Biochemistry 412

New Strategies, Technologies, & Applications For DNA Sequencing

27 January 2009
In 1980, the sequencing cost per finished bp \( \approx \$1.00 \)
In 2003, the sequencing cost per finished bp \( \approx \$0.01 \)

>>> a 100-fold reduction in 20-25 years

Figure 1 | **Exponential growth in computing and sequencing.** The dark-blue plot indicates the Kurzweil/Moore’s Law\(^{108}\): it describes the doubling of computer instructions per second per US dollar (IPS/US $) that has been occurring approximately every 18 months since 1900. The magenta plot indicates an exponential growth in the number of base pairs of accurate DNA sequence per unit cost (bp/US $) as a function of time\(^1\). To some extent, the doubling time for DNA mimics the IPS/US $ curve because it is dependent on it. An even steeper segment occurs in the orange curve; this depicts the number of web sites (doubling time of four months)\(^{109}\) and shows how quickly a technology can explode when a protocol that can be shared spreads through an existing infrastructure. The turquoise plot is an ‘Open Source’ case study of ‘FLUORESCENT IN SITU SEQUENCING’ with pololones\(^4\) (see main text for details of this DNA-sequencing technology) in bp/min on simple test templates (doubling time of one month).

Today, the cost of sequencing each additional human genome ("resequencing") is estimated to be about $1 million.

Q1: Can that cost be dropped by ≥ two orders of magnitude (i.e, cost ≤ $10K per genome)?*

Q2: If the answer to Q1 is “yes”, what new technologies will be needed to achieve this?

[*Note: some even talk about reaching the goal of the “$1000 genome”, which may or may not be achievable depending on what degree of completeness and usefulness of “formatting” you are talking about.]
Even more ambitious:

A $10 million prize for the first group that demonstrates the sequencing of 100 human genomes in 10 days or less!

Note: this was modeled on the Ansari X Prize for the first privately-financed human spaceflight, which was recently won by SpaceShipOne (at right -- built by Mojave Aerospace Ventures).
Just as “space tourism” is an industry that was spawned by rocketry research and the space program, personal genome sequences -- made possible by knowledge bases and technologies developed for the human genome project -- are now available to wealthy people and celebrities.
The Diploid Genome Sequence of an Individual Human

Samuel Levy1+, Granger Sutton1, Pauline C. Ng1, Lars Feuk2, Aaron L. Halpern1, Brian P. Walenz1, Nelson Axelrod1, Jiaqi Huang1, Ewen F. Kirkness1, Gennady Denisov1, Yuan Lin1, Jeffrey R. MacDonald2, Andy Wing Chun Pang2, Mary Shago2, Timothy B. Stockwell1, Alexia Tsiamouri1, Vineet Bafna3, Vikas Bansal3, Saul A. Kravitz1, Dana A. Busam1, Karen Y. Beeson1, Tina C. McIntosh1, Karin A. Remington1, Josep F. Abri1d, John Gill1, Jon Borman1, Yu-Hui Rogers1, Marvin E. Frazier1, Stephen W. Scherer2, Robert L. Strausberg1, J. Craig Venter1

1 J. Craig Venter Institute, Rockville, Maryland, United States of America, 2 Program in Genetics and Genomic Biology, The Hospital for Sick Children, and Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada, 3 Department of Computer Science and Engineering, University of California San Diego, La Jolla, California, United States of America, 4 Genetics Department, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain

Presented here is a genome sequence of an individual human. It was produced from ~32 million random DNA fragments, sequenced by Sanger dideoxy technology and assembled into 4,528 scaffolds, comprising 2,810 million bases (Mb) of contiguous sequence with approximately 7.5-fold coverage for any given region. We developed a modified version of the Celera assembler to facilitate the identification and comparison of alternate alleles within this individual diploid genome. Comparison of this genome and the National Center for Biotechnology Information human reference assembly revealed more than 4.1 million DNA variants, encompassing 12.3 Mb. These variants (of which 1,288,319 were novel) included 3,213,401 single nucleotide polymorphisms (SNPs), 53,823 block substitutions (2–206 bp), 292,102 heterozygous insertion/deletion events (indels) (1–571 bp), 559,473 homozygous indels (1–82,711 bp), 90 inversions, as well as numerous segmental duplications and copy number variation regions. Non-SNP DNA variation accounts for 22% of all events identified in the donor, however they involve 74% of all variant bases. This suggests an important role for non-SNP genetic alterations in defining the diploid genome structure. Moreover, 44% of genes were heterozygous for one or more variants. Using a novel haplotype assembly strategy, we were able to span 1.5 Gb of genome sequence in segments >200 kb, providing further precision to the diploid nature of the genome. These data depict a definitive molecular portrait of a diploid human genome that provides a starting point for future genome comparisons and enables an era of individualized genomic information.
J. Craig Venter (above) led the effort, and donated his DNA, to sequence and assemble the complete genome of a single individual.

Browse Craig Venter’s genome (!) at:
http://journals.plos.org/plosbiology/suppinfo/pbio.0050254/pbio.0050254.sd001.htm
know thyself.

The first personal genomics company to offer complete genome sequencing and analysis services for private individuals.

Our approach
Complete genome sequencing
Frequently asked questions
Recent news

About Knome

Based in Cambridge, Massachusetts, Knome works alongside leading geneticists, clinicians and bioinformaticians from Harvard and MIT to enable private individuals to obtain, understand, and share their genomic information in a manner that is both anonymous and secure.

We partner with our clients to help them understand what their genome can tell them about themselves. By being amongst the first individuals in history to have their complete genome sequenced, these individuals are helping pioneer the emerging field of personal genomics.

E-mail: info@knome.com

Knome in the media

The Book of Me
GQ
October 14, 2008

The Genetic Early Adopters
Technology Review
September 8, 2008

Mapping Out a Nascent Market
The Boston Globe
August 10, 2008

Gene Map Becomes a Luxury Item
The New York Times
March 4, 2008

View all news stories »

Learn More

KnomeExplorer™
A window into your genes.

GenomeKey™
Unlock your genome.

GeneReviews™
The most comprehensive personal genomics analysis.

Knome does not provide medical advice, diagnosis or treatment. Analyses and reports provided by Knome are for informational purposes only and are subject to change. You should consult your physician if you have questions regarding any medical condition, before starting any new treatment, and before stopping any treatment that has been prescribed for you.
Several companies are pursuing *massively parallel* (= cheaper) new DNA sequencing strategies, including some that involve *single molecule* analyses.

Some of players with commercially available systems are given below:

454 Life Sciences (now part of Roche)  
(http://www.454.com)

Solexa (now part of Illumina)  
(http://www.illumina.com/pages.ilmn?ID=203)

Applied Biosystems (SOLID system)  
(http://www.appliedbiosystems.com/?abhomepage=na)

Helicos BioSciences  
(http://www.helicosbio.com)
A quick overview....

(a) Sanger sequencing
(b) Sequencing by hybridization
(c) Pyrosequencing
(d) Sequencing by synthesis - base-by-base*

(e) Sequencing by synthesis - ligation*
(f) Nanopore technology*
(g) Anchored polymerase read-out*

*Note: legend incorrectly labeled in paper!

“Classic” Sanger Sequencing

1 Before Sanger-style sequencing, an original DNA strand is broken into smaller fragments and cloned within colonies of Escherichia coli bacteria. Once extracted from the bacteria, the DNA fragments will undergo another massive round of copying, known as amplification, by a process called polymerase chain reaction (PCR).

2 During PCR, fragments are heated so they will separate into single strands. A short nucleotide sequence called a primer is then annealed to each original template. Starting at the primer, polymerase links free-floating nucleotides [called dNTPs] into new complementary strands. The process is repeated over and over to generate millions of copies of each fragment.

3 Single-stranded fragments are next tagged in a process similar to PCR but with fluorescently labeled terminator nucleotides (ddNTPs) added to the mixture of primers, polymerase and dNTPs. Complementary strands are built until by chance a ddNTP is incorporated, halting synthesis. The resulting copy fragments have varying lengths and a tagged nucleotide at one end.

4 Capillary electrophoresis separates the fragments, which are negatively charged, by drawing them toward a positively charged pole. Because the shortest fragments move fastest, their order reflects their size and their ddNTP terminators can thus be "read" as the template’s base sequence. Laser light activates the fluorescent tags as the fragments pass a detection window, producing a color/readout that is translated into a sequence.

For a primer on how large scale genomic sequencing is currently done, please see the educational link below at the U. S. Department of Energy Joint Genome Institute.

http://www.jgi.doe.gov/education/how/
Box 2 | Applications of ultra-low-cost sequencing: a partial list

- Sequencing of individual human genomes as a component of preventative medicine.
- Rapid hypothesis testing for genotype–phenotype associations\textsuperscript{14,17,18}.
- \textit{In vitro} and \textit{in situ} gene-expression profiling at all stages in the development of a multicellular organism\textsuperscript{88,89}.
- Cancer research: for example, determining comprehensive mutation sets for individual clones\textsuperscript{90}, carrying out loss-of-heterozygosity analysis\textsuperscript{91} and profiling tumour sub-types for diagnosis and prognosis\textsuperscript{92,93}.
- Temporal profiling of B- and T-cell receptor diversity, both clinically and for antibody selection in the laboratory.
- Identification of known and new pathogens\textsuperscript{94}; development of biowarfare sensors\textsuperscript{95}.
- Detailed annotation of the human genome through \textsc{phylogenetic footprinting and shadowing}\textsuperscript{96}.
- Quantification of alternative splice variants in the transcriptomes of higher eukaryotes\textsuperscript{56,97}.
- Definition of epigenetic structures (such as chromatin modifications and methylation patterns)\textsuperscript{98}.
- \textit{In situ or ex vivo} discovery of cell-lineage patterns\textsuperscript{99,100}.
- Characterization of microbial strains that have been subjected to extensive \textsc{directed evolution}\textsuperscript{101,102}.
- Exploration of microbial diversity towards agricultural, environmental and therapeutic goals\textsuperscript{13,103}.
- Annotation of microbial genomes through the selectional analysis of tagged insertional mutants\textsuperscript{104,105}.
- Use of DNA or RNA oligonucleotides as agents to bind specific protein targets with high affinity and specificity (so-called ‘aptamer technology’) for diagnostics and therapeutics\textsuperscript{106}.
- DNA computing\textsuperscript{23,24} — that is, manipulating DNA libraries to carry out highly parallel computations. Potential solutions to the problem are often encoded in nucleotide sequence, and standard experimental manipulations (such as hybridization) are used to search the space of possible solutions.

Note: “Re-sequencing”, doing the n\textsuperscript{th} genome, or part of a genome, from an organism where at least one example of the complete genome is already known in the same (or a related) species, is almost always easier and cheaper than doing the \textit{de novo} sequencing to get the first example of a genome in that species.

\textit{Do you know why??}
A sample of some of the approaches that are being tried...

Figure 2 | Examples of microelectrophoretic sequencing and nanopore sequencing. a | Microelectrophoretic sequencing. Left: a microfabricated wafer for 384-well capillary electrophoretic sequencing. Reactions are injected at the perimeter and run towards the centre, where a rotary confocal fluorescence scanner carries out the detection. Reproduced with permission from REF 27 © (2000) American Chemical Society. Right: microelectrophoretic sequencing produces raw sequencing traces that are similar to those generated by electrophoretic sequencing. b | Nanopore sequencing. Left: single-stranded polynucleotides can only pass single-file through a hemolysin nanopore. Right: the presence of the polynucleotide in the nanopore is detected as a transient blockade of the baseline ionic current. pA, pico-Ampere.

Some more technological approaches…
(some of which really work!)

• Sequencing by hybridization (annealing)
• Sequencing by “ligase-edited” annealing
• Pyrosequencing
• Tethered polymerase in waveguide wells

Note: there are also higher tech versions of “classic” Sanger sequencing (see http://www.illumina.com/pages.ilmn?ID=203 and http://www.helicosbio.com)
Note re nomenclature: “hybridization” originally referred to duplexes between complementary strands of RNA and DNA, i.e., hybrid duplexes. But in the conventional sense that the term is now used, “hybridization” also often refers to the formation of DNA-DNA duplexes, particular when one half of the duplex is a so-called “probe” molecule and the other is the target. A more general term describing in vitro formation of nucleic acid duplexes stabilized by complementary base-pairing -- whether DNA-DNA, RNA-DNA, or RNA-RNA -- is “annealing” (but you will almost never hear this term used instead of “hybridization”).
For sequencing by hybridization, you typically need to have a fluorescently-labeled copy of your target DNA.

Sequencing by hybridization (Affymetrix chip)

**Figure 1**  Eight features (four for the forward strand and four for the reverse complement strand) are associated with every queried site. Each feature consists of a 25-base oligonucleotide. The 13th base is the query base and all possible genotypes are tested. Each feature is divided into 56 equal pixels, and the pixels are scanned individually. The outermost 26 pixels are “masked,” so that only the 30 interior pixels are used for any calculation.
Sequencing by synthesis

Base-by-base elongation using chain-terminating nucleotide analogs with cleavable 3’-OH blocking groups and fluorescent moieties. This is the primary approach used by Illumina (Solexa).

Note: the nucleotide analogs shown at left are from Ju et al [Proc. Natl. Acad. Sci. USA 103, 19636 (2006)] and are not the same compounds used by Illumina in their system.

Fig. 1. Structures of 3’-O-allyl-dCTP-allyl-Bodipy-FL-510 ($\lambda_{\text{abs(max)}} = 502 \text{ nm}; \lambda_{\text{em(max)}} = 510 \text{ nm}$), 3’-O-allyl-dUTP-allyl-R6G ($\lambda_{\text{abs(max)}} = 525 \text{ nm}; \lambda_{\text{em(max)}} = 550 \text{ nm}$), 3’-O-allyl-dATP-allyl-ROX ($\lambda_{\text{abs(max)}} = 585 \text{ nm}; \lambda_{\text{em(max)}} = 602 \text{ nm}$), and 3’-O-allyl-dGTP-allyl-bodipy-650 ($\lambda_{\text{abs(max)}} = 630 \text{ nm}; \lambda_{\text{em(max)}} = 650 \text{ nm}$).
Fig. 2. The polymerase extension scheme (Left) and MALDI-TOF MS spectra of the four consecutive extension products and their deallylated products (Right). Primer extended with 3'-O-allyl-dUTP-allyl-RGG (1), and its deallylated product 2; Product 2 extended with 3'-O-allyl-dGTP-allyl-bodipy-650 (3), and its deallylated product 4; Product 4 extended with 3'-O-allyl-dATP-allyl-ROX (5), and its deallylated product 6; Product 6 extended with 3'-O-allyl-dCTP-allyl-bodipy-FL-510 (7), and its deallylated product (8). After 30 s of incubation with the palladium/TPPTS mixture at 70°C, deallylation is complete with both the fluorophores and the 3'-O-allyl groups cleaved from the extended DNA products.

Figure 1  The general principle behind different pyrosequencing reaction systems. A polymerase catalyzes incorporation of nucleotid(e(s) into a nucleic acid chain. As a result of the incorporation, a pyrophosphate (PPI) molecule(s) is released and subsequently converted to ATP, by ATP sulfurylase. Light is produced in the luciferase reaction during which a luciferin molecule is oxidized.
Four enzyme pyrosequencing

Use of an apyrase wash to destroy residual nucleotides

**Figure 3** Schematic representation of the progress of the enzyme reaction in liquid-phase pyrosequencing. Primed DNA template and four enzymes involved in liquid-phase pyrosequencing are placed in a well of a microtiter plate. The four different nucleotides are added stepwise and incorporation is followed using the enzyme ATP sulfurylase and luciferase. The nucleotides are continuously degraded by nucleotide-degrading enzyme allowing addition of subsequent nucleotide. dXTP indicates one of the four nucleotides.

Note: in the 454 variant of this technology, an apyrase wash separates each addition of a dNTP.

Figure 4  Pyrogram of the raw data obtained from liquid-phase pyrosequencing. Proportional signals are obtained for one, two, three, and four base incorporations. Nucleotide addition, according to the order of nucleotides, is indicated below the pyrogram and the obtained sequence is indicated above the pyrogram.
Other ways to reduce costs:

• do things in parallel
  (rather than in series)

• reduce volumes
  (= reduced reagent costs)

A cool idea: *micro-emulsion PCR*
Microemulsions as tiny discrete reaction vessels for producing clonal populations of PCR-amplified DNA from single molecules

Fig. 1. Schematic of BEAMing. Step 1: Magnetic beads covalently coated with streptavidin are bound to biotinylated oligonucleotides (oligos). Step 2: An aqueous mix containing all the necessary components for PCR plus primer-bound beads and template DNA are stirred together with an oil/detergent mix to create microemulsions. The aqueous compartments (white circles in the gray oil layer) contain an average of less than one template molecule and less than one bead. Red and green templates represent two template molecules, the sequences of which differ by one or many nucleotides. Step 3: The microemulsions are temperature-cycled as in a conventional PCR. If a DNA template and a bead are present together in a single aqueous compartment, the bead-bound oligonucleotides act as primers for amplification. The straight red and green lines connected to the beads represent extension products from the two different kinds of templates. Step 4: The emulsions are broken, and the beads are purified with a magnet. Step 5: After denaturation, the beads are incubated with oligonucleotides that can distinguish between the sequences of the different kinds of templates. Fluorescently labeled antibodies then are used to label the bound hybridization probes, which renders the beads containing PCR product as red or green after appropriate laser excitation. Step 6: Flow cytometry is used to count the red and green beads.

Fig. 2. Photograph of a typical microemulsion. Microemulsions were made as described in *Materials and Methods* with the exception that the aqueous compartments contained cascade blue-labeled dCTP and the beads were prelabeled by binding to oligonucleotides coupled to R-phycoerythrin (red) or Alexa 488 (green). One microliter of microemulsion was deposited in 1 μl of oil on a microscope slide before photography. Of the seven aqueous compartments visible in this picture, two contain beads. Note the heterogeneous size of the aqueous compartments (beads are 1.05 μm in diameter).

Historical note: Bert Vogelstein and his colleagues at Johns Hopkins Medical School developed this technique in order to detect mutations in tumor suppressor genes (in cells shed from precancerous polyps) against a large background of unmutated genes (from shedding of normal cells) in urine samples, sputum samples, stool samples, etc.

*It has been estimated that half of all precancerous growths might be detected at an early enough stage by this approach to enable their surgical removal before they progressed to become malignant.*
Genome sequencing in microfabricated high-density picolitre reactors

Marcel Margulies\textsuperscript{1,4}, Michael Egholm\textsuperscript{1,4}, William E. Altman\textsuperscript{1}, Said Attiya\textsuperscript{1}, Joel S. Bader\textsuperscript{1}, Lisa A. Bemben\textsuperscript{1}, Jan Berka\textsuperscript{1}, Michael S. Braverman\textsuperscript{1}, Yi-Ju Chen\textsuperscript{1}, Zhoutao Chen\textsuperscript{1}, Scott B. Dewell\textsuperscript{1}, Lei Du\textsuperscript{1}, Joseph M. Fierro\textsuperscript{1}, Xavier V. Gomes\textsuperscript{1}, Brian C. Godwin\textsuperscript{1}, Wen He\textsuperscript{1}, Scott Helgesen\textsuperscript{1}, Chun He Ho\textsuperscript{1}, Gerard P. Irzyk\textsuperscript{1}, Szilveszter C. Jando\textsuperscript{1}, Maria L. I. Alenquer\textsuperscript{1}, Thomas P. Jarvie\textsuperscript{1}, Kshama B. Jirage\textsuperscript{1}, Jong-Bum Kim\textsuperscript{1}, James R. Knight\textsuperscript{1}, Janna R. Lanza\textsuperscript{1}, John H. Leamon\textsuperscript{1}, Steven M. Lefkowitz\textsuperscript{1}, Ming Lei\textsuperscript{1}, Jing Li\textsuperscript{1}, Kenton L. Lohman\textsuperscript{1}, Hong Lu\textsuperscript{1}, Vinod B. Makhijani\textsuperscript{1}, Keith E. McDade\textsuperscript{1}, Michael P. McKenna\textsuperscript{1}, Eugene W. Myers\textsuperscript{2}, Elizabeth Nickerson\textsuperscript{1}, John R. Nobile\textsuperscript{1}, Ramona Plant\textsuperscript{1}, Bernard P. Puc\textsuperscript{1}, Michael T. Ronan\textsuperscript{1}, George T. Roth\textsuperscript{1}, Gary J. Sarkis\textsuperscript{1}, Jan Fredrik Simons\textsuperscript{1}, John W. Simpson\textsuperscript{1}, Maithreyan Srinivasan\textsuperscript{1}, Karrie R. Tartaro\textsuperscript{1}, Alexander Tomasz\textsuperscript{3}, Kari A. Vogt\textsuperscript{1}, Greg A. Volkmer\textsuperscript{1}, Shally H. Wang\textsuperscript{1}, Yong Wang\textsuperscript{1}, Michael P. Weiner\textsuperscript{4}, Pengguang Yu\textsuperscript{1}, Richard F. Begley\textsuperscript{1} & Jonathan M. Rothberg\textsuperscript{1}

The proliferation of large-scale DNA-sequencing projects in recent years has driven a search for alternative methods to reduce time and cost. Here we describe a scalable, highly parallel sequencing system with raw throughput significantly greater than that of state-of-the-art capillary electrophoresis instruments. The apparatus uses a novel fibre-optic slide of individual wells and is able to sequence 25 million bases, at 99\% or better accuracy, in one four-hour run. To achieve an approximately 100-fold increase in throughput over current Sanger sequencing technology, we have developed an emulsion method for DNA amplification and an instrument for sequencing by synthesis using a pyrosequencing protocol optimized for solid support and picolitre-scale volumes. Here we show the utility, throughput, accuracy and robustness of this system by shotgun sequencing and \textit{de novo} assembly of the \textit{Mycoplasma genitalium} genome with 96\% coverage at 99.96\% accuracy in one run of the machine.
Cost savings by using massively parallel pico-liter scale sequencing (454 Life Sciences Corp.)

Figure 1 | Speeding up sequencing. Flow diagrams for a, traditional microlitre-scale Sanger DNA sequencing and electrophoresis, and b, the massively parallel picolitre-scale sequencing developed by Rothberg et al. The traditional microlitre-scale approach requires a longer processing time per production cycle, substantially more support equipment, a larger facility and more labour than the picolitre-scale approach.
Design of 454’s sequencing scheme

Figure 1 | Sample preparation. a. Genomic DNA is isolated, fragmented, ligated to adapters and separated into single strands (top left). Fragments are bound to beads under conditions that favour one fragment per bead, the beads are captured in the droplets of a PCR-reaction-mixture-in-oil emulsion and PCR amplification occurs within each droplet, resulting in beads each carrying ten million copies of a unique DNA template (top right). The emulsion is broken, the DNA strands are denatured, and beads carrying single-stranded DNA clones are deposited into wells of a fibre-optic slide (bottom right). Smaller beads carrying immobilized enzymes required for pyrophosphate sequencing are deposited into each well (bottom left).

b. Microscope photograph of emulsion showing droplets containing a bead and empty droplets. The thin arrow points to a 28-μm bead; the thick arrow points to an approximately 100-μm droplet. c. Scanning electron micrograph of a portion of a fibre-optic slide, showing fibre-optic cladding and wells before bead deposition.

454’s Sequencing Instrument

**Figure 2 | Sequencing instrument.** The sequencing instrument consists of the following major subsystems: a fluidic assembly (a), a flow chamber that includes the well-containing fibre-optic slide (b), a CCD camera-based imaging assembly (c), and a computer that provides the necessary user interface and instrument control.

Figure 3 | Flowgram of a 113-bases read from an *M. genitalium* run. Nucleotides are flowed in the order T, A, C, G. The sequence is shown above the flowgram. The signal value intervals corresponding to the various homopolymers are indicated on the right. The first four bases (in red, above the flowgram) constitute the ‘key’ sequence, used to identify wells containing a DNA-carrying bead.
Resequencing of a small bacterial genome (size = 580 kbp) in a single instrument run

Table 2 | Summary statistics for M. genitalium

<table>
<thead>
<tr>
<th>Sequencing summary</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of instrument runs</td>
<td>1</td>
</tr>
<tr>
<td>Size of fibre-optic slide</td>
<td>60 × 60 mm²</td>
</tr>
<tr>
<td>Run time/number of cycles</td>
<td>243 min/42</td>
</tr>
<tr>
<td>High quality reads</td>
<td>306,178</td>
</tr>
<tr>
<td>Average read length (bases)</td>
<td>110</td>
</tr>
<tr>
<td>Number of bases in high quality reads</td>
<td>33,655,553</td>
</tr>
<tr>
<td>Bases with a Phred score of 20 and above</td>
<td>26,753,540</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Re-sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reads mapped to single locations</td>
</tr>
<tr>
<td>Number of bases in mapped reads</td>
</tr>
<tr>
<td>Individual read insertion error rate</td>
</tr>
<tr>
<td>Individual read deletion error rate</td>
</tr>
<tr>
<td>Individual read substitution error rate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Re-sequencing consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average over sampling</td>
</tr>
<tr>
<td>Coverage, all (Z ≥ 4)</td>
</tr>
<tr>
<td>Consensus accuracy, all (Z ≥ 4)</td>
</tr>
<tr>
<td>Consensus insertion error rate, all (Z ≥ 4)</td>
</tr>
<tr>
<td>Consensus deletion error rate, all (Z ≥ 4)</td>
</tr>
<tr>
<td>Consensus substitution error rate, all (Z ≥ 4)</td>
</tr>
<tr>
<td>Number of contigs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>De novo assembly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage, all (Z ≥ 4)</td>
</tr>
<tr>
<td>Consensus accuracy, all (Z ≥ 4)</td>
</tr>
<tr>
<td>Number of contigs</td>
</tr>
<tr>
<td>Average contig size (kb)</td>
</tr>
</tbody>
</table>

The individual read error rates are referenced to the total number of bases in mapped reads.

Note: conventional Sanger sequencing is still the “gold standard” for de novo sequencing of a never-before-sequenced genome. These new, rapid methods of sequencing will probably be most useful for cheaply re-sequencing genomes of different individuals (or strains) of the same species.
Is very, very fast and cheap DNA sequencing on the horizon??
Real-Time DNA Sequencing from Single Polymerase Molecules

John Eid,* Adrian Fehr,* Jeremy Gray,* Khai Luong,* John Lyle,* Geoff Otto,* Paul Peluso,* David Rank,* Primo Baybayan, Brad Bettman, Arkadiusz Bibillo, Keith Bjornson, Bidhan Chaudhuri, Frederick Christians, Ronald Cicero, Sonya Clark, Ravindra Dalal, Alex deWinter, John Dixon, Mathieu Foquet, Alfred Gaertner, Paul Hardenbol, Cheryl Heiner, Kevin Hester, David Holden, Gregory Kearns, Xiangyu Kong, Ronald Kuse, Yves Lacroix, Steven Lin, Paul Lundquist, Congcong Ma, Patrick Marks, Mark Maxham, Devon Murphy, Insil Park, Thang Pham, Michael Phillips, Joy Roy, Robert Sebra, Gene Shen, Jon Sorenson, Austin Tomaney, Kevin Travers, Mark Trulson, John Vieceli, Jeffrey Wegener, Dawn Wu, Alicia Yang, Denis Zaccarin, Peter Zhao, Frank Zhong, Jonas Korlach,† Stephen Turner†

Pacific Biosciences, 1505 Adams Drive, Menlo Park, CA 94025
Pacific Biosciences technological approach

Fig. 1. Principle of single-molecule, real-time DNA sequencing. (A) Experimental geometry. A single molecule of DNA template-bound Φ29 DNA polymerase is immobilized at the bottom of a ZMW, which is illuminated from below by laser light. The ZMW nanostructure provides excitation confinement in the zeptoliter (10⁻²¹ liter) regime, enabling detection of individual phospholinked nucleotide substrates against the bulk solution background as they are incorporated into the DNA strand by the polymerase. (B) Schematic event sequence of the phospholinked dNTP incorporation cycle, with a corresponding expected time trace of detected fluorescence intensity from the ZMW. (1) A phospholinked nucleotide forms a cognate association with the template in the polymerase active site, (2) causing an elevation of the fluorescence output on the corresponding color channel. (3) Phosphodiester bond formation liberates the dye-linker-pyrophosphate product, which diffuses out of the ZMW, thus ending the fluorescence pulse. (4) The polymerase translocates to the next position, and (5) the next cognate nucleotide binds the active site beginning the subsequent pulse.

Actual 4-color sequencing with the PacBio system

Is very, very fast and cheap DNA sequencing on the horizon??

See Pacific Biosciences