly quickly and efficiently. Because of its increased reliability in data recovery, it is ideal for aDNA sequencing in which this is a major issue. Usage is still somewhat limited since the other major issue with aDNA sequencing is high contamination rates (due to bacterial and fungal contamination over time, and researcher contamination, though this can be minimized), but as techniques improve, we can expect huge strides in the field. (Knapp et al., 2010) The use of ultra-clean rooms and unique tags to distinguish aDNA from contamination from researchers were some of the ways used by the team investigated Neanderthal DNA to minimize the effects of contamination. (Max Planck Society, 2010) The study of aDNA will provide a great deal of insight in evolutionary biology and may even allow us to bring species back from extinction.

The other main field of interest for NGS is the medical field. As discussed earlier, the digitization of NGS data makes comparison via bioinformatic software significantly easier. This allows us to search for mutations in a much more efficient manner, which will greatly benefit patients with genetic disorders. In combination with CRISPR gene editing technology, NGS could potentially be used to find harmful mutations, which we could then edit with gene editing techniques. NGS can search for multiple mutations at one time, which would be particularly beneficial for examinations of cancers caused by multiple mutations. (Qin, 2019) Overall, the efficiency and broad scope of NGS makes it particularly useful in clinical settings where broader scope and speed are essential.

Although it has been around for around twenty years, NGS is still very much in development. As with all sequencing methods, there are concerns regarding its accuracy, and it

can only combat sample contamination so much. However, it has been proven to be a more advanced version of most other sequencing methods available today, particularly the similar but less efficient Sanger sequencing. As the field of genetics continues to advance, we can expect to see increased usage of NGS, and as it becomes more widespread, more usage in clinical field. I believe that to make full use of this technique, we should focus on its clinical applications and lowering the cost of sequencing machines and tests to make them more available for individuals, as well as investing into research preventing or reversing contamination of samples. Clean samples, both modern and ancient DNA, can tell us about current health issues for individual patients and genetic groups, and it can tell us about where we came from- and how much of you is Neanderthal.

Works Cited

Heather, James M., and Benjamin Chain. “The sequence of sequencers: The history of sequencing DNA.” *Genomics*, vol. 107, no. 1, 2016, pp. 1-8. *ScienceDirect*, https://www.sciencedirect.com/science/article/pii/S0888754315300410#bb0035. Accessed 1 March 2024.

Klein, Richard G. “Profile of Svante Pääbo: 2022 Nobel laureate in physiology or medicine.” *PNAS*, vol. 120, no. 1, 2022. *PNAS*, https://www.pnas.org/doi/10.1073/pnas.2217025119#:~:text=Svante%20P%C3%A4%C3%A4bo%2C%20a%20Swedish%20geneticist,all%20those%20of%20the%20Neanderthals. Accessed 1 March 2024.

Knapp, Michael, and Michael Hofreiter. “Next Generation Sequencing of Ancient DNA: Requirements, Strategies and Perspectives.” *Genes* vol. 1,2 227-43. 28 Jul. 2010, doi:10.3390/genes1020227. Accessed 29 February 2024

Max Planck Society. *The Neandertal in us*. Press release. 6 May 2010, https://web.archive.org/web/20100516135912/http://www.eva.mpg.de/neandertal/press/presskit-neandertal/pdf/PR_MPI_Neandertal_EN.pdf. Accessed 4 March 2024.

McGroarty, Patrick. “Team in Germany maps Neanderthal genome.” *The Associated Press*, 2 December 2009, https://web.archive.org/web/20120926021129/http://readingeagle.com/article.aspx?id=125240. Accessed 4 March 2024.

Noonan, James P., et al. “Sequencing and Analysis of Neanderthal Genomic DNA.” *Science*, 2008,

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2583069/pdf/nihms72167.pdf. Accessed 29 February 2024.

Orlando, L., Allaby, R., Skoglund, P. *et al.* Ancient DNA analysis. *Nat Rev Methods Primers* 1, 14 (2021). https://doi.org/10.1038/s43586-020-00011-0

Qin, Dahui. "Next-generation sequencing and its clinical application." *Cancer biology & medicine* vol. 16,1 (2019): 4-10. doi:10.20892/j.issn.2095-3941.2018.0055

Sigma-Aldrich. "Sanger Sequencing Steps & Method." *Sigma-Aldrich*, https://www.sigmaaldrich.com/US/en/technical-documents/protocol/genomics/sequencing/sanger-sequencing. Accessed 4 March 2024.

Svante Paabo. *Nobel Prize lecture: Svante Pääbo, Nobel Prize in Physiology or Medicine 2022*. 8 December 2022. *YouTube.com*, YouTube, https://www.youtube.com/watch?v=FXgW1WZV0rc&t=2s. Accessed 29 February 2024.