Aquaculture Genome Technologies
Zhanjiang (John) Liu
Copyright © 2007 by Blackwell Publishing

Chapter 25
DNA Sequencing Technologies

Zhanjiang Liu

An understanding of the organization, expression and function, and evolutionary history of the aquaculture genomes requires knowing their primary structure—the linear order of the nucleotide base pairs of the genomes. Currently, demand for low-cost sequencing far outstrips what existing sequencing technology can provide. New sequencing technologies must be developed to dissect genomes of species with relatively small research communities.

DNA sequencing technology fundamentally reshaped the field of biology. It has been one of the largest driving forces in the genomics revolution. In the last 30 years, sequencing capacity has dramatically increased, while the cost of sequencing has been drastically reduced (Figure 25.1). However, the principles behind the various sequencing platforms have remained the same, relying on Sanger's dideoxy chain termination sequencing technology.

Back in the 1970s, two methods were independently developed for DNA sequencing. One method was developed by an American team, and the other, by an English team. The Americans, led by Maxam and Gilbert, used a "chemical cleavage protocol," while the English, led by Sanger, used an enzymatic protocol involving the use of nucleotide analogues for incorporation into DNA, which terminates the growing

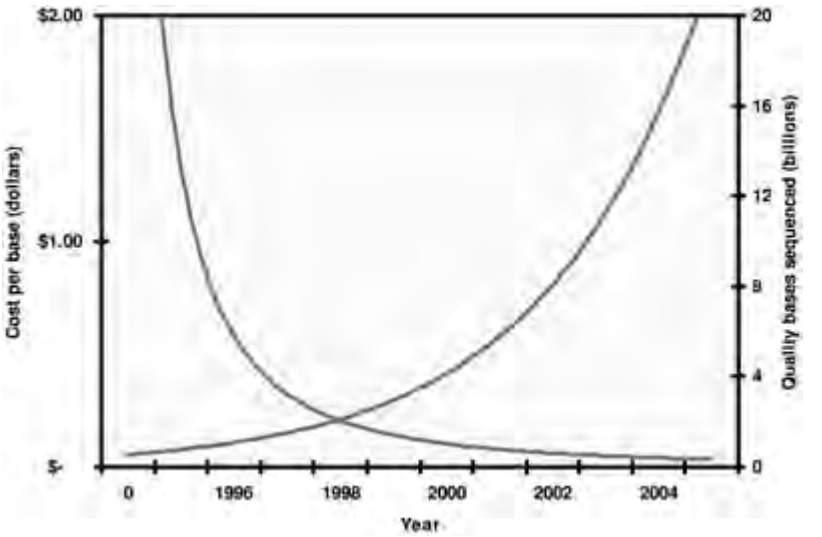

Figure 25.1. The trend of DNA sequencing capacity and cost. The figure is adopted from the Human Genome Project Information Web site (http://www.ornl.gov/sci/techresources/Human_Genome/research/instrumentation.shtml).

chain of DNA synthesis. Even though both teams shared the 1980 Nobel Prize, Sanger's method became the standard because of its practicality, while Maxam-Gilbert's chemical method has become obsolete and is no longer used for large-scale DNA sequencing.

In addition to these two major sequencing strategies, several emerging sequencing technologies are showing great promise. In particular, the 454 sequencing and the Solexa sequencing platforms, both based on the principles of sequencing by synthesis, are gaining momentum currently, and may meet increasing sequencing demand while reducing the cost. In this chapter, we will review the principles of sequencing technologies that brought us where we are today, and also study the potential of new sequencing technologies that may meet the demands of future genome sequencing.

Sanger's dideoxynucleotide chain termination method (the enzymatic method)

Sanger's method, also referred to as the dideoxynucleotide chain termination method, is based on the use of dideoxynucleotides (ddNTP) in addition to the normal nucleotides (dNTP) found in DNA. There are two key elements to Sanger's sequencing method: first, the discovery and application of nucleotide analogs; and second, a balanced proportion of dNTPs and their analogs ddNTPs in such a way that the chain termination occurs randomly at each base position creating a ladder representing each base being synthesized during DNA chain elongation.

dNTPs are building blocks of DNA. The ddNTPs are analogs of the dNTP that are essentially the same as nucleotides except they contain a hydrogen group (H) on the ribose 3′ carbon instead of a hydroxyl group (OH). Because they are the structural analogs of dNTPs, they can be used as substrates by DNA polymerases to be incorporated into DNA. ddNTPs, once incorporated into DNA, prevent the addition of further nucleotides because of their lack of a hydroxyl group at the 3′ of the DNA complex, leading to the termination of the DNA chain. If we suppose that only dNTPs are used for a synthesis reaction, the DNA chain would be synthesized continuously as long as the template and substrate exist. In contrast, if only ddNTPs are used, the DNA chain would be terminated at the first base when a dideoxynucleotide is incorporated. The key to Sanger's method is to use a proper proportion of dNTP and ddNTPs so that a ladder is created representing DNA chains terminated at every base position during DNA synthesis.

Historically, Sanger's method required the preparation of a single-stranded template from the DNA to be sequenced. This requirement was demanded by the DNA polymerase used for sequencing DNA. DNA polymerases all have a couple of important requirements for their polymerase activities: first, DNA polymerases require a template to copy from; second, they require the presence of a primer annealed to the template with the remaining portion of the template single-stranded so the DNA chain can be extended from the primer. When these conditions are present, DNA polymerase can synthesize DNA with a proper buffer system and in the presence of dNTPs.

The preparation of single-stranded templates was a great challenge back in the 1970s and the early 1980s. Use of the single-stranded phage M13 led to the innovation

of technology for the production of single-stranded templates. A group of scientists headed by Dr. Joachim Messing identified a region in the M13 that was required for the packaging of the M13 single-stranded DNA within the intergenic region (Messing et al. 1977). This discovery led to the development of a series of pUC cloning vectors (such as pUC118 and pUC119) aimed at production of single-stranded templates from cloned DNA. The most popular plasmid vectors used today all contain the intergenic region for the production of single-stranded DNA.

The discovery of thermostable DNA polymerases such as *Taq* DNA polymerase, and, thereafter, the invention of polymerase chain reaction (PCR) in 1986 completely eliminated the need for the production of single-stranded DNA. In a typical PCR reaction, the double-stranded DNA templates are denatured using heat. The denatured single-stranded DNA is annealed with the sequencing primer, and the primer is extended by the *Taq* polymerase. Based on the principles of PCR and sequencing, the cycle sequencing procedures were developed to sequence DNA using a double-stranded DNA template.

A typical cycle sequencing reaction contains several steps. Before the DNA can be sequenced, it has to be denatured into single strands using heat. Next, a primer is annealed to one of the template strands. This primer is specifically designed so that its 3′ end is located next to the DNA sequence of interest. Most often, the sequencing primer is designed based on the cloning vector at a region immediately outside of the poly-cloning sites. Typical sequencing primers are the T3 and T7 primers used with pBlueScript cloning vectors, or T7 and Sp6 sequencing primers used with pGEM cloning vectors. To detect the randomly terminated sequencing ladder, the sequencing primer or one of the nucleotides should be radioactively or fluorescently labeled. To illustrate the procedures of sequencing, let us use the classical sequencing reactions as an example. In a typical four-lane sequencing reaction, once the primer is annealed to the DNA, the solution is divided into four tubes labeled "A," "C," "G," "T" and then reagents are added to these samples as follows:

- "A" tube: all four dNTPs, ddATP, and DNA polymerase
- "C" tube: all four dNTPs, ddCTP, and DNA polymerase
- "G" tube: all four dNTPs, ddGTP, and DNA polymerase
- "T" tube: all four dNTPs, ddTTP, and DNA polymerase

As the DNA is synthesized, nucleotides are added on to the growing DNA chain by the polymerase. However, remember that the reaction is using a population of molecules in large numbers. On occasion, a proportional number of molecules incorporate a dideoxynucleotide into the chain in place of a normal nucleotide, resulting in a chain termination. Recall that we had four separate reactions for A, C, G, and T, respectively. In the reaction labeled "A," chain termination products should generate products terminating at positions with A (at the template positions containing T). Very similarly, in the reactions labeled "C," "G," and "T," newly synthesized DNA segments should be terminated at positions with C, G, and T, respectively. The sequencing ladder is resolved by denaturing polyacrylamide electrophoresis on a sequencing gel (Figure 25.2).

Although four-lane sequencing is still used by several platforms such as automated sequencers from LI-COR, the availability of fluorescent labels has allowed the development of single-lane sequencers, which increase sequencing efficiency at least four-fold. The sequencing platforms used by ABI automated sequencers such as ABI

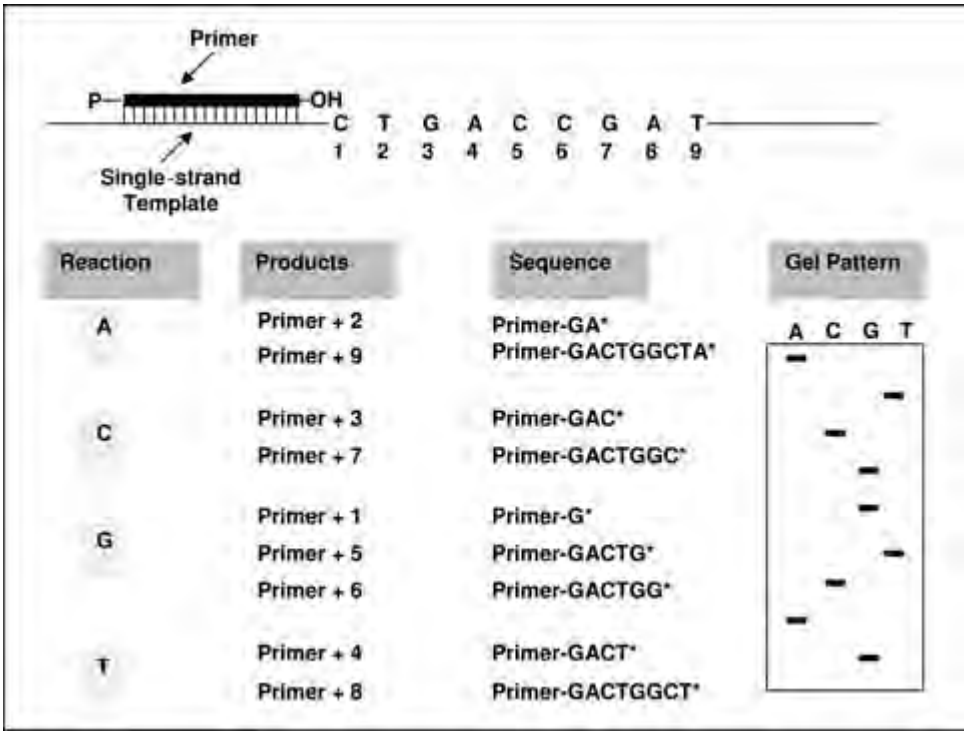

Figure 25.2. Principles of Sanger's dideoxynucleotide chain termination sequencing. Single-stranded template is annealed with a sequencing primer. After the dNTP and ddNTP mix is provided to the reaction containing DNA polymerase with proper buffer, synthesis of the new strand of DNA starts by primer extension. The incorporation of dideoxynucleotide (* in the figure) into DNA terminates the growing DNA chain, creating a sequencing ladder at each base position, depending on the base to be incorporated.

PRISM 3730 XL or 3130 XL, are all single-lane sequencing using BigDye labeling. The sequencing principles of single-lane sequencing are exactly the same as four-lane sequencing except that each dideoxynucleotide is labeled with a fluorescent dye with a different wavelength of excitation and emission so that they can be mixed in the same labeling reaction. The labeled fluorescent products are separated on a sequencing gel and subsequently detected by appropriate laser detectors.

Each sequencing reaction has a limited read length on sequencing gels, usually around 300–500 base pairs (bp). During the early days of sequencing, the ability to sequence long stretches of DNA depended on the development of nested overlapping clones allowing the entire DNA segment to be sequenced because the cost of oligonucleotide synthesis was very high at that time (as high as $30 per base). A number of methodologies for the production of nested clones were then developed (e.g., Anderson 1981, Henikoff 1984, Dale et al. 1985, Liu and Hackett 1993). Because these methodologies are no longer in use for sequencing today, these are only mentioned here as a part of the history of sequencing.

The primer walking approach is routinely used to completely sequence a long segment of DNA. Starting in 1994, primer synthesis costs reached a level of $1 per base from a historical high of $30 per base, and continued technological advancements

since then have allowed further reduction in the cost of primer synthesis. Lower primer costs meant that primer walking for sequencing a long segment of DNA could be cost effective. The primer walking method uses a previously generated sequence for the design of an additional primer for continued sequencing (Kim et al. 2000). The cycle of sequencing primer design followed by sequencing is continued until the desired segment of DNA is completely sequenced. Although cost effective, primer walking is slow when applied to very large pieces of DNA. In these cases, a shotgun sequencing approach is used, and complete sequences are then assembled by overlapping the generated sequences. This approach will be further discussed in the next chapter, Sequencing the Genome.

Maxam-Gilbert sequencing method

The Maxam-Gilbert sequencing method is no longer used for large-scale DNA sequencing, therefore, it is only briefly described here as a part of the historical perspective. This chemical cleavage method involves modification of the bases in DNA followed by base-specific cleavage using chemicals.

The Maxam-Gilbert sequencing method starts with double-stranded DNA. Most often, the DNA to be sequenced is radioactively labeled by attaching a radioactive phosphorus (^{32}P) group to the 5′ end, using a polynucleotide kinase enzyme and γ-^{32}P-ATP. Through exchange reactions of the polynucleotide kinase, the γ-^{32}P is transferred to the DNA molecule. The next step is to separate the strands of DNA. The two strands of the DNA are separated by treatment with dimethyl sulphoxide and heating to 90°C before being purified by electrophoresis. The third step is to conduct chemical modifications. The single-stranded sample is split into four separate samples, and each is treated with one of the cleavage reagents. This step involves alteration of bases. The four reactions follow: (1) G only, treated with dimethylsulphate (DMS); (2) A + G, treated with DMS plus formic acid; (3) C only, treated with hydrazine in 1.5M NaCl; and (4) C + T, treated with hydrazine. After base modifications, all modified bases are cleaved with piperdine. The cleavage products are electrophoresed on a polyacrylamide denaturing gel in four separate lanes with the smallest fragments run fastest. The sequence is read from the bottom (5′) to top (3′) of the gel.

When first developed, the Maxam-Gilbert sequencing method was thought to have the advantage of being able to sequence stretches of DNA that could not be sequenced by the enzymatic method due to high guanine-cytosine (GC) content or strong secondary structures. However, the toxic chemicals involved and the difficulties of adapting such a method for automation doomed this historically Nobel Prize winning technique to eventual obscurity.

Pyrosequencing and the 454 sequencing platform

Sequence determination is most commonly performed using the Sanger method. Decades of technological development and refining of the sequencing platforms made the dideoxy chain termination method the gold standard in the DNA sequencing business. All of the genomes sequenced to date including those of bacteria, archeal,

and eukaryotes (http://www.ncbi.nlm.nih.gov, and http://www.tigr.org) were sequenced using the chain termination method. However, this technique faces limitations in both throughput and cost for most future applications. Research needs have changed from sequencing a single human genome, as done in the human genome project, to sequencing thousands of genomes from groups of people with certain diseases to provide insight into genetic determinants of for example, cancer, heart disease, and high blood pressure. Many research groups around the world have put great effort into the development of alternative principles for DNA sequencing (Ronaghi 2001). Three methods that hold promise are sequencing by hybridization (Bains and Smith 1988, Drmanac et al. 1989, Khrapko et al. 1989), parallel signature sequencing based on ligation and cleavage (Brenner et al. 2000), and pyrosequencing (Ronaghi et al. 1998). Of these, pyrosequencing has emerged as the new sequencing methodology closest to being ready for widespread, practical applications. Pyrosequencing has the potential advantages of accuracy, flexibility, parallel processing, and can be easily automated. Furthermore, the technique dispenses with the need for labeled primers, labeled nucleotides, and gel electrophoresis. Because the newly developed 454 sequencing platform is based on pyrosequencing, here let us first introduce the principles of pyrosequencing.

Pyrosequencing is based on the detection of released pyrophosphate (PP_i) during DNA synthesis. Its principles, procedures, strengths, and weaknesses, as well as potential applications are discussed in a recent review by Ronaghi (2001). Pyrosequencing involves measurement of light generated through a cascade of enzymatic reactions after each nucleotide base incorporation. Technically, pyrosequencing can be divided into several steps for the sake of understanding. Step 1, a sequencing primer is hybridized to a single stranded, PCR amplified, DNA template, and incubated with the enzymes, DNA polymerase, adenosine 5′-triphosphate (ATP) sulfurylase, luciferase and apyrase, and the substrates, adenosine 5′-phosphosulfate (APS), and luciferin. Step 2, the synthesis of the first base by incorporating the needed nucleotide generates PP_i. Pyrophosphate is the natural product of DNA polymerization (Figure 25.3). Step 3, PP_i is converted by sulfurylase into ATP that reacts with luciferin to generate light in the presence of luciferase. The light is recorded by a charge-coupled device (CCD) camera and quantified as pyrograms. Step 4, the last step of the cascade after the recording of light, is to clear up the system, allowing the start of sequencing for the second base. This requires the degradation of all existing nucleotides and ATP by injecting apyrase. The sequential injection of one nucleotide at a time, coupled to the generation of light when the base is incorporated into the growing chain of DNA, allows the determination of sequences. The overall reaction from polymerization to light detection takes place within 3–4 seconds at room temperature. One pmol of DNA in a pyrosequencing reaction yields 6×10^{11} ATP molecules which, in turn, generate more than 6×10^9 photons at a wavelength of 560 nanometers (Ronaghi 2001). When mononucleotide repeats are encountered in the sequence, the pyrosequencing reaction continuously incorporates the repeated nucleotide until it reaches a different nucleotide. The light signal produced is proportional to the number of mononucleotides incorporated up to 8 bases. Mononucleotide repeats greater than 8 bp cannot be accurately sequenced by pyrosequencing (see below).

The theory behind pyrosequencing is sequencing-by-synthesis (Melamede 1985). It has not been used previously for sequencing because of the technical challenges of

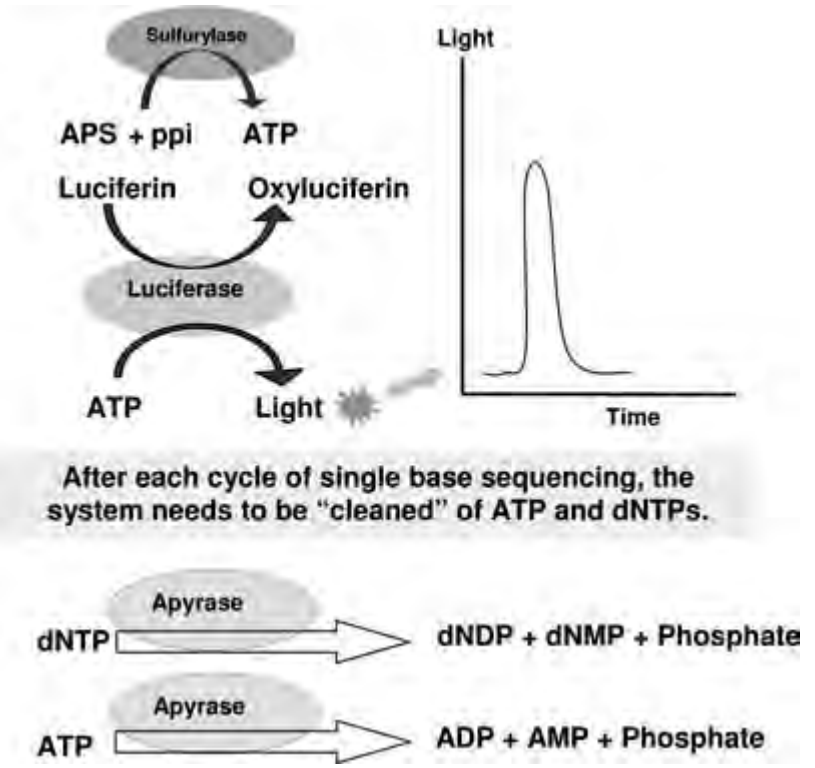

Figure 25.3. Schematic presentation of the principles of pyrosequencing.

synchronization of all the enzymatic reactions. Recent progress in luciferase chemistry and the availability of apyrase and sulfurylase have contributed to the technological advances toward the application of pyrosequencing as a sequencing strategy. However, pyrosequencing still faces great challenges in several areas (Ronaghi 2001). An inherent problem with the described method is de novo sequencing of polymorphic regions in heterozygous DNA material. In most cases, it will be possible to detect the polymorphism. If the polymorphism is a base substitution, it will be possible to obtain a synchronized extension after the substituted nucleotide. If the polymorphism is a deletion or insertion of the same nucleotide (e.g., AAAAAAA versus AA) as the adjacent nucleotide on the DNA template, the sequence after the polymorphism will be synchronized. However, if the polymorphism is a deletion or insertion of another type of nucleotide, the sequencing reaction can become out of phase, making the interpretation of the subsequent sequence difficult. If the polymorphism is known, it is always possible to use programmed nucleotide delivery to keep the extension of different alleles synchronized after the polymorphic region. Another problem is the difficulty in determining the number of incorporated nucleotides in homopolymeric regions, due to the nonlinear light response following incorporation of more than 5–8 identical nucleotides. Finally, but most importantly, pyrosequencing must overcome the problem of short read lengths. Currently, between 100–200 bp can be generated using pyrosequencing-derived sequencing technologies. For genome

sequencing and genome sequence assembly, long reads are definitely desirable. In spite of these inherent problems, pyrosequencing technology has recently been developed into a new sequencing platform called "454 sequencing." The platform has the potential to revolutionize genome sequencing, providing higher throughput and significantly lower costs than current standards.

The 454 sequencing platform

454 Life Sciences has developed a revolutionary technology, producing tens of millions of raw bases per hour on a single instrument. Many biologically meaningful and complex regions of genomes can be analyzed with this system without the time or cost constraints of current DNA sequencing methods. Through this technology, 454 Life Sciences provides an enabling solution for ultra-high-throughput DNA sequencing (http://www.454.com/). As only limited information is available at this time from a publication in *Nature* (Margulies et al. 2005), I will briefly describe the 454 sequencing principles and processes here based on the available information.

The 454 sequencing platform is based on the principles of pyrosequencing and uses microfabricated high-density picolitre reactors (Margulies et al. 2005, Figure 25.4). No cloning is necessary for 454 sequencing, and thus certain terms such as DNA library, etc., have specific meanings here for the description of the 454 sequencing system. The clonal DNA used for sequencing is obtained by clonal PCR amplification of a single molecule in emulsified water-in-oil microreactors. Preparation of the DNA library consists of a few simple steps. Genomic DNA is fractionated into smaller fragments (300–500 bps) that are subsequently filled in to polished ends (blunted), allowing ligation of adaptors to the genomic DNA for PCR amplification. To prevent intermolecular ligation of the genomic DNA fragments, the DNA fragments are dephosphorylated. Short adaptors (A and B) are then ligated onto the ends of the fragments. After ligation, the gap needs to be repaired, presumably using a DNA ligase. The adaptors provide priming sequences for both amplification and sequencing of the sample-library fragments. The two adaptors are different. Adaptor B contains a 5′-biotin tag that enables immobilization of one strand of the library onto streptavidin-coated beads. The nonbiotinylated strand is released and used as a single-stranded template DNA library.

The single-stranded template DNA library is immobilized onto beads carrying short primers complementary to the Adaptor A sequences by base pairing. The key element here is attaining the correct proportion of beads to DNA molecules so that only one molecule is captured by each bead. The beads containing a single molecule of the single-stranded template are emulsified with the amplification reagents in a water-in-oil mixture. Each bead is captured within its own microreactor where PCR amplification occurs. This results in bead-immobilized, clonally amplified DNA fragments.

The single-strand template DNA library beads are added to the DNA Bead Incubation Mix (containing DNA polymerase) and are layered with enzyme beads (containing sulfurylase and luciferase) onto the PicoTiterPlate device. The device is centrifuged to deposit the beads into the wells. The layer of enzyme beads ensures that the DNA beads remain positioned in the wells during the sequencing reaction. Due to the size of the wells in relation to the beads, only one bead containing a

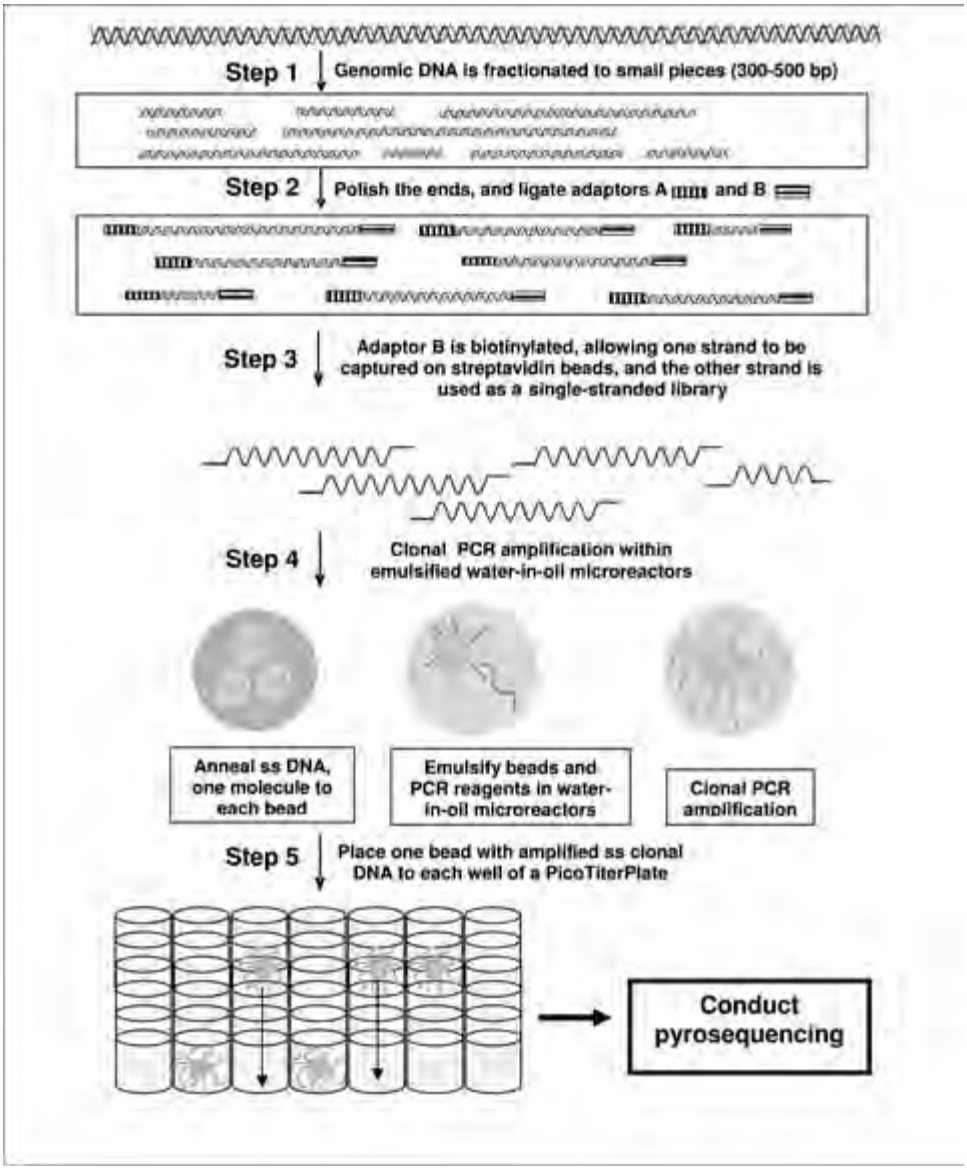

Figure 25.4. Schematic presentation of the principles of the 454 sequencing platform. (Also see color plate.)

specific clonally amplified genomic DNA segment should be placed into each well of the PicoTiterPlate device. The loaded PicoTiterPlate device is placed into the "454 sequencer," the Genome Sequencer 20 Instrument, which performs "pyrosequencing-like" reactions. See above under Pyrosequencing. Unlike a traditional pyrosequencing reaction, hundreds of thousands of beads, each with millions of copies of clonally amplified DNA, are sequenced in parallel. Each well of the PicoTiterPlate device is a separate pyrosequencing reaction. If a nucleotide complementary to the template strand is flowed into a well, the polymerase extends the existing DNA strand

by adding nucleotide(s). Addition of one (or more) nucleotide(s) results in a reaction that generates a light signal that is recorded by the CCD camera in the instrument. The signal strength is proportional to the number of nucleotides incorporated in a single nucleotide flow. Typically, 200,000 reactions are conducted in parallel in a single run. Assuming generation of 100 bp by a single reaction, each run should generate 20 million bp or more of sequence in several hours using a single instrument. Imagine, this is approximately a 10× coverage of a bacterial genome!

From the above introduction, it is clear that the 454 sequencing platform holds much potential. The current major problems prohibiting its application for the sequencing of complex genomes are its relatively short sequencing reads, and difficulties in accurate determination of homopolymeric run in the DNA. The short reads complicate genome sequence assembly, while the inability to determine the number of bases within a long homopolymeric run prohibits accurate sequencing of genomes. These problems are more significant for complex genomes with high levels of repeat structure. However, the technologies' high throughput and low costs are very attractive, especially for aquaculture species. As the technology is perfected to minimize these drawbacks, the 454 sequencing platform will show even greater promise.

The Solexa sequencing platform

Solexa's core technology, the Clonal Single Molecule Array™ technology, allows simultaneous analysis of hundreds of millions of individual molecules. With its unparalleled data density, Solexa's platform will dramatically improve the speed and reduce the cost of a range of genetic analysis applications, including whole-genome resequencing and expression profiling. Although Solexa is developing and plans on commercializing a new platform based on Sequencing-By-Synthesis (SBS), the company currently offers a range of expression profiling and small RNA analysis based on its Massively Parallel Signature Sequencing (MPSS) technology. Both technologies (MPSS and SBS) leverage massively parallel sequencing of short DNA and complementary DNA (cDNA) fragments to generate data from millions of fragments simultaneously. However, by leveraging its Clonal Single Molecule Array™ technology and reversible terminator chemistry, Solexa's SBS approach is anticipated to generate up to 1 billion bases of data per run at costs far more economical than that of MPSS. Given the data density and economical advantages of SBS over MPSS, Solexa plans on transitioning its MPSS-based service offerings to SBS. As limited information is available in published sources, here I introduce some basic concepts of the Solexa sequencing technology using Web-based information (http://www.solexa.com/).

Solexa's SBS uses four proprietary fluorescently labeled modified nucleotides. These specially created nucleotides, which also possess a reversible termination property, allow each cycle of the sequencing reaction to occur simultaneously in the presence of all four nucleotides (A, C, T, G). In the presence of all four nucleotides, the polymerase is able to select the correct base to incorporate, with the natural competition between all four alternatives leading to higher accuracy than methods where only one nucleotide is present in the reaction mix at a time, which require the enzyme to reject an incorrect nucleotide. Sequences where a particular base is repeated one after another ("homopolymer repeats") are dealt with as for any other

sequence and with high accuracy; this avoids the problems of measuring intensity and deducing how many bases were present in the repeat that are the cause of uncertainty seen with "one base per reaction" methods, as described above with the 454 sequencing platform.

Conclusion

Sequencing technology has been applied in research for 30 years. Although multiple sequencing strategies have been invented during this time, Sanger's sequencing strategy has been used for sequencing all the genomes to date. Decades of optimization, refining, and streamlined operations have allowed Sanger's method to be applied on an industrial scale at a reasonable cost for worldwide mutual efforts such as the Human Genome Project. However, the genome projects of most eukaryotic and prokaryotic species have no worldwide interest. In addition, interest in the biomedical community has shifted from obtaining a single human genome sequence to comparing thousands of human genomes in the context of susceptibility to cancer, heart disease, diabetes, high blood pressure, and so on. The current sequencing technologies cannot meet these tremendous sequencing demands with current levels of research funding. Even though the cost of sequencing has been reduced by at least 100 times in the last decade, sequencing a genome of a comparable size to the mammalian genomes still takes $10 million or more using current sequencing technology. This price tag is too expensive for individual human genome sequencing and far too expensive for the underfunded aquaculture community. Current sequencing costs need to be reduced another 100-fold before genome sequencing can become routine. Emerging sequencing technologies such as 454 sequencing and Solexa sequencing have the potential to meet these demands. Aquaculture genome sequencing requires high efficiency and low cost, because we have a small research community. Hopefully the emerging sequencing technologies will become mature soon to provide low-cost and high-quality sequencing. With the current state of technology, 454 sequencing and Solexa sequencing would allow the entire genomes of many fish species to be sequenced for less than $1 million. Although genome sequence assembly may be difficult with current read lengths, it may be possible to anchor "thousands of sequence islands" created by 454 sequencing or Solexa sequencing onto well-developed physical and genetic maps, which may prove to be sufficient for agricultural research objectives.

References

Anderson S. 1981. Shotgun DNA sequencing using cloned DNase I-generated fragments. Nucl Acids Res, 9, pp. 3015–3027.

Bains W and GC Smith. 1988. A novel method for nucleic acid sequence determination. J Theoret Biol, 135, pp. 303–307.

Brenner S, SR Williams, EH Vermaas, T Storck, K Moon, C McCollum, JI Mao, S Luo, JJ Kirchner, and S Eletr. 2000. In vitro cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs. Proc Natl Acad Sci, 97, pp. 1665–1670.

Dale RM, BA McClure, and JP Houchins. 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18 S rDNA. Plasmid, 13, pp. 31–40.

Drmanac R, I Labat, I Brukner, and R Crkvenjakov. 1989. Sequencing of megabase plus DNA by hybridization: Theory of the method. Genomics, 4, pp. 114–128.

Henikoff S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene, 28, pp. 351–359.

Khrapko KR, YP Lysov, AA Khorlyn, VV Shick, VL Florentiev, and AD Mirzabekov. 1989. An oligonucleotide hybridization approach to DNA sequencing. FEBS Lett, 256, pp. 118–122.

Kim S, A Karsi, RA Dunham, and Z Liu. 2000. The skeletal muscle alpha-actin gene of channel catfish (*Ictalurus punctatus*) and its association with piscine specific SINE elements. Gene, 252, pp. 173–181.

Liu Z and PB Hackett. 1993. Reverse cloning procedure for generation of subclones for DNA sequencing. Methods Enzymol, 218, pp. 47–58.

Margulies M, M Egholm, WE Altman, S Attiya, JS Bader, LA Bemben, J Berka, MS Braverman, YJ Chen, Z Chen, SB Dewell, L Du, JM Fierro, XV Gomes, BC Godwin, W He, S Helgesen, CH Ho, GP Irzyk, SC Jando, ML Alenquer, TP Jarvie, KB Jirage, JB Kim, JR Knight, JR Lanza, JH Leamon, SM Lefkowitz, M Lei, J Li, KL Lohman, H Lu, VB Makhijani, KE McDade, MP McKenna, EW Myers, E Nickerson, JR Nobile, R Plant, BP Puc, MT Ronan, GT Roth, GJ Sarkis, JF Simons, JW Simpson, M Srinivasan, KR Tartaro, A Tomasz, KA Vogt, GA Volkmer, SH Wang, Y Wang, MP Weiner, P Yu, RF Begley, and JM Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature, 437, pp. 326–327.

Maxam A and W Gilbert. 1977. A new method of sequencing DNA. Proc Natl Acad Sci USA, 74, pp. 560–564.

Melamede RJ. 1985. Automatable process for sequencing nucleotide. US Patent No. US4863849.

Messing J, B Gronenborn, B Muller-Hill, and PH Hofschneider. 1977. Filamentous coliphage M13 as a cloning vehicle: insertion of a *Hin*d II Fragment of the *lac* regulatory region in M13 replicative form in vitro. Proc Natl Acad Sci USA PNAS, 74, pp. 3642–3646.

Ronaghi M, M Uhlæn, and P Nyræn. 1998. A sequencing method based on real-time pyrophosphate. Science, 281, pp. 363–365.

Ronaghi M. 2001. Pyrosequencing Sheds Light on DNA Sequencing. Genome Research, 11, pp. 3–11.

Sanger F, S Nicklen, and AR Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA, 74, pp. 5463–5467.