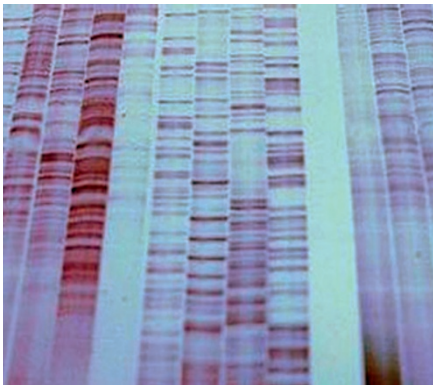
A brief history of (DNA sequencing) time

Elaine R. Mardis

Each spring, I co-teach a course to undergraduates that aims to introduce them to the brief history of genome sequencing and then immerses them in its current practices. Looking out at their eager faces, I try (perhaps in vain) to capture for them in a 1 hour lecture what has been my life's focus for the past 20 years or so — DNA sequencing. If forced to summarize this time period succinctly, I would say, "Never a dull moment!" This is largely owing to the technological advances that have catapulted genome sequencing from a cottage industry to a high-throughput enterprise, akin to a factory. That said, the constant winds of change that swirl around DNA sequencing are about to change the paradigm yet again, owing to next-generation approaches.

The history of DNA sequencing is a brief one — first described by Fred Sanger in 1977, merely 30 years ago. In that brief time, progress has been staggering, owing in large part to interdisciplinary innovations that have built on the fundamental and elegant concept of dideoxynucleotide termination. The milestones that I describe to my students are familiar territory to me, and I strive to convey the ingenuity and intelligence that were required to make each one a reality, and to capture the definitive impact of each discovery on our capabilities. Indeed, where would DNA sequencing be today without synthetic oligonucleotides, fluorescence resonance energy transfer dideoxynucleotides, high-fidelity polymerases and cycle sequencing? Taken alone or together, each of these discoveries allowed a quantum leap in sequencing data quality and throughput. Can my students ever appreciate the transition from radioactive to fluorescent labelling? Although it took sequencing laboratories several years to trust the data quality from fluorescent sequencers, this

> "… the technological advances have catapulted genome sequencing from a cottage industry to a high-throughput enterprise, akin to a factory."



transition not only allowed automated detection and base calling of sequence reads, but also made sequencing safer and more readily automated. What about the development of automated picking robots that replaced manual harvesting with sterile toothpicks? The automated imaging of agar plates coupled with actuator-driven steel pins for harvesting colonies was key to high-throughput pipelines, and saved the eyesight and careers of many technicians. How about the transition from hand-held to robotic multi-pipettors? This provided another crucial step in high-throughput automated pipelines, especially at 384-well plate densities. Or consider, only 8 years ago, when slab-gel sequencers were replaced by capillary-array sequencers?

Eliminating hand loading, image files and gel-lane tracking allowed a tremendous acceleration of throughput and capacity with a drastic decrease in labour. What about the profound impact of finely-tuned bioinformatics tools that allow the assembly, annotation and analysis of a genome sequence, teasing out its every secret by clever approaches? Genomics simply would not be possible without bioinformatics.

It was this delicate integration of chemistry, engineering, enzymology, separation sciences and software that allowed us to first describe the human genome sequence, and that now has set the stage for even greater accomplishments at a previously unimagined pace. In fact, with next-generation technology, we now have an amazingly accurate genome-wide readout for many types of experiments, not merely for re-sequencing genomes. With next-generation-based applications, we rapidly will begin to annotate the human genome with the positions of regulatory protein-binding sites and other functional sites using sequencing as a means of evaluating chromatin immunoprecipitation (ChIP) experiments. A variant of ChIP will elucidate how genome-wide patterns of histone binding and DNA methylation relate to gene-expression regulation in various cell states, such as differentiation or disease. The short read-lengths of next-generation platforms also are ideal for discovering the sequences of non-coding RNA genes that cannot be elucidated by *in silico* methods. Bioinformatics-based analysis of genome-wide DNA copy number, DNA-sequence variation and RNA-expression data sets, all produced by next-generation sequencers, will be integrated to stitch together the pieces of the biological 'story' being told by genomic data. These stories will, in turn, help us to make sense of our clinical observations, and suggest cures to aberrant biological states. Present day sequencing technology advances have the potential to revolutionize the translation of genomic information into clinical practice.

To have been a part of sequencing up to now is to have a unique perspective on how far we have come and how our accomplishments rest on the shoulders of those innovators. I pause to wonder what I will be teaching my students in the years to come, and quickly realize that presently I might not even be able to imagine it. So, at the end of our hour together, I choose to leave the past behind, and attempt to invigorate my students with the same sense of potential for questions yet to be answered that I first felt in graduate school and still feel to this day. The context and potential are not that different now from then. Certainly, they will have an ever-increasing number of next-generation sequencing instruments, applications and bioinformatics tools available to them; moreover, the reach of sequencing as an approach to answering genome-scale biological questions will be greater than ever. My hope is that they not only will respect the brief but illustrious past of DNA sequencing, but also will choose to become a part of its promising future.

Elaine R. Mardis is Associate Professor in Genetics and Molecular Microbiology, and Co-Director of the Genome Sequencing Center at the Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, Missouri 63108, USA.
e-mail: emardis@watson.wustl.edu

doi: 10.1038/mrg2240