Quarterly Reviews of Biophysics **35**, 2 (2002), pp. 169–200. © 2002 Cambridge University Press
DOI: 10.1017/S0033583502003797 Printed in the United Kingdom

A review of DNA sequencing techniques

Lilian T. C. França[1], Emanuel Carrilho[2] and Tarso B. L. Kist[3]*

[1] Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil and Instituto de Biofísica Carlos Chagas Filho, CCS, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil (E-mail: lila@biof.ufrj.br)
[2] Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, SP, Brazil (E-mail: emanuel@iqsc.sc.usp.br)
[3] Departamento de Biofísica, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, 91501–970, Porto Alegre, RS, Brazil (E-mail: tarso@orion.ufrgs.br)

1. Summary

The four best known DNA sequencing techniques are reviewed. Important practical issues covered are read-length, speed, accuracy, throughput, cost, as well as the automation of sample handling and preparation. The methods reviewed are: (i) the Sanger method and its

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

most important variants (enzymic methods); (ii) the Maxam & Gilbert method and other chemical methods; (iii) the Pyrosequencing™ method – DNA sequencing in real time by the detection of released pyrophosphate (PPi); and (iv) single molecule sequencing with exonuclease (exonuclease digestion of a single molecule composed of a single strand of fluorescently labelled deoxynucleotides). Each method is briefly described, the current literature is covered, advantages, disadvantages, and the most suitable applications of each method are discussed.

2. Introduction

DNA sequencing techniques are key tools in many fields. A large number of different sciences are receiving the benefits of these techniques, ranging from archaeology, anthropology, genetics, biotechnology, molecular biology, forensic sciences, among others. A silent and remarkable revolution is under way in many disciplines; DNA sequencing is promoting new discoveries that are revolutionizing the conceptual foundations of many fields. At the same time new and very important issues are emerging with these developments, such as bioethical questions and questions related to public health and safety.

In this review we will follow the chronological development of the methods. We will start in Section 3 with the methods developed by Sanger and his collaborators in the 1970s. The Maxam & Gilbert method and other chemical methods are reviewed in Section 4. The PPi method – based on detection of PPi released on nucleotide incorporation during chain extension by polymerase – is reviewed in Section 5. The methods based on single molecule detection are reviewed in Section 6. Finally, the concluding remarks are given in Section 7.

3. Sanger's method and other enzymic methods

The first method described by Sanger and Coulson for DNA sequencing was called 'plus and minus' (Sanger & Coulson, 1975). This method used *Escherichia coli* DNA polymerase I and DNA polymerase from bacteriophage T4 (Englund, 1971, 1972) with different limiting nucleoside triphosphates. The products generated by the polymerases were resolved by ionophoresis on acrylamide gels. Due to the inefficacy of the 'plus and minus' method, 2 yr later, Sanger and his co-workers described a new breakthrough method for sequencing oligonucleotides via enzymic polymerization (Sanger *et al.* 1977). This method, which would revolutionize the field of genomics in the years to come, was initially known as the chain-termination method or the dideoxynucleotide method. It consisted of a catalysed enzymic reaction that polymerizes the DNA fragments complementary to the template DNA of interest (unknown DNA). Briefly, a ^{32}P-labelled primer (short oligonucleotide with a sequence complementary to the template DNA) was annealed to a specific known region on the template DNA, which provided a starting point for DNA synthesis. In the presence of DNA polymerases, catalytic polymerization of deoxynucleoside triphosphates (dNTP) onto the DNA occurred. The polymerization was extended until the enzyme incorporated a modified nucleoside [called a terminator or dideoxynucleoside triphosphate (ddNTP)] into the growing chain.

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

whereas the residue at the 3′-end was determined by the dideoxynucleotide used in the reaction. After all four reactions were completed, the mixture of different-sized DNA fragments was resolved by electrophoresis on a denaturing polyacrylamide gel, in four parallel lanes. The pattern of bands showed the distribution of the termination in the synthesized strand of DNA and the unknown sequence could be read by autoradiography. For a better understanding of the Sanger reaction, see Fig. 1. The enzymic method for DNA sequencing has been used for genomic research as the main tool to generate the fragments necessary for sequencing, regardless of the sequencing strategy. Two different approaches, shotgun and primer walking sequencing, are the most used (Griffin & Griffin, 1993). The main aspects of each strategy are described below in more detail.

3.1 Random approach

Also known as shotgun sequencing, this is a random process because there is no control of the region that is going to be sequenced, at least in the usual procedures (there are exceptions, for instance see the procedure described by Lander *et al.* 2001). Genomic DNA is randomly fragmented (by sonication, nebulization, or other scission methods) into smaller pieces, normally ranging from 2 to 3 kb. The fragments, inserted into a vector, are replicated in a bacterial culture. Several positive amplifications are selected, and the DNA is extensively sequenced. Due to the random nature of this process, the sequences generated overlap in many regions (Adams *et al.* 1996). The process of overlaying or alignment of the sequences is called sequence assembly. Shotgun sequencing normally produces a high level of redundancy (the same base is sequenced 6–10 times, in different reactions) which affects the total cost. A new variation of the method introduced by Venter *et al.* (1996) involved shotgunning a whole genome at once. This strategy depended enormously on computational resources to align all generated sequences. However, the efforts were rewarded with the sequencing of the *Haemophilus influenzae* genome in only 18 months (Fleischmann *et al.* 1995) and, more recently, the human genome (Venter *et al.* 2001).

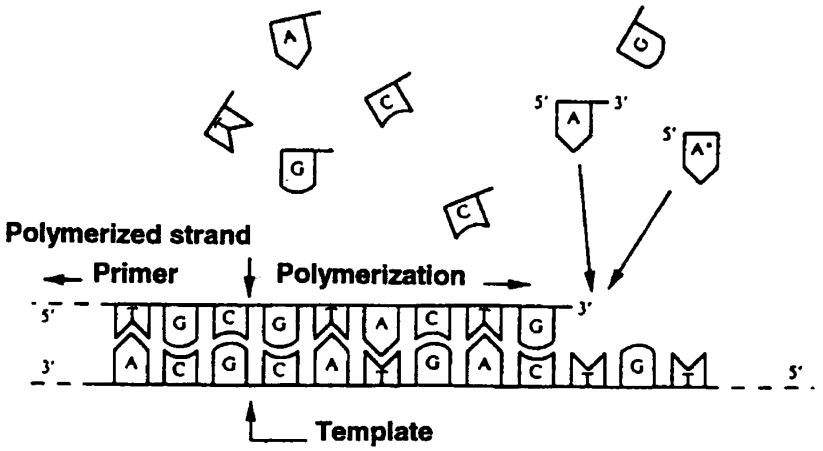
Shotgun sequencing is well established, with ready availability of optimized cloning vectors, fluorescently labelled universal primers, and software for base calling and sequence assembly. The whole process has a high level of automation, from the cloning of the vectors and colony selection to the bases called. A simplified diagram of the shotgun process is summarized in Fig. 2. Although the random approach is fully compatible with automation, it can produce gaps in the sequence that can only be completed by direct sequencing of the region.

3.2 Direct approach

The other approach for genomic sequencing is the direct sequencing of unknown DNA within sites in which the sequence is known. For example, an unknown sequence of DNA is inserted into a vector and amplified. The first sequencing reaction is performed using the primers that hybridize to the vector sequence and polymerize the strand complementary to the template. A second priming site is then chosen inside the newly generated sequence, following the same direction as the first one. This approach is known as primer walking

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

(*a*)



(*b*)

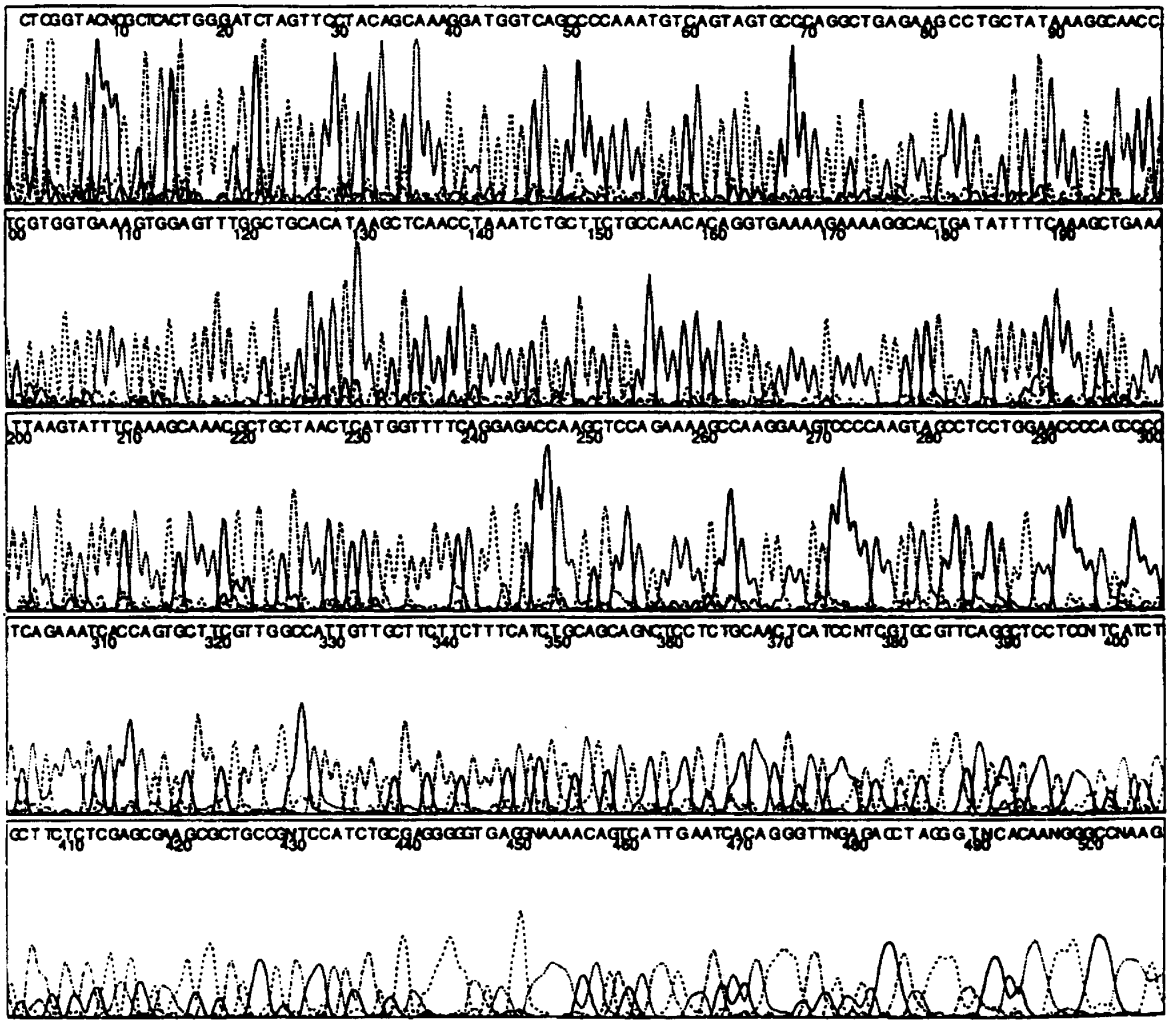


Fig. 1. Schematic representation of a sequencing process ('four-colour Sanger'): starting from many copies of the ssDNA to be sequenced, bearing a known 'marker' at the beginning of the unknown sequence, a short oligonucleotide 'primer' complementary to this marker is hybridized (i.e. paired) to the marker, in the presence of DNA polymerase and free nucleotides. This hybridization initiates reconstruction by the polymerase of a single strand complementary to the unknown sequence (*a*). Including in the nucleotide bath in which the polymerization takes place a small fraction of fluorescently labelled dideoxynucleotides (one different dye for each nucleotide type), which lack the OH group

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

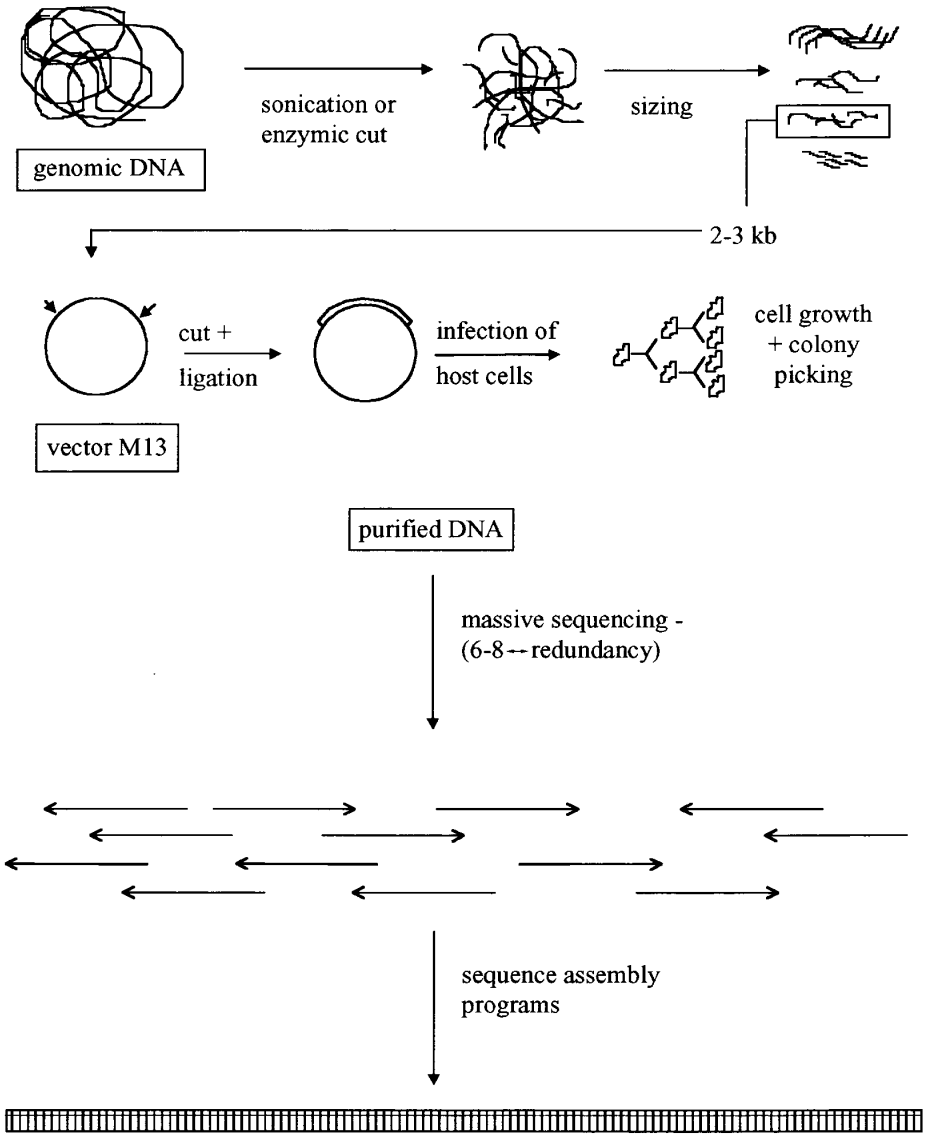


Fig. 2. Random sequencing approach or shotgun. The distinct processes involve first fragmentation of the DNA into 2–3 kbp range, fragments are then cloned into vectors and introduced into host cells for amplification. After purification, the DNA from individual colonies is sequenced, and the results are lined up with sequence-assembly programs.

random), as seen in Fig. 3. However, it requires the synthesis of each new primer, which, in the past, was time consuming and expensive, especially when dye-labelled primers were used.

Some alternatives were introduced to overcome the problems of time and cost (Ruiz-Martinez *et al.* 1996). Although slightly different, these approaches shared the same idea of using a short oligonucleotide library as a means to create a longer primer. The number of all sequences possible for an oligonucleotide with n bases is equal to 4^n. It was proposed by Kieleczawa *et al.* (1992) that a hexamer library containing 4096 oligonucleotides could be cost effective. While each new 18-mer primer is used only once for each new reaction site

---

newly synthesized ssDNAs are then separated by size electrophoretically [see electropherogram in (*b*)]: consecutive peaks correspond to DNA fragments differing by one base, and each line corresponds to

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.