

10-1-2010

Technology in Genomics and Bioinformatics

Timothy Hall

Olivet Nazarene University, mhall2@live.olivet.edu

Follow this and additional works at: http://digitalcommons.olivet.edu/csis_stec

 Part of the [Bioinformatics Commons](#), [Biotechnology Commons](#), and the [Genomics Commons](#)

Recommended Citation

Hall, Timothy, "Technology in Genomics and Bioinformatics" (2010). *Technology Essay Contest Winners*. Paper 4.
http://digitalcommons.olivet.edu/csis_stec/4

This Essay is brought to you for free and open access by the Computer Science at Digital Commons @ Olivet. It has been accepted for inclusion in Technology Essay Contest Winners by an authorized administrator of Digital Commons @ Olivet. For more information, please contact kboyens@olivet.edu.

Timothy Hall

Technology Essay

22 November 2010

Technology in Genomics and Bioinformatics

Since the discovery of DNA structure, given by the Watson/Crick model, in 1953, technological advances in the field has skyrocketed understanding of the subject. However, as further advances are made it becomes ever-increasingly more imminent that there is less and less that is actually understood. Many Mendelian advances and basic protein property discoveries in the 1960s and 70s led to DNA recombinant technologies, or single-gene study procedures, and sequencing techniques in the 80s and 90s. The Sanger Method, developed by Frederick Sanger in 1977, utilized radioactive isotope tagging and radiography to sequence DNA by hand. Although a rather visionary method, the sequencing throughput of the method was very low. Much more time was saved with the invention of computational laboratory tools such as capillary-based fluorescence sequencers, which utilizes fluorescent tagging and laser technology to analyze base pair sequences. Faster sequencing techniques such as these made the dream of sequencing entire genomes a reality. In 2001, the entire human genome was completely sequenced. The human genome, along with several other recently sequenced genomes, are available to anyone without charge on the National Center for Biotechnology Information (NCBI) database. Because of the availability and convenience of not only DNA sequences but also gene and mutation information, it is evident that the future of bioinformatics will have a prominent impact on the medical community.

The numerous technological advances in genomics in the last 50 years have occurred on multiple levels of the discipline. Three levels of important technological advances are laid out: sequencing, DNA recombinant technology, and databasing. In 1977, the Sanger Method introduced sequencing techniques. The method used isotopes to radioactively label nucleotides. Modified nucleotides, called dideoxynucleotides, were in DNA replication techniques to create many fragments of DNA of varied sizes. With the addition of the modified nucleotide to any given sequence, the process was stopped. Then, the fragment sampling, being radioactively labeled, was sent through a polyacrylamide gel that separated the fragments by size. The basis of this gel is that smaller fragments can slip through the porous gel further than larger fragments can. There are four wells in the gel that code for each of the four nucleotide bases (A, G, C, T). The gel is then viewed as an x-ray, because of the radioactivity, and the sequence can be attained. Because the smaller fragments are at the bottom, the sequence is read from bottom to top.

The invention of fluorescence labeling of DNA ushered in a new era of automated sequencing. This area of automated sequencing was led by capillary-based fluorescence sequencing. The process follows the same principles of size separation, labeling, and sequencing. However, this is carried out in a safer and easier procedure. Fluorescent labeling carries nearly no dangers with it, as opposed to the dangers of radioactive labeling. Tools such as thermocyclers run the DNA replication process with the fluorescently labeled nucleotides. The samples are then separated by capillary electrophoresis and sequenced by DNA sequencers. Fluorescent labels come in four varieties, one color for each nucleotide base. Chromatograms are recorded by the sequencer as colored peaks, from which a sequence is produced.

In order to be able to sequence the entire human genome, it had to be broken up into smaller, more manageable chunks and put back together like a puzzle. This process of dividing up genomic chunks was made possible through DNA recombinant technology. This area of the genomics field includes techniques such as Polymerase Chain Reaction (PCR) and the utilization of vectors like Bacterial Artificial Chromosomes (BAC). PCR is a protocol that takes a specified DNA segment, usually a target gene, and creates thousands to millions of copies of it so it can be isolated and studied. BAC's are specific vectors that can be used to carry target genes into organisms for further study. These specific techniques along with several others aided in the development of the Human Genome Project. One example of these developments was realized by Craig Venter. In 1991, he discovered an easier way to sequence genes without sequencing the entire genome. He suggested the idea of isolating mRNA that is produced from exons on DNA, expressed DNA. Then, by utilizing reverse transcriptase, an enzyme used by some viruses to create DNA from RNA, creating a copy of the original DNA sequence used to create to mRNA. After these DNA segments are created, computer programs put them together like puzzle pieces to create a genome map. Venter joined Francis Collins and the rest of the Human Genome Program to catalyze the sequencing of the human genome. It was finished two years in advance, in 2001, due to technological advances and Venter's new technique.

After the metaphorical research journey to the human genome summit was completed, a cornucopia of new questions, topics, and research opportunities opened up. The human genome sequence was compiled in an online database hosted by the National Center for Biotechnological Information (NCBI). This information is presented on this database alongside other protein, RNA, and DNA information provided by PubMed, a

medical journal database, and several other mapping and genomic comparison databases. In the NCBI database, one can search for a certain sequence, gene, or gene product, along with many other search options. Informative pages are available for each gene and gene product that provide relevant links to other databases and websites. PubMed provides medical information in the form of links to medical journal articles about specified genes and mutations. GenBank provides sequencing data for comparison of sequences and protein coding regions within a genome. BLAST allows the user to compare analogous genes and gene products between several different completely sequenced organism genomes. There are also several genomic mapping tools to allow a user to see the location of the specified search item on its specific chromosome in the genome. All of these databases compiled in one location on the NCBI website allows users to utilize many avenues of research to further the development of the discipline.

The future of genomics looks to be cut-out for researchers. There are many avenues that have been made available with the advent of the human genome. One of the largest growing areas in this subject that has been gaining much publicity is the area of Single Nucleotide Polymorphisms (SNP). SNP's are the single base differences within species that makes each individual different. Genetic mutations in chromosome crossing over and other outside factors provide for these changes. The knowledge of the complete human genome and the development and possibility of newer, faster, and cheaper sequencing protocols have opened the door for new medical possibilities. If each person could calculate his or her specific genome, including single nucleotide differences (SNP's), that could lead to development of more personalized treatment therapies. This could also lead to estimation of certain diseases based one SNP differences in genetic make-up.

The advantages that these new advancements provide are numerous and varied. The advent of new technologies provides faster sequencing throughput, making the opportunity available to allow for the sequencing of an entire genome to be completed in twenty-four hours. The input of sequencing data and information into large databases distributes it across the world and provides the ability for comparisons between genes, gene products, mutations, and comparisons between species. The fact that these databases can be accessed instantly will help further catalyze not only developments in genomics but also in the medical field.

References

- McPherson, John. "Next-generation gap." *Nature* 6 (2009): 2-5. *Nature: Methods*. Web. 23 Feb. 2010.
- OMICS. "Whole Genome Sequencing in 24 Hours." *Genetic Engineering and Biotechnology News* 25.15 (2005): n. pag. *GEN: Omics Channel*. Web. 22 Feb. 2010.

