Downloaded from genome.cshlp.org on October 7, 2020 - Published by Cold Spring Harbor Laboratory Press

Perspective

Emerging technologies in DNA sequencing

Michael L. Metzker

Human Genome Sequencing Center and Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA

Demand for DNA sequence information has never been greater, yet current Sanger technology is too costly, time consuming, and labor intensive to meet this ongoing demand. Applications span numerous research interests, including sequence variation studies, comparative genomics and evolution, forensics, and diagnostic and applied therapeutics. Several emerging technologies show promise of delivering next-generation solutions for fast and affordable genome sequencing. In this review article, the DNA polymerase-dependent strategies of Sanger sequencing, single nucleotide addition, and cyclic reversible termination are discussed to highlight recent advances and potential challenges these technologies face in their development for ultrafast DNA sequencing.

More than just a mapping and sequencing endeavor, the Human Genome Project (HGP) has altered the mindset and approach to many basic and applied research efforts. Early skepticism and controversy (Koshland 1989; Luria et al. 1989; Roberts 1989b; Fox et al. 1990) were soon laid to rest by well-developed strategies (Roberts 1989a; Collins and Galas 1993; Collins et al. 1998) that led to the successful execution of mankind's largest biology project. At the core of the HGP was technology development that advanced the pace of sequencing a mammalian-size genome from years to months. Along the way, numerous strategies emerged that hold promise for rapid, efficient, and inexpensive delivery of DNA sequence information. For the HGP, a brute-force approach was adopted for completing the job by coupling the core technologies of Sanger sequencing and fluorescence detection. The completion of the sequencing phase could not have been accomplished without major innovations in recombinant protein engineering, fluorescent dye development, capillary electrophoresis, automation, robotics, informatics, and process management. The result was completion of a high-quality, reference sequence of the human genome in April, 2003 (Collins et al. 2003), marking the 50-year anniversary of the discovery of the double-helix structure. For many outside the genome community, that heroic milestone signaled the end of this international scientific project, but for the rest of us, it only marked the beginning of things to come.

The need for sequencing has never been greater than it is today, with applications spanning diverse research sectors including comparative genomics and evolution, forensics, epidemiology, and applied medicine for diagnostics and therapeutics. Arguably, the strongest rationale for ongoing sequencing is the quest for identification and interpretation of human sequence variation as it relates to health and disease. The most common form of variation is the single nucleotide polymorphism (SNP). Although two unrelated people share, on average, 99.9% sequence identity (i.e., one difference in a thousand base pairs), the average occurrence of an SNP in the general population is once every few hundred base pairs. As such, more than nine million unique SNPs have been cataloged in the public database, dbSNP (Crawford and Nickerson 2005), with many more expected to be found in large-scale resequencing efforts.

A great deal of attention has been focused on common SNPs with a minor allele frequency >5% and their potential role in common disease (Lander 1996; Risch and Merikangas 1996; Collins et al. 1997). Recent, large-scale genotyping efforts of these common SNPs have shown that much of the human genome can be parsed into common haplotype blocks (Daly et al. 2001; Patil et al. 2001; Gabriel et al. 2002). The International HapMap Consortium (2003) was formed to characterize common patterns of sequence variation by determining allele frequencies and the degree of association between SNPs among geographically distinct groups, leading to the identification of "tagSNPs" for genome-wide, disease-based association studies. With this method of characterization, however, rare SNPs/haplotypes may be overlooked, as highlighted by Liu et al. (2005), who described an association of rare variants/haplotypes with osteoporosis.

A shift in large-scale strategies from genotyping to resequencing is currently taking place to explore the significance of less-common SNPs to human biology and disease. The "re" in this approach is the sequencing of additional genomes related to a reference genome for de novo SNP discovery and comparative genomics application. The ENCODE Project Consortium (2004) has described significant efforts toward resequencing megabase-sized blocks of the human genome. Consequently, genome centers are now diverting at least 10%–20% of their resources, which currently translates to ~5% capacity, to resequencing hundreds to thousands of gene regions. This increase in momentum for high-throughput resequencing will greatly facilitate studies to determine the genetic basis of susceptibility to common disease, cancer biology, and disease association in model and nonmodel organisms.

Current sequencing technologies are too expensive, labor intensive, and time consuming for broad application in human sequence variation studies. Genome center cost is calculated on the basis of dollars per 1000 Q_{20} bases (defined below) and can be generally divided into the categories of instrumentation, personnel, reagents and materials, and overhead expenses. Currently, these centers are operating at less than one dollar per 1000 Q_{20} bases, with at least 50% of the cost resulting from DNA sequencing instrumentation alone. Developments in novel detection methods, miniaturization in instrumentation, microfluidic separation technologies, and an increase in the number of assays per run will most likely have the biggest impact on reducing cost. It should be emphasized, however, that new sequencing strategies will be needed to use these high-throughput platforms effectively. In September, 2004, the National Human Genome Re-

E-mail mmetzker@bcm.tmc.edu; fax (713) 798-5741.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.3770505.

QIAGEN Exhibit 1026.0001

Downloaded from genome.cshlp.org on October 7, 2020 - Published by Cold Spring Harbor Laboratory Press

search Institute (NHGRI) initiated two new programs aimed at bringing the cost of whole-genome sequencing down to $100,000 (http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-002.html), with the eventual goal being $1000 (http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-003.html).

Numerous strategies and platforms for ultrafast DNA sequencing currently under development include sequencing-by-hybridization (SBH), nanopore sequencing, and sequencing-by-synthesis (SBS), the latter of which encompasses many different DNA polymerase-dependent strategies. Use of the term SBS has become increasingly ambiguous in the literature; therefore, I propose a classification of DNA polymerase-dependent strategies into three major categories: *Sanger sequencing*, *single nucleotide addition* (*SNA*), and *cyclic reversible termination* (*CRT*) (Text Box 1). In this review, I will focus only on DNA polymerase-dependent strategies, which represent the broadest area of research and development. For the SNA and CRT strategies, I will emphasize the chemistry in an effort to illustrate the advantages and challenges of these methods. Because of the competitive nature of technology development, the exchange of scientific ideas is often thwarted, as many companies do not readily publish results. Although this review will highlight recent advances reported in the literature, readers are directed to the Web sites of companies who are active in the sequencing field (Table 1). A recent review by Shendure et al. (2004) provides a comprehensive overview of SBH and nanopore sequencing technologies. Important issues surrounding whole-genome sequencing, such as ownership, consent, privacy, and legal, ethical, and social implications, will not be addressed here (Foster and Sharp 2002; Robertson 2003; Bonham et al. 2005).

Table 1. Companies involved in DNA sequencing technology development

Company names	Web site addresses
454 Life Sciences Corp.	www.454.com
Agencourt Biosciences Corp.	www.agencourt.com
GE Healthcare, formerly Amersham Biosciences	www.amershambiosciences.com
Applied Biosystems, Inc.	www.appliedbiosystems.com
Genovoxx	www.genovoxx.de
Helicos Bioscience Corp.	www.helicosbio.com
LaserGen, Inc.	www.lasergen.com
Li-Cor, Inc.	www.licor.com
Microchip Biotechnologies, Inc.	www.mcbiotech.com
Nanofluidics	www.nanofluidics.com
SeqWright	www.seqwright.com
Solexa-Lynx	www.solexa.com
Visigen Biotechnologies, Inc.	www.visigenbio.com

Sanger sequencing: State-of-the-art technology

The Sanger method is a mixed-mode process involving synthesis of a complementary DNA template using natural 2′-deoxynucleotides (dNTPs) and termination of synthesis using 2′,3′-dideoxynucleotides (ddNTPs) by DNA polymerase (Sanger et al. 1977). Balanced appropriately, competition between synthesis and termination processes results in the generation of a set of nested fragments, which differ in nucleoside monophosphate units. The ratio of dNTP/ddNTP in the sequencing reaction determines the frequency of chain termination, and hence the distribution of lengths of terminated chains. The nested fragments are then separated by their size using high-resolution gel electrophoresis and analyzed to reveal the DNA sequence. Advancements in fluorescence detection (Smith et al. 1986; Prober et al. 1987), enzymology (Tabor and Richardson 1989, 1995), fluorescent dyes (Ju et al. 1995; Metzker et al. 1996; Lee et al. 1997), dynamic-coating polymers and their derivatives (Ruiz-Martinez et al. 1993; Carrilho et al. 1996; Madabhushi et al. 1996, 1999; Madabhushi 1998; Salas-Solano et al. 1998; Guttman 2002a, 2002b), and capillary array electrophoresis (CAE) (Takahashi et al. 1994; Kheterpal et al. 1996) have helped to define current DNA sequencing platforms.

For automated Sanger sequencing, either the primer or the terminating ddNTP is tagged with a specific fluorescent dye (e.g., ddATP is labeled with the green dye). As these dye-labeled fragments pass through the detection region, fluorophores are excited by the laser in the DNA sequencer, producing fluorescence emissions of four different colors. The determination of the color is the underlying method for assigning a base call, and the order of the fluorescent fragments reveals the DNA sequence. The "raw" fluorescence signals, however, must be transformed. Removal of cross-talk, correction for dye mobility alterations, and normalization of emission intensities must be performed before readable DNA sequence information can be obtained (Smith et al. 1987). Base-calling and error probability assignment (Ewing and Green 1998; Ewing et al. 1998) applications are then used to call the DNA sequence and assess the accuracy of the call. A Phred_{20} or Q_{20} score, equivalent to an error probability of 1% for a given base call, is considered a high-quality base and serves as the commodity standard throughout the sequencing community.

Text Box 1. DNA polymerase-dependent strategies

In the broadest sense, all methods involving a DNA polymerase could be considered a SBS approach, if synthesis alone was the defining process. The defining element of these DNA polymerase-dependent methods, however, is not really synthesis at all but rather the means by which DNA synthesis terminates. From this point of view, the DNA sequencing approaches highlighted here have been organized according to their termination strategies. *Sanger sequencing* and "dideoxy" sequencing are frequently used as synonymous terms.

These unnatural ddNTP terminators replace the OH with an H at the 3′-position of the deoxyribose molecule and irreversibly terminate DNA polymerase activity, unless the nucleotide is removed by the process of phosphorolysis. This process is mediated by high concentrations of pyrophosphate or ATP and is a major cause of "drop-outs" in DNA sequence data.

Single nucleotide addition (SNA) methods such as pyrosequencing use limiting amounts of individual natural dNTPs to cause DNA synthesis to pause, which, unlike the Sanger method, can be resumed with the addition of natural nucleotides. Limiting the amount of a given dNTP is required to minimize misincorporation effects observed at higher concentrations. A major drawback with the SNA approach is the incomplete extension through homopolymer repeats.

Cyclic reversible termination (CRT) uses reversible terminators containing a protecting group attached to the nucleotide that terminates DNA synthesis. For the reversible terminator, removal of the protecting group restores the natural nucleotide substrate, allowing subsequent addition of reversible terminating nucleotides. One example of a reversible terminator is a 3′-*O*-protected nucleotide (Fig. 4B), although protecting groups can be attached to other sites on the nucleotide as well. This stepwise base addition approach, which cycles between coupling and deprotection, mimics many of the steps of automated DNA synthesis of oligonucleotides.

QIAGEN Exhibit 1026.0002

Downloaded from genome.cshlp.org on October 7, 2020 - Published by Cold Spring Harbor Laboratory Press

High-throughput DNA sequencing is conducted primarily at large genome centers that continue to refine the sequencing process and strive for Q_{20} bases at lower cost. For example, the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC) produces approximately four million sequencing reactions per month (R.A. Gibbs, pers. comm.). The current production efficiency or pass rate is approximately 89% (after removal of failed reactions, vector sequences, etc.), with sequencing reads averaging 805 Q_{20} bases in length. These metrics translate into the equivalent of sequencing one mammalian-size genome per month. Redundancy is required to improve the base-calling accuracy and contiguity of assembled genomes, resulting in the generation of six times the genome size in Q_{20} bases for production of a draft-quality sequence. Thus, delivery of a mammalian-size, draft-quality sequence requires approximately six months and $12 million. Ongoing advances in new technologies will be critical to meet the goal of rapid, genome-scale sequencing for the price of $100,000 and, ultimately, $1000 per genome.

Sanger sequencing: Recent advances

Microfluidic separation platforms

Technology development remains active for the fluorescence-based Sanger approach with emphasis on producing faster and cheaper sequencing reads. One key area of research is the application of microfluidic separation devices to DNA sequencing. These microfluidic devices can be fabricated using a variety of substrate materials, with several molecular biology processes integrated onto a single device (e.g., lab-on-a-chip). A number of reviews have been devoted to microfluidic devices (Becker and Gartner 2000; Carrilho 2000; McDonald et al. 2000; Quake and Scherer 2000; Boone et al. 2002; Paegel et al. 2003; Kan et al. 2004), recent advances of which I will highlight as they relate to DNA sequencing. These miniature devices have several advantages over CAE, including improved sample injection and faster separation times.

The separation principles of microfabricated devices are similar to those of conventional CAE, however, their injection methods are very different. With CAE, the sample is introduced by electrokinetic injection into the capillary. The injection time, which defines the length of the sample plug, is typically short and allows only a minute fraction of the sample to be analyzed. A further drawback is that data quality is compromised with increasing impurities in the sample and an intrinsic bias in favor of shorter DNA fragments over longer ones. Microfluidic devices, on the other hand, are less susceptible to these injection problems because the sample is introduced via a channel network by a variety of process strategies (Zhang and Manz 2001). Although early microfabricated chips employed a "T"-injector design (Harrison et al. 1992), the cross-T design (Harrison et al. 1993) is widely used today because of its superior sample control (Fig. 1A). The narrow width of the injector affords greater control in selection of sample plug size, which contributes to higher resolution separations with shorter separation lengths compared with CAE.

Most microfabricated devices use borofloat glass or fused silica substrates, which have the advantages of (1) high-quality optical properties, (2) good thermal conductivity, (3) well-documented surface chemistry, and (4) effective translation of capillary innovations. Woolley and Mathies (1995) demonstrated the first application of DNA sequencing using a microfabricated glass device in 1995, reporting single-base resolution

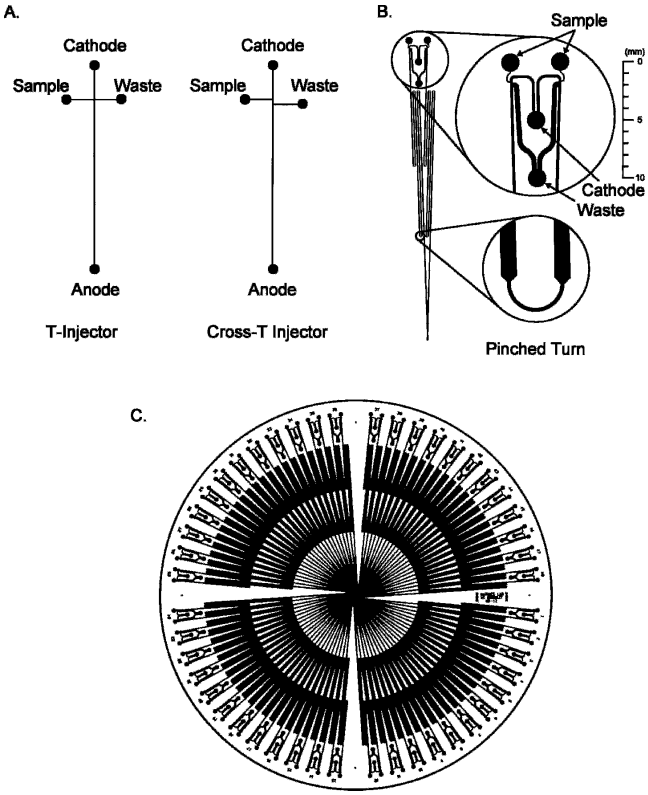


Figure 1. Microfabricated technologies. (*A*) Examples of a T-injector and cross-T injector layout. (*B*) Expanded view of the sample injector and pinched turn. (*C*) Schematic of the 96 channels in a radial chip design. (*B,C*) Reprinted with permission from National Academy of Sciences, U.S.A. © 2002, Paegel et al. 2002.

using their four-color scanner technology. Data quality and read-lengths have improved significantly since then, because of an increase in the effective separation lengths with run times of 30 minutes or less (Table 2) (Woolley and Mathies 1995; Liu et al. 1999; Schmalzing et al. 1999; Backhouse et al. 2000; Koutny et al. 2000; Liu et al. 2000; Salas-Solano et al. 2000; Simpson et al. 2000; Boone et al. 2002; Paegel et al. 2002; Shi and Anderson 2003). For example, Liu et al. (1999) reported 99.4% accuracy over 500 bases in 20 minutes, with an increase in separation length from 3.5 cm to 6.5 cm. More recent developments by Boone et al. (2002) and Shi and Anderson (2003) have shown the first DNA sequencing applications on plastic chips (Table 2). These chips can be fabricated with high geometric aspect ratios (i.e., deep and narrow channels) at significantly lower cost. Deep and narrow channel structures have the advantages of improved electrophoretic resolution (i.e., longer read-length) and better detection sensitivity.

While single-channel devices are useful for demonstrating feasibility, the construction of multiple channel arrays is essential for high-throughput DNA sequencing. A summary of DNA sequence metrics from several microfabricated multiple channel array devices is presented in Table 2. While Backhouse et al. (2000) and Koutny et al. (2000) reported improved read-lengths by increasing the effective separation lengths to 46.5 cm and 40 cm, respectively, these microfabricated channels were constructed on glass plates ≥50 cm in length, which is out of line with current efforts to miniaturize devices. One approach to circumvent this dilemma has been the introduction of turns along

QIAGEN Exhibit 1026.0003

Downloaded from genome.cshlp.org on October 7, 2020 - Published by Cold Spring Harbor Laboratory Press

Table 2. Summary of microfabricated devices for DNA sequencing applications

Research group	Number of channels	Template source	Separation length (cm)	Accuracy (%)	Read-length (bp)	Read-out time (min)
Single channel						
Woolley & Mathies (1995)	1	M13mp18	3.5	97	147	9
Liu et al. (1999)	1	M13mp18	6.5	99.4	500	20
Schmalzing et al. (1999)	1	M13 clones[a]	11.5	99	505	27
Salas-Solano et al. (2000)	1	M13mp18	11.5	98.5	640	30
Boone et al. (2002)	1	M13mp18	18.0	98	640	30
Shi & Anderson (2003)	1	Unknown	4.5	99.1	320	13
Multiple channel arrays						
Liu et al. (2000)	16	M13mp18	7.5	99	457	16
Simpson et al. (2000)	48	M13mp18	10.0	97	400	25–45
Backhouse et al. (2000)	48	BigDye Std	46.5	98	640	150
Koutny et al. (2000)	32	M13 clones[a]	40.0	98	800	78
Paegel et al. (2002)	96	M13mp18	15.9	99	430	24

[a]Mixture of M13mp18 vector or twelve M13 clones from human chromosome 17 project.

the length of the separation channel. Early studies, however, reported lower separation efficiency in channel turns due to band broadening (Jacobson et al. 1994) and differential field strength effects (Culbertson et al. 1998). Paegel et al. (2000) introduced a "pinched-turn" design (Fig. 1B) with an effective separation length of 15.9 cm on a 15-cm-diameter silica disc, which has been multiplexed into a 96-channel radial device (Fig. 1C) showing tremendous potential for increasing throughput in DNA sequencing applications (Paegel et al. 2002). Most of the data shown in Table 2, however, were derived using the standard M13mp18 vector as the sequencing template, and similar performance is not typically observed under the same conditions with "real-world" samples such as those from genome center production lines.

Fluorescence detection

The most widely used detection method for four-color DNA sequencing was initially described almost 20 years ago (Smith et al. 1986; Prober et al. 1987). This method is based on resolution of the emission signal from a dye-labeled nucleotide into color, with subsequent assignment in the DNA sequence. While successful for the sequencing of numerous higher and lower eukaryotic and prokaryotic genomes, these four-color systems have several disadvantages, including inefficient excitation of the fluorescent dyes, significant spectral overlap, and inefficient collection of the emission signals. The issue of inefficient excitation has been partially addressed by the use of fluorescence resonance energy-transfer (FRET) dyes (Ju et al. 1995; Metzker et al. 1996; Lee et al. 1997). At present, FRET dye-labeled ddNTP terminators are widely used throughout the sequencing community. The resulting improvements in acceptor dye signal intensities, however, are suboptimal compared with those of single dyes excited at their absorption maxima by the appropriate laser source.

To overcome these deficiencies, some investigators have proposed strategies using additional properties such as fluorescence life-time (Nunnally et al. 1997; Lieberwirth et al. 1998; Lassiter et al. 2000; Zhu et al. 2003, 2004) and radio frequency (RF) modulation (Alaverdian et al. 2002). For DNA sequencing applications, fluorescence life-time measurements have been described using pulsed lasers with high repetition rates (picosecond time-scale) with detection in the photon-counting mode. Soper and colleagues have recently demonstrated a combined approach of emission wavelength and fluorescence life-time measurements, with the potential to increase the number of fluorescent components in DNA sequencing assays (Zhu et al. 2003, 2004). Alaverdian et al. (2002) proposed using four continuous wave (CW) mode lasers, which are modulated at different RFs. To estimate the fluorescence signal for each dye, however, the resulting emission intensity pattern must be demodulated, which introduces a significant computational load for each capillary signal channel. Coupled with repetition rates on the order of ≥100 Hz, the RF method does not appear to be compatible with conventional CCD technology, limiting its scalability for detection of high-density capillary arrays.

Recently, Lewis et al. (2005) described a simple but effective method for multifluorescence discrimination called pulsed multiline excitation (PME). The underlying principle of this four-laser system is the correlation of sequential laser pulses with detector response (Fig. 2A). Advantages of PME are such that (1) absorption maxima for the four fluorescent dyes are matched to the excitation sources yielding maximum signal intensities, (2) temporal separation of the laser pulses and expansion of the dye set across the visible spectrum eliminate cross-talk between the dyes, and (3) collection of emission signals is improved by eliminating the requirement for dispersing elements (prisms or gratings) in color separation. In other words, PME measures multicomponent fluorescence assays in a color-blind manner. To demonstrate these advantages, Lewis et al. (2005) applied the PME technology to capillary electrophoresis for DNA sequencing. Figure 2B shows the unprocessed signals from the four PME laser waveforms for a portion of the PCR amplicon for the *TCF1* (formerly known as *HNF1A*) exon 10. Transformation of the data into unambiguous sequence data (Fig. 2C) is accomplished by applying only dye mobility correction software, eliminating the need for cross-talk and signal normalization software transformation. The PME technology holds promise for real-time field applications for DNA sequencing.

SNA methodologies

Pyrosequencing

Arguably the most successful non-Sanger method developed to date is pyrosequencing, first described in the literature by Hyman (1988). Pyrosequencing is a nonfluorescence technique that measures the release of inorganic pyrophosphate, which is proportionally converted into visible light by a series of enzymatic reactions (Ronaghi et al. 1996, 1998). Unlike other sequencing approaches that use 3′-modified dNTPs to terminate DNA

QIAGEN Exhibit 1026.0004

Downloaded from genome.cshlp.org on October 7, 2020 - Published by Cold Spring Harbor Laboratory Press

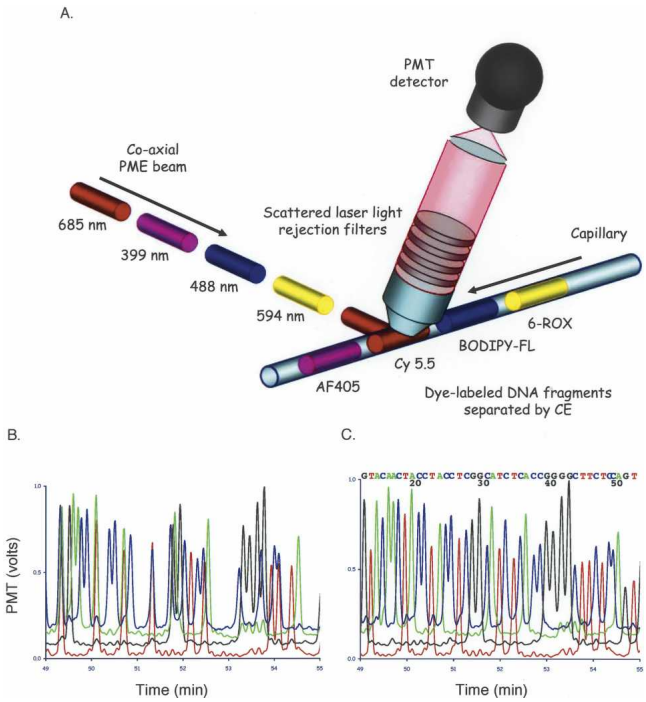


Figure 2. (*A*) Illustration of the PME technology. Here, each laser operates in a CW mode with mechanical shutters pulsing the different excitation beams in sequential order. The single coaxial PME beam interrogates the fluorescently labeled DNA fragments, which are separated by capillary gel electrophoresis. Scattered laser light is rejected via specific long-pass or wavelength notch filters, with pulsed emission signals from the dye-labeled DNA fragments being detected by the photomultiplier tube (PMT) without use of any dispersing elements. (*B*) Unprocessed fluorescence data obtained during the electrophoretic run for the *TCF1* exon 10 gene region using PME dye-primers. Blue, green, black, and red traces are AF-405, BODIPY-FL, 6-ROX, and Cy5.5 dye-primers terminated with ddCTP, ddATP, ddGTP, and ddTTP respectively. (*C*) Transformation of the raw trace data derived from the experiment described in *B* into readable, DNA sequence data using mobility software correction. Reprinted with permission from National Academy of Sciences, U.S.A. © 2005, Lewis et al. 2005.

synthesis, the pyrosequencing assay manipulates DNA polymerase by single addition of dNTPs in limiting amounts. Upon addition of the complementary dNTP, DNA polymerase extends the primer and pauses when it encounters a noncomplementary base. DNA synthesis is reinitiated following the addition of the next complementary dNTP in the dispensing cycle. The light generated by the enzymatic cascade is recorded as a series of peaks called a pyrogram, which corresponds to the order of complementary dNTPs incorporated and reveals the underlying DNA sequence. Applications for pyrosequencing have been reviewed by Ronaghi (2001) and Langaee and Ronaghi (2005).

Although elegant in design, the pyrosequencing approach has several limitations. For example, sequence reads are typically fewer than 100 bases in length, which has application in sequence tag identification such as serial analysis of gene expression (SAGE) (Velculescu et al. 1995), mini-sequencing for known SNPs, and mapping related genomes to a reference sequence, but limited application for whole-genome sequencing. Recent reports describe the use of single-stranded binding protein (Ronaghi 2000) and the isomeric Sp form of the dATPαS nucleotide (Gharizadeha et al. 2002), which may improve read-lengths up to 100 bases in routine settings. Secondly, homopolymer repeats greater than five nucleotides cannot be quantitatively measured. This is attributed to incomplete extension by DNA polymerase, which results from limiting the dNTP concentration to minimize nucleotide misincorporation effects. It has been suggested that re-addition of the same dNTP may be performed to ensure complete polymerization (Ronaghi 2001), although its practicality for high-throughput sequencing is unclear. Finally, the dispensing order of dNTPs determines the pyrogram profile, which must be carefully designed to avoid asynchronistic extensions of heterozygous sequences.

For a given dispensing order, approximately one half of all heterozygous sequences will result in asynchronistic extensions past the variable site. A survey of heterozygous variants detected by direct DNA sequencing of the *TCF1* gene revealed that 16 of 37 SNPs would result in nonsynchronistic extension after the heterozygous base (data not shown). If one allele extends past the heterozygous base position before the other and advances to the next nucleotide cycle, the nonsynchronicity becomes permanent. An illustration of the effect of dispensing order on asynchronistic extension is shown in Figure 3A. This observation is further highlighted by Entz et al. (2005) with the identification of more than 40 unique dispensing orders for the accurate typing of *HLA-DQB1* and *HLA-DRB1* alleles. Pyrosequencing may, therefore, be suited for pattern matching of known SNP profiles, while its application for de novo SNP discovery is less certain. Not surprisingly, base-calling for de novo SNPs is problematic and still performed manually (Langaee and Ronaghi 2005).

The 454 Corporation has recently introduced a whole-genome sequencing strategy by integrating pyrosequencing with their PicoTiterPlate (PTP) platform, which has been shown to amplify and image approximately 300,000 PCR templates captured on Sepharose beads (Leamon et al. 2003). The PTP is manufactured by anisotropic etching of a fiber optic faceplate with a well diameter of approximately 40 µm. The 454 group has developed a solution-based emulsion strategy to create microreactors for clonal amplification of single DNA molecules and attachment to these beads. One advantage of the clonal amplification strategy is that it addresses the dependence issue of dispensing order for sequencing of heterozygous bases discussed above. Following an enrichment step, DNA positive beads are loaded into individual PTP wells, which contain additional beads coupled with the necessary enzymes to perform the pyrosequencing chemistry (Margulies et al. 2005). Recently, the company announced its first complete genome sequencing of a recombinant adenoviral construct and the shotgun sequencing of the *Mycoplasma genitalium* genome.

The assembly of non-Sanger sequencing data will represent new challenges because the input read will differ in length, quantity, and quality. The complexity of the genome under analysis may also prove more difficult for assemblies compared with Sanger data, even when the offset is higher coverage of shorter reads. Chaisson et al. (2004) recently performed a simulated assembly study (short, error-free reads sampled at 30× coverage) using genome sequences from adenovirus, two mouse BACs, and two bacteria: *Campylobacter jejuni*, which contains very few repeat sequences (Parkhill et al. 2000b), and *Neisseria meningitidis*, which contains several hundred repetitive elements (Parkhill et al. 2000a). Compared with Sanger data, Chaisson et al. (2004) found that the read-length was inversely proportional to the number of contigs in the assembly (i.e., longer reads gave fewer contigs). Increasing genome complexity, on the other hand, directly increases the number of contigs. Here, they found that 95% of the genome was contained within 9–10 contigs for the

QIAGEN Exhibit 1026.0005

Downloaded from genome.cshlp.org on October 7, 2020 - Published by Cold Spring Harbor Laboratory Press

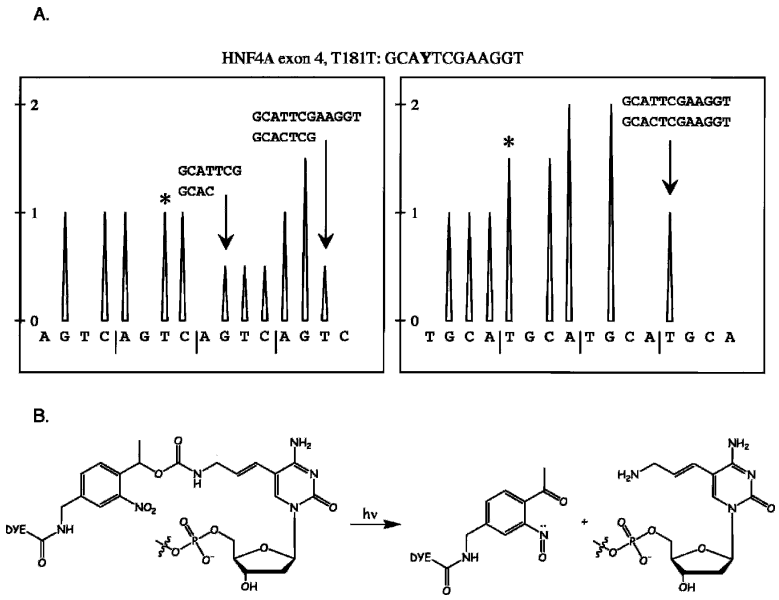


Figure 3. SNA technologies. (*A*) Simulated effects of two different dNTP dispensing orders on the outcome of the pyrogram profile. (*B*) The photocleavage reaction of a fluorescently labeled dNTP coupled with a photocleavable linker.

BAC clones, and the number of contigs increased from 21 to 344 for *C. jejuni* and *N. meningitides* genome sequences, respectively. Observed errors for real sequence data will undoubtedly decrease assembly performance for short reads. Thus, the success of the non-Sanger strategies for whole-genome sequencing applications will be highly dependent on the degree of its complexity, which appears to traverse all three phylogenetic domains.

Other single addition dNTP strategies

Methods other than pyrophosphate detection can be used to monitor single dNTP additions. For example, Braslavsky et al. (2003) used the technique of single-pair FRET (spFRET) to determine the order of nonconsecutive nucleotide additions. With this single molecule approach, Cy3-labeled-UTP was initially incorporated into the primer strand, serving as the donor dye. Subsequent incorporation of a complementary Cy5-labeled-UTP or Cy5-labeled-dCTP substrate resulted in the spFRET signal. Following photobleaching of the Cy5 dye, the natural nucleotides dATP and dGTP were added to increase the nucleotide distance between subsequent Cy5-labeled dNTP additions, which would otherwise have resulted in a significant reduction in incorporation efficiencies due to steric hindrance effects. For the DNA template sequence, written 3′-**A**TC**G**TC**A**TC**G**-5′ for convenience, the read-out would be the fingerprint sequence of 5′-UCUC. Levene et al. (2003) have recently described a zero-mode waveguide approach to single-molecule detection of R110-labeled-dCTP and coumarin-labeled-dCTP incorporation events by DNA polymerase.

Taking advantage of the steric effects observed in consecutively incorporated dye-labeled dNTPs, Mitra et al. (2003) introduced fluorescently labeled dNTPs, which contained cleavable linkers, to remove the bulky fluorescent group following incorporation by DNA polymerase. This method, called *fluorescent in situ sequencing* (FISSEQ), used linkers containing either a disulfide bridge, which is efficiently cleaved with a reducing agent, or a photocleavable group (Fig. 3B). Using the polony technology (Mitra and Church 1999), Church and colleagues elegantly demonstrated the addition of single Cy5-SS-dNTPs followed by dye cleavage for accurate DNA sequencing of several templates. The presence of a fluorescence signal corresponding to the dispersing order of the Cy5-SS-dNTPs revealed the DNA sequence. Although read-lengths up to eight bases were demonstrated, several miscalls were reported. One such call resulted from nucleotide read-through. That is, consecutive incorporations of dye-labeled dNTPs can occur (e.g., the sequence 5′-CAGCC was read as 5′-CAGC), presumably with different efficiencies that are dependent on the local DNA sequence context. A second error occurred as a result of a single nucleotide insertion (e.g., the sequence 5′-ATGT was read as 5′-AGTGT). Although more difficult to interpret, it is possible that the residual linker structure, remaining on the nucleobases following dye cleavage, could alter nucleotide specificity and incorporation efficiency of subsequent incoming dNTPs in a sequence-dependent manner. More recently, Seo et al. (2004, 2005) described a similar strategy using four different dye-labeled dNTPs with photocleavable linkers (Fig. 3B) and reported read-lengths of 12 bases. A key advantage of the four-color approach is that all four dNTPs can be assayed simultaneously, although both reports demonstrated use of the single dNTP addition method.

Kartalov and Quake (2004) proposed a different approach to overcome the steric effects of consecutive dye-labeled bases by use of single-addition, same-nucleobase mixtures (e.g., dCTP/TAMRA-labeled ddCTP) as a method for DNA sequencing. The nucleobase mixture strategy serves the dual purpose of dye-labeling for fluorescence detection (reporter phase) and ongoing DNA synthesis of the complementary nucleotide (extension phase). The dNTP and dye-labeled ddNTP concentrations are balanced appropriately so that only a fraction of the primer strands incorporate the dye-labeled ddNTP. The presence of a fluorescence signal reveals the complementary nucleotide in the DNA sequence, but reporters are eliminated from subsequent dNTP additions. With each nucleotide addition, signal loss is inversely proportional to the increased accumulation of termination products. The fluorescence is then quenched by photobleaching before the next nucleobase mixture is dispensed to repeat the process. Configured in a microfluidic device, the average read-length for the mixed nucleobase addition scheme was three bases, which can be partially attributed to signal loss with subsequent base additions. The accuracy of the method is highly dependent on the reporter phase mimicking the extension phase. For example, a simple homopolymer repeat of two bases will be undercalled in the DNA sequence, as the reporter phase will reflect a single base addition while the extension phase will incorporate two bases.

CRT

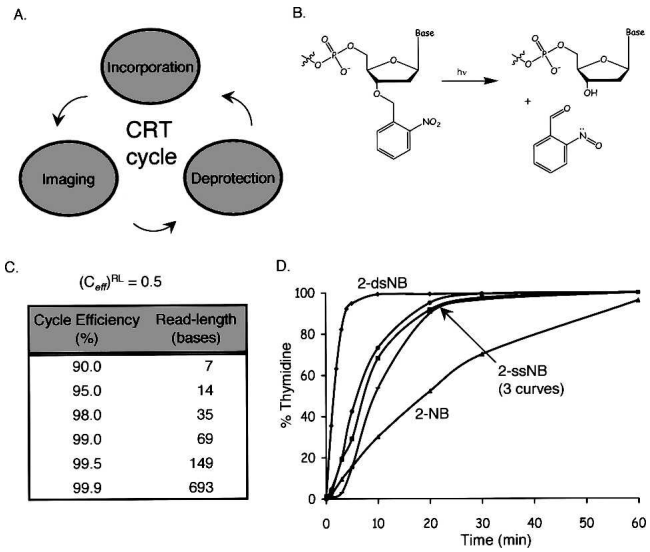
While CRT technology represents tremendous potential for whole-genome sequencing, this strategy still faces significant challenges in its implementation. The CRT cycle is comprised of

QIAGEN Exhibit 1026.0006

Downloaded from genome.cshlp.org on October 7, 2020 - Published by Cold Spring Harbor Laboratory Press

three steps: *incorporation*, *imaging*, and *deprotection*, as illustrated in Figure 4A. The advantages of CRT over Sanger are (1) elimination of gel electrophoresis and (2) formatting of the CRT assay in a highly parallel fashion. Its advantages over pyrosequencing are that (1) all four bases are present during the incorporation phase, (2) step-wise control allows for single-base additions through homopolymer repeats, and (3) synchronistic extensions are maintained past heterozygous bases. An additional advantage is that unlike the pyrosequencing assay, which must be contained within a defined reaction well, the CRT assay can be performed on a number of highly parallel platforms, such as high-density oligonucleotide arrays (Pease et al. 1994; Albert et al. 2003), PTP arrays, (Leamon et al. 2003), polony arrays (Mitra and Church 1999), or random dispersion of single molecules. Albert et al. (2003) have demonstrated the 5′→3′ synthesis of oligonucleotide on a high-density array and the application of incorporation of dye-labeled ddNTPs by DNA polymerase. These advantages of the CRT technology could represent significant improvements in speed, throughput, and accuracy over Sanger and SNA approaches.

At the center of the CRT chemistry is the reversible terminator. Ideally, these terminators should exhibit fast and efficient deprotection kinetics, efficient incorporation kinetics by DNA polymerase, and labels with desired characteristics, such as fluorophores with good fluorescence properties. Of the challenges associated with CRT for high-throughput genome sequencing, creating these reversible terminators with the desired properties and identifying DNA polymerases that recognize these substrates with high affinities are the most demanding aspects of the technology. The latter point is exemplified by the presence of competing natural nucleotides, which can readily cause asynchronistic base extensions (Metzker et al. 1998). The first examples of reversible terminators using commercially available DNA polymerases were reported by Canard and Sarfati (1994) and Metzker et al. (1994).



Cycle Efficiency (%)	Read-length (bases)
90.0	7
95.0	14
98.0	35
99.0	69
99.5	149
99.9	693

Figure 4. CRT technologies. (*A*) The CRT cycle. (*B*) The photocleavage reaction of a 3′-*O*-2-nitrobenzyl-nucleoside. (*C*) Effect of cycle efficiency on CRT read-length. (*D*) Kinetic study of protocleavage reaction for single substituted (2-SSNB) and double substituted (2-dsNB) 2-nitrobenzyl thymidine analogs. Percentage thymidine (%Thy) was calculated according to the equation: $\%Thy = A_{Thy}/(A_{Thy} + A_{s2NB})$, where A_{Thy} and A_{s2NB} are the integrated peak areas from RP-HPLC analysis for thymidine and substituted 2-nitrobenzyl thymidine analogs, respectively.

For CRT terminators to function properly, the protecting group must be efficiently cleaved under mild conditions while coupled to the primer. Removal of the protecting group generally involves either treatment with strong acid or base, catalytic or chemical reduction, or a combination of these methods. Unfortunately, these conditions may chemically perturb the DNA polymerase, nucleotides, oligonucleotide-primed template, or the solid support. Use of photocleavable protecting groups is an attractive alternative to rigorous chemical treatment and can be employed in a noninvasive manner. Of the various photocleavable protecting groups (Pillai 1980), the light-sensitive 2-nitrobenzyl group has been widely used. For example, it has been applied to natural nucleotides (Metzker et al. 1994, 1998), to the linker structure coupling a fluorescent dye to nucleobases (Li et al. 2003; Mitra et al. 2003), and to other nucleic acid structures as well (Ohtsuka et al. 1974; Pease et al. 1994; Chaulk and MacMillan 1998; Singh-Gasson et al. 1999). Under appropriate deprotection conditions (e.g., ultraviolet light >300 nm), the 2-nitrobenzyl group can be efficiently cleaved (Fig. 4B) without affecting either the pyrimidine or purine bases (Bartholomew and Broom 1975; Pease et al. 1994).

Other protecting groups have been described for reversible terminators as well. For example, Metzker et al. (1994) first described the synthesis and incorporation of a 3′-*O*-allyl-dATP by DNA polymerase, with the *O*-allyl group being removed using the well-known palladium (Pd) catalyst chemistry (Hayakawa et al. 1986, 1993; Honda et al. 1997). Recently, Ruparel et al. (2005) reported the synthesis of the first fluorescently labeled 3′-*O*-allyl-dNTPs. These unique reversible terminators require dual deprotection steps using UV light to cleave the fluorophore from the nucleotide (Fig. 3B), and the Pd catalyst reaction to restore the natural 3′-OH substrate. At this year's *Advances in Genome Biology and Technology/Automation in Mapping and Sequencing* meeting, Solexa reported on a similar CRT chemistry with a sequence read-length of approximately 20 bases (http://www.agbt.org) and recently reported the complete sequencing of the φχ174 genome (http://www.solexa.com).

Earlier concerns regarding short read-lengths and assemblies for SNA strategies will prove relevant to CRT as well. To overcome this issue, research efforts in CRT technology development will continue to focus on the cycle efficiency. The CRT read-length is governed by the overall cycle efficiency, which is highly dependent on the product of deprotection and incorporation efficiencies. For example, if one considers the conservative loss of 50% signal as the assay's end-point, the read-length is a function of the cycle efficiency (C_{eff}) (Fig. 4C). Here, a read-length of only seven bases will be achieved with an overall cycle efficiency of 90% and can be increased beyond 100 bases in length by improving cycle efficiency to >99%. Figure 4D illustrates the effect that chemical modifications of the 2-nitrobenzyl ring system have on deprotection efficiency and thymidine production (V.A. Litosh, W. Wu, B. Stupi, and M. Metzker, unpubl.). Thus, recent improvements in chemical engineering of reversible terminators are important developments for CRT as an emerging technology for DNA sequencing applications.

Conclusions

Recent developments in DNA polymerase-dependent strategies highlight the central role these methods play in determination of the overall success of the sequencing assay. Although the standards for current Sanger technology have set the mark for emerg-

QIAGEN Exhibit 1026.0007

Downloaded from genome.cshlp.org on October 7, 2020 - Published by Cold Spring Harbor Laboratory Press

ing SNA and CRT technologies, these measures have evolved over several decades and from numerous research laboratories. The integration of additional technologies will be key for development of robust DNA sequencing platforms, including instrumentation, microfluidics, robotics, automation, software control, data acquisition, and informatics.

Beyond the integrated instrumentation built around the chemistry, the method by which genomes are sequenced will be important. Most strategies described in this review will employ the random approach of whole-genome shotgun sequencing and assembly (Weber and Myers 1997), including resequencing efforts for human sequence variation studies. While the random approach has the advantage of simplicity, it will require a tremendous number of sequence reads (i.e., a minimum of 900 million, 100-base reads will be needed to achieve a 30× assembly for a mammalian-size genome) to produce comprehensive sequence data for comparative studies between genomes. A directed approach, which targets specific regions across the genome, can effectively reduce genome size and complexity and, therefore, the number of sequencing reads needed to produce these comprehensive data sets. One example of a directed strategy for human resequencing could be the application of the CRT method to 5′→3′ synthesized high-density oligonucleotide arrays (Albert et al. 2003) by relying on the reference sequence as anchor points along the genome. The careful selection of unique and functional priming sites would represent an oligonucleotide tiling path across the genome. Priming CRT reactions from these anchor points and sequencing to adjacent priming sites would provide contiguous coverage of the targeted regions of interest. CRT reads could then be aligned to the known positions along the reference genome in a straightforward manner. This approach could also be used for mapping sequence reads to related genomes for comparative genomics studies. Alignment of random reads could be performed using conventional assembly algorithms, guided by the reference sequence, to produce contiguous DNA sequence information.

Although in its infancy, the potential for these emerging sequencing strategies to deliver next-generation technologies looks promising. Improvements in speed, efficiency, throughput, and sensitivity will all contribute to a reduction in cost over the next several years. The timing of these strategies coincides with an increasing demand for resequencing capacity, which will provide valuable insight into the role of specific sequence variation with common disease. Integration of multidisciplinary technologies will translate into practical and affordable sequencing devices capable of whole-genome analyses. Application of genome sequence information to health benefits could revolutionize disease prevention measures, early disease interventions, and make the possibility of personalized therapies routine.

Acknowledgments

I am extremely grateful to Richard A. Gibbs, Donna M. Muzny, and Sherry Metzker for critical review of the manuscript; Steven A. Soper for technical discussion; and NHGRI for their support from grants R01 HG003573, R41 HG003072, R41 HG003265, and R21 HG002443.

References

Alaverdian, L., Alaverdian, S., Bilenko, O., Bogdanov, I., Filippova, E., Gavrilov, D., Gorbovitski, B., Gouzman, M., Gudkov, G., Domratchev, S., et al. 2002. A family of novel DNA sequencing instruments based on single-photon detection. *Electrophoresis* **23:** 2804–2817.

Albert, T.J., Norton, J., Ott, M., Richmond, T., Nuwaysir, K., Nuwaysir, E.F., Stengele, K.-P., and Green, R.D. 2003. Light-directed 5′→3′ synthesis of complex oligonucleotide microarrays. *Nucleic Acids Res.* **31:** e35.

Backhouse, C., Caamano, M., Oaks, F., Nordman, E., Carrillo, A., Johnson, B., and Bay, S. 2000. DNA sequencing in a monolithic microchannel device. *Electrophoresis* **21:** 150–156.

Bartholomew, D.G. and Broom, A.D. 1975. One-step chemical synthesis of ribonucleosides bearing a photolabile ether protecting group. *J. Chem. Soc. Chem. Commun.* **Issue 2:** 38.

Becker, H. and Gartner, C. 2000. Polymer microfabrication methods for microfluidic analytical applications. *Electrophoresis* **21:** 12–26.

Bonham, V.L., Warshauer-Baker, E., and Collins, F.S. 2005. Race and ethnicity in the genome era: The complexity of the constructs. *Am. Psychol.* **60:** 9–15.

Boone, T., Fan, Z., Hooper, H., Ricco, A., Tan, H., and Williams, S. 2002. Plastic advances microfluidic devices. *Anal. Chem.* **74:** 78A–86A.

Braslavsky, I., Hebert, B., Kartalov, E., and Quake, S.R. 2003. Sequence information can be obtained from single DNA molecules. *Proc. Natl. Acad. Sci.* **100:** 3960–3964.

Canard, B. and Sarfati, R. 1994. DNA polymerase fluorescent substrates with reversible 3′-tags. *Gene* **148:** 1–6.

Carrilho, E. 2000. DNA sequencing by capillary array electrophoresis and microfabricated array systems. *Electrophoresis* **21:** 55–65.

Carrilho, E., Ruiz-Martinez, M.C., Berka, J., Smirnov, I., Goetzinger, W., Miller, A.W., Brady, D., and Karger, B.L. 1996. Rapid DNA sequencing of more than 1000 bases per run by capillary electrophoresis using replaceable linear polyacrylamide solutions. *Anal. Chem.* **68:** 3305–3313.

Chaisson, M., Pevzner, P., and Tang, H. 2004. Fragment assembly with short reads. *Bioinformatics* **20:** 2067–2074.

Chaulk, S. and MacMillan, A. 1998. Caged RNA: Photo-control of a ribozyme reaction. *Nucleic Acids Res.* **26:** 3173–3178.

Collins, F. and Galas, D. 1993. A new five-year plan for the U.S. Human Genome Project. *Science* **262:** 43–46.

Collins, F.S., Guyer, M.S., and Chakravarti, A. 1997. Variations on a theme: Cataloging human DNA sequence variation. *Science* **278:** 1580–1581.

Collins, F.S., Patrinos, A., Jordon, E., Chakravarti, A., Gesteland, R., Walters, L., and Members of the DOE and NIH Planning Groups. 1998. New goals for the U.S. Human Genome Project: 1998–2003. *Science* **282:** 682–689.

Collins, F.S., Green, E.D., Guttmacher, A.E., and Guyer, M.S. 2003. A vision for the future of genomics research. *Nature* **422:** 835–847.

Crawford, D.C. and Nickerson, D.A. 2005. Definition and clinical importance of haplotypes. *Annu. Rev. Med.* **56:** 303–320.

Culbertson, C.T., Jacobson, S.C., and Ramsey, J.M. 1998. Dispersion sources for compact geometries on microchips. *Anal. Chem.* **70:** 3781–3789.

Daly, M.J., Rioux, J.D., Schaffner, S.F., Hudson, T.J., and Lander, E.S. 2001. High-resolution haplotype structure in the human genome. *Nat. Genet.* **29:** 229–237.

The ENCODE Project Consortium. 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306:** 636–640.

Entz, P., Toliat, M.R., Hampe, J., Valentonyte, R., Jenisch, S., Nürnberg, P., and Nagy, M. 2005. New strategies for efficient typing of HLA class-II loci DQB1 and DRB1 by using pyrosequencing. *Tissue Antigens* **65:** 67–80.

Ewing, B. and Green, P. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res.* **8:** 186–194.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res.* **8:** 175–185.

Foster, M.W. and Sharp, R.R. 2002. Race, ethnicity, and genomics: Social classifications as proxies of biological heterogeneity. *Genome Res.* **12:** 844–850.

Fox, M.S., Magasanik, B., Signer, E.R., Solomon, F., Gellert, M.F., Haber, J.E., Daniel, J., Koshland, E., and Muschel, L.H. 1990. The Genome Project: Pro and con. *Science* **247:** 270.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., et al. 2002. The structure of haplotype blocks in the human genome. *Science* **296:** 2225–2229.

Gharizadeha, B., Nordströma, T., Ahmadiana, A., Ronaghi, M., and Nyrén, P. 2002. Long-read pyrosequencing using pure 2′-deoxyadenosine-5′-*O*′-(1-thiotriphosphate) Sp-isomer. *Anal. Biochem.* **301:** 82–90.

Guttman, A. 2002a. Capillary electrophoresis using replaceable gels. U.S.

QIAGEN Exhibit 1026.0008

Downloaded from genome.cshlp.org on October 7, 2020 - Published by Cold Spring Harbor Laboratory Press

patent no. RE37,606.
———. 2002b. Capillary electrophoresis using replaceable gels. U.S. patent no. RE37,941.
Harrison, D.J., Manz, A., Fan, Z., Luedi, H., and Widmer, H.M. 1992. Capillary electrophoresis and sample injection systems integrated on a planar glass chip. *Anal. Chem.* **64:** 1926–1932.
Harrison, D.J., Fluri, K., Seiler, K., Fan, Z., Effenhauser, C.S., and Manz, A. 1993. Micromachining a miniaturized capillary electrophoresis-based chemical analysis system on a chip. *Science* **261:** 895–897.
Hayakawa, Y., Kato, H., Uchiyama, M., Kajino, H., and Noyori, R. 1986. Allyloxycarbonyl group: A versatile blocking group for nucleotide synthesis. *J. Org. Chem.* **51:** 2400–2402.
Hayakawa, Y., Hirose, M., and Noyori, R. 1993. *O*-Allyl protection of guanine and thymine residues in oligodeoxyribonucleotides. *J. Org. Chem.* **58:** 5551–5555.
Honda, M., Morita, H., and Nagakura, I. 1997. Deprotection of allyl groups with sulfinic acids and palladium catalyst. *J. Org. Chem.* **62:** 8932–8936.
Hyman, E.D. 1988. A new method of sequencing DNA. *Anal. Biochem.* **174:** 423–436.
The International HapMap Consortium. 2003. The International HapMap Project. *Nature* **426:** 789–796.
Jacobson, S.C., Hergenroder, R., Koutny, L.B., Warmack, R.J., and Ramsey, J.M. 1994. Effects of injection schemes and column geometry on the performance of microchip electrophoresis devices. *Anal. Chem.* **66:** 1107–1113.
Ju, J., Ruan, C., Fuller, C., Glazer, A., and Mathies, R. 1995. Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. *Proc. Natl. Acad. Sci.* **92:** 4347–4351.
Kan, C.-W., Fredlake, C.P., Doherty, E.A.S., and Barron, A.E. 2004. DNA sequencing and genotyping in miniaturized electrophoresis systems. *Electrophoresis* **25:** 3564–3588.
Kartalov, E.P. and Quake, S.R. 2004. Microfluidic device reads up to four consecutive base pairs in DNA sequencing-by-synthesis. *Nucleic Acids Res.* **32:** 2873–2879.
Kheterpal, I., Scherer, J., Clark, S., Radhakrishnan, A., Ju, J., Ginther, C., Sensabaugh, G.F., and Mathies, R.A. 1996. DNA sequencing using a four-color confocal fluorescence capillary array scanner. *Electrophoresis* **17:** 1852–1859.
Koshland, D.E. 1989. Sequences and consequences of the human genome. *Science* **246:** 189.
Koutny, L., Schmalzing, D., Salas-Solano, O., El-Difrawy, S., Adourian, A., Buonocore, S., Abbey, K., McEwan, P., Matsudaira, P., and Ehrlich, D. 2000. Eight hundred-base sequencing in a microfabricated electrophoretic device. *Anal. Chem.* **72:** 3388–3391.
Lander, E.S. 1996. The new genomics: Global views of biology. *Science* **274:** 536–539.
Langaee, T. and Ronaghi, M. 2005. Genetic variation analyses by pyrosequencing. *Mutat. Res.* **573:** 96–102.
Lassiter, S.J., Stryjewski, W., Benjamin, J., Legendre, L., Erdmann, R., Wahl, M., Wurm, J., Peterson, R., Middendorf, L., and Soper, S.A. 2000. Time-resolved fluorescence imaging of slab gels for lifetime base-calling in DNA sequencing applications. *Anal. Chem.* **72:** 5373–5382.
Leamon, J.H., Lee, W.L., Tartaro, K.R., Lanza, J.R., Sarkis, G.J., deWinter, A.D., Berka, J., Weiner, M., Rothberg, J.M., and Lohman, K.L. 2003. A massively parallel PicoTiterPlate™ based platform for discrete picoliter-scale polymerase chain reactions. *Electrophoresis* **24:** 3769–3777.
Lee, L., Spurgeon, S., Heiner, C., Benson, S., Rosenblum, B., Menchen, S., Graham, R., Constantinescu, A., Upadhya, K., and Cassel, J. 1997. New energy transfer dyes for DNA sequencing. *Nucleic Acids Res.* **25:** 2816–2822.
Levene, M.J., Korlach, J., Turner, S.W., Foquet, M., Craighead, H.G., and Webb, W.W. 2003. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299:** 682–686.
Lewis, E.K., Haaland, W.C., Nguyen, F., Heller, D.A., Allen, M.J., MacGregor, R.R., Berger, C.S., Willingham, B., Burns, L.A., Scott, G.B.I., et al. 2005. Color-blind fluorescence detection for four-color DNA sequencing. *Proc. Natl. Acad. Sci.* **102:** 5346–5351.
Li, Z., Bai, X., Ruparel, H., Kim, S., Turro, N.J., and Ju, J. 2003. A photocleavable fluorescent nucleotide for DNA sequencing and analysis. *Proc. Natl. Acad. Sci.* **100:** 414–419.
Lieberwirth, U., Arden-Jacob, J., Drexhage, K.H., Herten, D.P., Muller, R., Neumann, M., Schulz, A., Siebert, S., Sagner, G., Klingel, S., et al. 1998. Multiplex dye DNA sequencing in capillary gel electrophoresis by diode laser-based time-resolved fluorescence detection. *Anal. Chem.* **70:** 4771–4779.
Liu, S., Shi, Y., Ja, W., and Mathies, R.A. 1999. Optimization of high-speed DNA sequencing on microfabricated capillary electrophoresis channels. *Anal. Chem.* **71:** 566–573.
Liu, S., Ren, H., Gao, Q., Roach, D.J., Loder Jr., R.T., Armstrong, T.M., Mao, Q., Blaga, I., Barker, D.L., and Jovanovich, S.B. 2000. Automated parallel DNA sequencing on multiple channel microchips. *Proc. Natl. Acad. Sci.*. **97:** 5369–5374.
Liu, P.-Y., Zhang, Y.-Y., Lu, Y., Long, J.-R., Shen, H., Zhao, L.-J., Xu, F.-H., Xiao, P., Xiong, D.-H., Liu, Y.-J., et al. 2005. A survey of haplotype variants at several disease candidate genes: The importance of rare variants for complex diseases. *J. Med. Genet.* **42:** 221–227.
Luria, S.E., Cooper, D.M., and Berkowitz, A. 1989. Human Genome Project. *Science* **246:** 873–874.
Madabhushi, R.S. 1998. Separation of 4-color DNA sequencing extension products in noncovalently coated capillaries using low viscosity polymer solutions. *Electrophoresis* **19:** 224–230.
Madabhushi, R.S., Menchen, S.M., Efcavitch, J.W., and Grossman, P.D. 1996. Polymers for separation of biomolecules by capillary electrophoresis. U.S. patent no. 5,567,292.
———. 1999. Polymers for separation of biomolecules by capillary electrophoresis. U.S. patent no. 5,916,426.
Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:** 376–380.
McDonald, J.C., Duffy, D.C., Anderson, J.R., Chiu, D.T., Wu, H., Schueller, O.J.A., and Whitesides, G.M. 2000. Fabrication of microfluidic systems in poly(dimethylsiloxane). *Electrophoresis* **21:** 27–40.
Metzker, M.L., Raghavachari, R., Richards, S., Jacutin, S.E., Civitello, A., Burgess, K., and Gibbs, R.A. 1994. Termination of DNA synthesis by novel 3′-modified deoxyribonucleoside triphosphates. *Nucleic Acids Res.* **22:** 4259–4267.
Metzker, M.L., Lu, J., and Gibbs, R.A. 1996. Electrophoretically uniform fluorescent dyes for automated DNA sequencing. *Science* **271:** 1420–1422.
Metzker, M.L., Raghavachari, R., Burgess, K., and Gibbs, R.A. 1998. Elimination of residual natural nucleotides from 3′-*O*-modified-dNTP syntheses by enzymatic Mop-Up. *BioTechniques* **25:** 814–817.
Mitra, R. and Church, G. 1999. In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res.* **27:** e34.
Mitra, R.D., Shendure, J., Olejnik, J., Edyta-Krzymanska-Olejnik, and Church, G.M. 2003. Fluorescent in situ sequencing on polymerase colonies. *Anal. Biochem.* **320:** 55–65.
Nunnally, B.K., He, H., Li, L.-C., Tucker, S.A., and McGown, L.B. 1997. Characterization of visible dyes for four-decay fluorescence detection in DNA sequencing. *Anal. Chem.* **69:** 2392–2397.
Ohtsuka, E., Tanaka, S., and Ikehara, M. 1974. Studies on transfer ribonucleic acids and related compounds. IX(1) Ribooligonucleotide synthesis using a photosensitive *o*-nitrobenzyl protection at the 2′-hydroxyl group. *Nucleic Acids Res.* **1:** 1351–1357.
Paegel, B.M., Hutt, L.D., Simpson, P.C., and Mathies, R.A. 2000. Turn geometry for minimizing band broadening in microfabricated capillary electrophoresis channels. *Anal. Chem.* **70:** 3030–3037.
Paegel, B.M., Emrich, C.A., Wedemayer, G.J., Scherer, J.R., and Mathies, R.A. 2002. High throughput DNA sequencing with a microfabricated 96-lane capillary array electrophoresis bioprocessor. *Proc. Natl. Acad. Sci.* **99:** 574–579.
Paegel, B.M., Blazej, R.G., and Mathies, R.A. 2003. Microfluidic devices for DNA sequencing: Sample preparation and electrophoretic analysis. *Curr. Opin. Biotechnol.* **14:** 42–50.
Parkhill, J., Achtman, M., James, K.D., Bentley, S.D., Churcher, C., Klee, S.R., Morelli, G., Basham, D., Brown, D., Chillingworth, T., et al. 2000a. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404:** 502–506.
Parkhill, J., Wren, B.W., Mungall, K., Ketley, J.M., Churcher, C., Basham, D., Chillingworth, T., Davies, R.M., Feltwell, T., Holroyd, S., et al. 2000b. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403:** 665–668.
Patil, N., Berno, A.J., Hinds, D.A., Barrett, W.A., Doshi, J.M., Hacker, C.R., Kautzer, C.R., Lee, D.H., Marjoribanks, C., McDonough, D.P., et al. 2001. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294:** 1719–1723.
Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P., and Fodor, S.P.A. 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci.* **91:** 5022–5026.
Pillai, V.N.R. 1980. Photoremovable protecting groups in organic synthesis. *Synthesis* **Issue 2:** 1–26.
Prober, J., Trainor, G., Dam, R., Hobbs, F., Robertson, C., Zagursky, R.,

QIAGEN Exhibit 1026.0009

Downloaded from genome.cshlp.org on October 7, 2020 - Published by Cold Spring Harbor Laboratory Press

Cocuzza, A., Jensen, M., and Baumeister, K. 1987. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238:** 336–341.
Quake, S. and Scherer, A. 2000. From micro- to nanofabrication with soft materials. *Science* **290:** 1536–1540.
Risch, N. and Merikangas, K. 1996. The future of genetic studies of complex human diseases. *Science* **273:** 1516–1517.
Roberts, L. 1989a. New game plan for genome mapping. *Science* **245:** 1438–1440.
———. 1989b. Watson versus Japan. *Science* **246:** 576–578.
Robertson, J.A. 2003. The $1000 genome: Ethical and legal issues in whole genome sequencing of individuals. *Am. J. Bioeth.* **3:** W-IF1.
Ronaghi, M. 2000. Improved performance of pyrosequencing using single-stranded DNA-binding protein. *Anal. Biochem.* **286:** 282–288.
———. 2001. Pyrosequencing sheds light on DNA sequencing. *Genome Res.* **11:** 3–11.
Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., and Nyrén, P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* **242:** 84–89.
Ronaghi, M., Uhlén, M., and Nyrén, P. 1998. A sequencing method based on real-time pyrophosphate. *Science* **281:** 363, 365.
Ruiz-Martinez, M.C., Berka, J., Belenkii, A., Foret, F., Miller, A.W., and Karger, B.L. 1993. DNA sequencing by capillary electrophoresis with replaceable linear polyacrylamide and laser-induced fluorescence detection. *Anal. Chem.* **65:** 2851–2858.
Ruparel, H., Bi, L., Li, Z., Bai, X., Kim, D.H., Turro, N.J., and Ju, J. 2005. Design and synthesis of a 3′-*O*-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. *Proc. Natl. Acad. Sci.* **102:** 5932–5937.
Salas-Solano, O., Carrilho, E., Kotler, L., Miller, A.W., Goetzinger, W., Sosic, Z., and Karger, B.L. 1998. Routine DNA Sequencing of 1000 Bases in Less Than One Hour by Capillary Electrophoresis with Replaceable Linear Polyacrylamide Solutions. *Anal. Chem.* **70:** 3996–4003.
Salas-Solano, O., Schmalzing, D., Koutny, L., Buonocore, S., Adourian, A., Matsudaira, P., and Ehrlich, D. 2000. Optimization of high-performance DNA sequencing on short microfabricated electrophoretic devices. *Anal. Chem.* **72:** 3129–3137.
Sanger, F., Nicklen, S., and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74:** 5463–5467.
Schmalzing, D., Tsao, N., Koutny, L., Chisholm, D., Srivastava, A., Adourian, A., Linton, L., McEwan, P., Matsudaira, P., and Ehrlich. D. 1999. Toward real-world sequencing by microdevice electrophoresis. *Genome Res.* **9:** 853–858.
Seo, T.S., Bai, X., Ruparel, H., Li, Z., Turro, N.J., and Ju, J. 2004. Photocleavable fluorescent nucleotides for DNA sequencing on a chip constructed by site-specific coupling chemistry. *Proc. Natl. Acad. Sci.* **101:** 5488–5493.
Seo, T.S., Bai, X., Kim, D.H., Meng, Q., Shi, S., Ruparel, H., Li, Z., Turro, N.J., and Ju, J. 2005. Four-color DNA sequencing by synthesis on a chip using photocleavable fluorescent nucleotides. *Proc. Natl. Acad. Sci.* **102:** 5926–5931.
Shendure, J., Mitra, R.D., Varma, C., and Church, G.M. 2004. Advanced sequencing technologies: Methods and goals. *Nat. Rev. Genet.* **5:** 335–344.
Shi, Y. and Anderson, R.C. 2003. High-resolution single-stranded DNA analysis on 4.5 cm plastic electrophoretic microchannels. *Electrophoresis* **24:** 3371–3377.
Simpson, J.W., Ruiz-Martinez, M.C., Mulhern, G.T., Berka, J., Latimer, D.R., Ball, J.A., Rothberg, J.M., and Went, G.T. 2000. Transmission imaging spectrograph and microfabricated channel system for DNA analysis. *Electrophoresis* **21:** 135–149.
Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., and Cerrina, F. 1999. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat. Biotechnol.* **17:** 974–978.
Smith, L., Sanders, J., Kaiser, R., Hughes, P., Dodd, C., Connell, C., Heiner, C., Kent, S., and Hood, L. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* **321:** 674–679.
Smith, L.M., Kaiser, R.J., Sanders, J.Z., and Hood, L.E. 1987. The synthesis and use of fluorescent oligonucleotides in DNA sequence analysis. *Methods Enzymol.* **155:** 260–301.
Tabor, S. and Richardson, C.C. 1989. Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and *Escherichia coli* DNA polymerase I. *Proc. Natl. Acad. Sci.* **86:** 4076–4080.
———. 1995. A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl. Acad. Sci.* **92:** 6339–6343.
Takahashi, S., Murakami, K., Anazawa, T., and Kambara, H. 1994. Multiple sheath-flow gel capillary-array electrophoresis for multicolor fluorescent DNA detection. *Anal. Chem.* **66:** 1021–1026.
Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. 1995. Serial analysis of gene expression. *Science* **270:** 484–487.
Weber, J.L. and Myers, E.W. 1997. Human whole-genome shotgun sequencing. *Genome Res.* **7:** 401–409.
Woolley, A.T. and Mathies, R.A. 1995. Ultra-high-speed DNA sequencing using capillary electrophoresis chips. *Anal. Chem.* **67:** 3676–3680.
Zhang, C.-X. and Manz, A. 2001. Narrow sample channel injectors for capillary electrophoresis on microchips. *Anal. Chem.* **73:** 2656–2662.
Zhu, L., Stryjewski, W., Lassiter, S., and Soper, S.A. 2003. Fluorescence multiplexing with time-resolved and spectral discrimination using a near-IR detector. *Anal. Chem.* **75:** 2280–2291.
Zhu, L., Stryjewski, W.J., and Soper, S.A. 2004. Multiplexed fluorescence detection in microfabricated devices with both time-resolved and spectral-discrimination capabilities using near-infrared fluorescence. *Anal. Biochem.* **330:** 206–218.

Web site references

http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-002.html; RFA-HG-04-002. 2004. $100,000 genome RFA.
http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-003.html; RFA-HG-04-003. 2004. $1000 genome RFA.
http://www.agbt.org; Home page for the Advances in Genome Biology and Technology meeting.
http://www.solexa.com; Home page for Solexa, Inc.

QIAGEN Exhibit 1026.0010

Downloaded from genome.cshlp.org on October 7, 2020 - Published by Cold Spring Harbor Laboratory Press

GENOME RESEARCH

Emerging technologies in DNA sequencing

Michael L. Metzker

Genome Res. 2005 15: 1767-1776
Access the most recent version at doi:10.1101/gr.3770505

References This article cites 103 articles, 38 of which can be accessed free at:
http://genome.cshlp.org/content/15/12/1767.full.html#ref-list-1

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

ThruPLEX® HV
failproof DNA-seq of FFPE & cfDNA
TaKaRa
Clontech Takara cellartis

To subscribe to *Genome Research* go to:
http://genome.cshlp.org/subscriptions

Cold Spring Harbor Laboratory Press

QIAGEN Exhibit 1026.0011