ELSEVIER

Journal of Chromatography B, 782 (2002) 67–72

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Review

Pyrosequencing for microbial typing

Mostafa Ronaghi[a,*], Elahe Elahi[b]

[a]*Stanford Genome Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA*
[b]*Faculty of Science, Tehran University, Tehran, Iran*

Abstract

Pyrosequencing is a real-time DNA sequencing technique generating short reads rapidly and inexpensively. This technology has the potential advantage of accuracy, ease-of-use, high flexibility and is now emerging as a popular platform for microbial typing. Here, we review the methodology and the use of this technique for viral typing, bacterial typing, and fungal typing. In addition, we describe how to use multiplexing for accurate and rapid typing.

Keywords: Reviews; Pyrosequencing; Genotyping; Multiplexing

Contents

*Corresponding author. Tel.: +1-650-812-1971; fax: +1-650-812-1975.
E-mail address: mostafa@stanford.edu (M. Ronaghi).

1570-0232/02/$ – see front matter
PII: S1570-0232(02)00693-1

1. Introduction

Microbial typing provides the ability to discriminate between and catalogue microbial nucleic acid molecules. Typing is essential in several disciplines of microbiology research including taxonomy, studies of evolutionary mechanisms and phylogenetic relationships, population genetics of microorganisms, and microbial epidemiology. In microbial epidemiology, information on the distribution of types is used for defining the source and route of infection, for studying the persistence and re-infection rates, clonal selection in the host and microbial evolution. For typing, phenotypic methods such as biotyping, serotyping and hemagglutinin typing have little discriminatory power compared to genotypic methods [1]. Therefore, great efforts have been made to establish genotypic methods. These include creation of plasmid profiles [2] or DNA restriction profiles [3], ribotyping [4], pulsed-field gel electrophoresis [5], allele-specific polymerase chain reaction [6], restriction fragment length polymorphism analysis [7], ligase chain reaction [8], dideoxy chain termination DNA sequencing [9] and Pyrosequencing [10,11]. Among the above-mentioned techniques, only the sequencing techniques have sufficient discrimination potential to be used for all microbial typing and for undisputable detection of new subtypes. Additional parameters to be considered when choosing a platform for microbial typing include accuracy, reproducibility, throughput, ease-of-use and interpretation, cost and toxic procedures. We believe that Pyrosequencing fulfills the above-mentioned criteria for use in microbial typing.

Pyrosequencing employs coupled enzymatic reactions to detect inorganic pyrophosphate (PPi) released as a result of nucleotide incorporation by DNA polymerase. The released PPi is converted to ATP by ATP sulfurylase, which provides the energy for luciferase to oxidize luciferin and generate light (Fig. 1). Unincorporated nucleotides are degraded by apyrase prior to addition of the next nucleotide allowing iterative addition of nucleotides. Since the added nucleotide is known, the sequence of the template can be determined. Here, we review recent developments in Pyrosequencing and their application toward viral typing, bacterial typing, and fungal typing. We also present a scheme for multiplexing for rapid large-scale microbial typing.

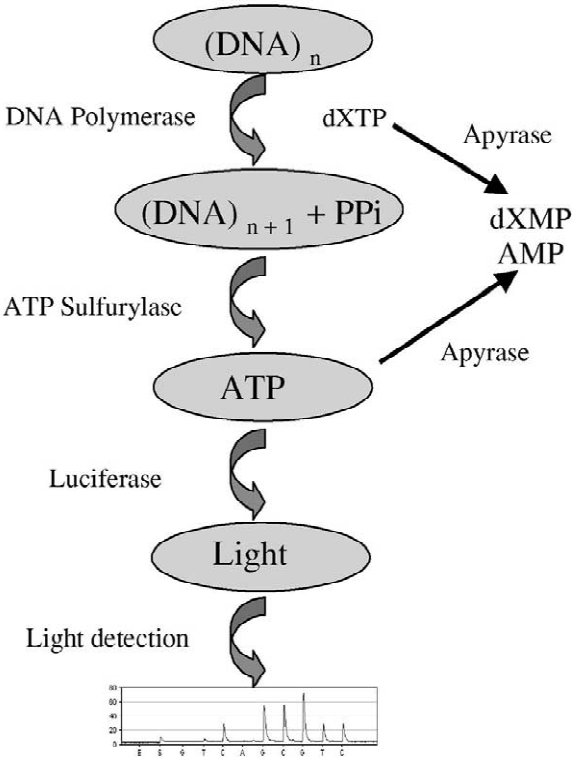

Fig. 1. The general principle of Pyrosequencing reaction system. A polymerase catalyzes incorporation of nucleotide(s) into a nucleic acid chain. As a result of the incorporation, a PPi molecule(s) is released and subsequently converted to ATP, by ATP sulfurylase. Light is produced in the luciferase reaction during which a luciferin molecule is oxidized. A photon detector device then detects the light and raw data will be available for further analysis. dXTP and ATP will be degraded by apyrase allowing iterative addition of nucleotides.

2. Recent developments for Pyrosequencing analysis

Pyrosequencing enables de novo sequencing of short reads of DNA. Sequence analysis by pyrosequencing involves three steps; PCR amplification, template preparation and pyrosequencing reactions. These three steps will briefly be reviewed here.

2.1. In vitro amplification

PCR amplification is usually performed on a variable or semi-conserved region of the microbial

genomes. Five to 50 μl PCR products are currently needed depending on the pyrosequencing machine to be used. Multiplex PCR has also been used for Pyrosequencing [12].

2.2. Template preparation

Prior to Pyrosequencing, the PCR product should be purified. Unincorporated nucleotides and PCR primers from the PCR reaction must be removed as they interfere with the pyrosequencing reaction. Salts present in the PCR reaction may inhibit the multi-enzyme system and should be removed or diluted. Three strategies for solid-phase template preparation were recently described [13]. For solid-phase strategies, we recommend the use of Sepharose beads for higher signal-to-noise ratios, cheapness, ease of template preparation scheme, and adaptability for automation.

2.3. Pyrosequencing analysis

Pyrosequencing methodology has gone through major technical developments to achieve a robust platform for DNA analysis. These include use of alfa-thio dATP instead of dATP [10], inclusion of nucleotide-degrading enzyme, apyrase, in the reaction system [11], the use of single-stranded DNA binding protein [14], and use of the purified Sp isomer of alfa-thio dATP [15]. These improvements have had a great impact on increasing the average length of reads to 50–60 nucleotides per read (Fig. 2). We have also developed a multiplexing strategy for pyrosequencing, which reduces the cost and time for analysis [12].

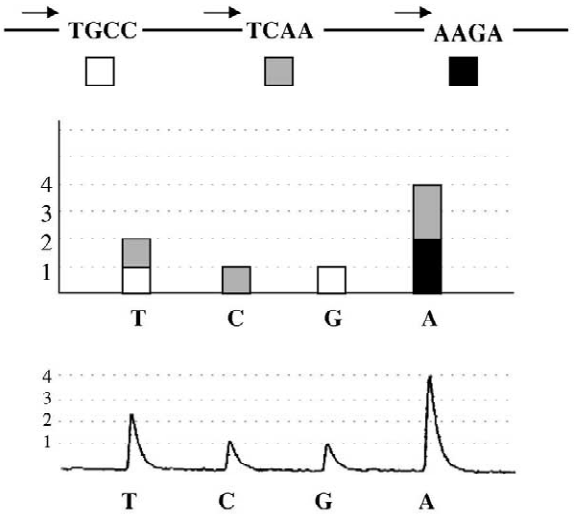

Fig. 3. Sequence (top panel), theoretical pattern and pyrogram (lower panel) obtained from subtype 1a using multiplex pyrosequencing. Height of peaks on the *y*-axis is determined by the number of incorporated nucleotides and the *x*-axis shows time. Nucleotides are added according to the specified order and are added at 1-min intervals. The added nucleotide is indicated below the pyrogram.

2.4. Multiplex Pyrosequencing

Multiplexing in Pyrosequencing employs simultaneous extension of several primers hybridized to one or several target DNA templates [12]. Pyrograms obtained from multiplex pyrosequencing reflect number of nucleotides incorporated by all primers and can easily be deconvoluted to singleplex pyrosequencing data (Fig. 3). In microbial typing, where many positions on a variable region may need to be checked, multiplexing provides a powerful system for rapid typing. We recently reported on the

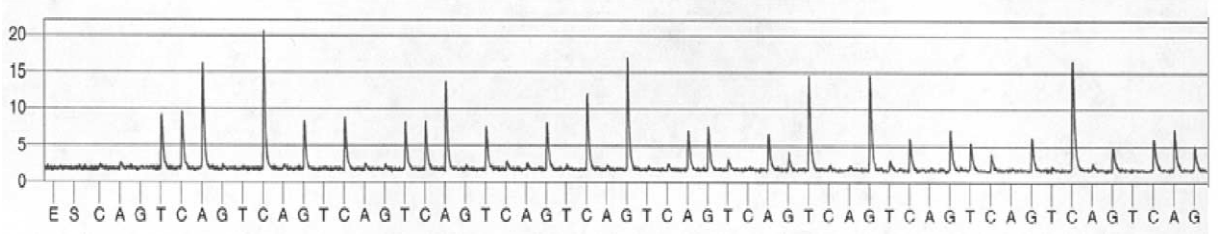

Fig. 2. Pyrogram obtained using standard pyrosequencing on subtype 1a of HCV. The order of nucleotide addition is indicated in the lower panel of the pyrograms.

use of this technique for typing of major subtypes of hepatitis C virus (HCV) and determination of multiple single nucleotide polymorphisms [12].

3. Pyrosequencing for viral typing

3.1. Hepatitis C virus (HCV) typing

For HCV typing semi-conserved regions such as NS5 or 5′ untranslated region have been used as a target for subtyping. We recently reported on the use of pyrosequencing for typing of Hepatitis C virus [16]. Pyrosequencing of the region could exactly determine all possible subtypes (Fig. 2). For population-specific subtyping, multiplex pyrosequencing could successfully be applied [12]. Fig. 3 demonstrates the theoretical and raw data obtained from multiplex Pyrosequencing.

3.2. Human papilloma virus (HPV) typing

The L1 region in HPV shows high heterogeneity and sequencing of stretches as short as 14 nucleotides in this gene has shown to be sufficient for classification of viruses [17]. In this study, 67 clinical samples were typed with three methods. Pyrosequencing and Sanger sequencing showed comparable data while mis-typing could be observed with allele-specific PCR.

4. Bacterial typing

4.1. Single nucleotide polymorphisms for bacterial typing

It has been shown that gene duplication may occur in bacteria, generating additional copies of an operon or generating gene families. Gene duplication has extensively been shown for the 16S rRNA gene and the existence of more than 10 copies has been reported [18]. Multi operon genes may vary in sequence by one or more nucleotides. These variations can be used for typing of different strains. Polymorphic genes in bacteria can also be used for genotyping. Unnerstad et al. recently showed the use of Pyrosequencing for grouping of *Listeria monocytogenes* strains using surface protein internalin genes [19].

4.2. Polymorphic bacterial genes for typing

It is common to use the 16S rRNA gene for identification of bacterial strains and species. Monstein et al. demonstrated Pyrosequencing of V1 and V3 regions is an accurate means for identification and subtyping of *Helicobacter pylori* [20]. However, quite often the ribosomal genes are found to be identical in closely related strains. When comparison of such strains is of interest, other polymorphic genes are considered. We recently reported on the use of the promotor region of pertusis toxin gene as an accurate means of typing of *Bordetella pertusis* and *parapertusis* [21,22]. Pyrosequencing has also been used for identification of *Bacillus anthracis* by sequence determination of 20 nucleotides of the Ba813 gene (www.pyrosequencing.com).

5. Microflora composition determination

Determination of microbial flora should enable the development of a detailed knowledge of microbial ecology. This knowledge is essential to derive scientifically valid probiotics or to monitor important microbes in a flora. The best studied microflora is probably the bacterial flora. However, many bacteria do not grow on agar plates, making this protocol inadequate for some investigations. A DNA-based technique would allow a better understanding on the composition of the flora. Pyrosequencing is an efficient technique for determination of the structure of flora. For this purpose, semi-conservative genes such as 16S rRNA gene in bacteria can be amplified using universal primer pairs. After amplification, the product can be cloned. The generated library can be screened by Pyrosequencing. Sequencing of 50–60 nucleotides provides sufficient data to taxonomically group the bacteria. For accurate determination, at least hundreds to thousands of colonies need to be screened since many bacteria which have a major role in a flora are present at below the percent level.

6. Fungal typing

Pyrosequencing technology has been used for identification of different clinically relevant fungi. The 18S rRNA gene serves as an informative target region and can be used for accurate typing. Pyrosequencing of up to 40 bases in a variable region using a general primer has been successfully used to type the relevant yeast strains [23].

7. Challenges in Pyrosequencing for microbial typing

There are two limitations in the applicability of this system for any microbial typing. The first is the limitation in the read length and the second is the limitation in complete incorporation of nucleotides in homopolymeric regions. In some typing applications, longer than 50–60 nucleotides are required. We suggest sequencing of multiple regions to increase the accuracy in typing. There is also difficulty in determining the number of incorporated nucleotides in homopolymeric regions, due to a non-linear light response to the sequential incorporation of more than 5–6 identical nucleotides. The polymerization efficiency through homopolymeric regions has been investigated and the results indicate that it is possible to incorporate up to 10 identical adjacent nucleotides in the presence of apyrase [14,24]. However, to elucidate the correct number of incorporated nucleotides, it may be necessary to use specific software algorithms for signal integration. For de novo sequencing of microbes, these regions are however unique and will not influence the accuracy of typing. It is also possible to add the nucleotide twice for a homopolymeric region to ensure complete polymerization.

8. Features of Pyrosequencing

Pyrosequencing offers the same accuracy as conventional DNA sequencing for short reads while it is more flexible and a large number of samples can be processed in parallel. Furthermore, pyrosequencing circumvents electrophoresis, size separation and the need for labeled nucleotides and labeled primers. The reaction is performed in real-time and the raw data is directly analyzed. This technique can be multiplexed which enables rapid and accurate screening of a large number of samples inexpensively. The cost of this technique for large-scale microbial typing is unmatched by the conventional DNA sequencing technique. Using the 384-well format of pyrosequencing (PTP system) one can analyze between 10,000 and 50,000 samples per day with a cost of 20–30 cents per sample. Currently, this is at least 10 times cheaper than the conventional sequencing technique. Enzymatic preparation of PCR products will simplify and reduce the cost. Pyrosequencing on miniaturized arrays are also presently being developed which will further reduce the cost of pyrosequencing by at least two orders of magnitude.

9. Instrumentation

Three commercial versions of multi-channel pyrosequencing instruments are currently in use (www.pyrosequencing.com). These instruments utilize 96-well and 384-well plate formats. All versions use a disposable inkjet cartridge for precise pneumatic delivery of small volumes of nucleotides into a temperature-controlled microtiter plate. The microtiter plate is continuously agitated during pyrosequencing to increase the rate of the enzymatic reactions. A lens array focuses the light signal generated from each reaction well on the microtiter plate onto a specific locus of a CCD-camera. Nucleotides are dispensed into alternating wells with a pulse delay to minimize cross-talk of generated light between adjacent wells. A cooled high sensitivity CCD-camera images the plate every second to follow the progress of the pyrosequencing reactions. Data acquisition modules and an interface for PC-connection are used in this instrument. Software running in a Windows 2000™ environment enables experimental control of the dispensation order for each well. The signals in a pyrogram typically generate high quality raw sequence data with excellent signal-to-noise ratio with peak height proportional to the number of incorporated nucleotides. These instruments facilitate analysis of between 5000 and 50,000 analyses per day. Future formats including array-

based pyrosequencing and microfluidics reaction devices are presently in development.

10. Software for pyrogram analysis

Four different software programs for SNP genotyping, allele quantification, tag sequencing, and multiplex analysis are now available (www.pyrosequencing.com). All these software programs are useful for the studies of microbial systems. All software programs are based on pattern recognition. The assignment of quality values is based on a number of different parameters, including difference in match between the best and next best choice of genotypes, agreement between expected and obtained sequence around the analyzed nucleotide, signal-to-noise ratios, variance in peak heights of adjacent nucleotides, and peak width. The tag software allows multiple addition of the same nucleotide to ensure complete polymerization in homopolymeric regions.

Acknowledgements

The authors are supported by NIH grant ≠ 2POI HG00205. Dr. Elahi is supported by a grant from the Iranian Blood Transfusion Center.

References

[1] U. Busch, H. Nitschko, J. Chromatogr. B 722 (1999) 263.
[2] L.W. Mayer, Clin. Microbiol. Rev. 1 (1988) 228.
[3] E.M. Southern, J. Mol. Biol. 98 (1975) 503.
[4] T. Popovic, C.A. Bopp, O. Olsvik, J.A. Kiehlbauch, in: D.H. Pershing, T.F. Smith, F.C. Tenover, T.J. White (Eds.), Diagnostic Molecular Microbiology, ASM Press, Washington, DC, 1993, p. 573.
[5] M.W. McDonell, M.N. Simon, F.W. Studier, J. Mol. Biol. 110 (1977) 119.
[6] H. Okamoto, S. Okada, Y. Sugiyama, K. Kurai, H. Iizuka, A. Machida, Y. Miyakawa, M. Mayumi, J. Gen. Virol. 72 (1991) 2697.
[7] C.A. Gaydos, L. Bodo, L. Welsh, E.W. Hook III, Sex. Transm. Dis. 19 (1992) 303.
[8] F. Barany, Proc. Natl. Acad. Sci. USA 88 (1991) 189.
[9] F. Sanger, S. Nicklen, A.R. Coulson, Proc. Natl. Acad. Sci. 74 (1977) 5463.
[10] M. Ronaghi, S. Karamouhamed, B. Pettersson, M. Uhlén, P. Nyrén, Anal. Biochem. 242 (1996) 84.
[11] M. Ronaghi, M. Uhlén, P. Nyrén, Science 281 (1998) 363.
[12] N. Pourmand, E. Elahi, R.W. Davis, M. Ronaghi, Multiplex pyrosequencing, Nucl. Acids Res. 1 (2002) E31.
[13] H. Fakhrai-Rad, N. Pourmand, M. Ronaghi, Hum. Mutat. 19 (2002) 479.
[14] M. Ronaghi, Anal. Biochem. 286 (2000) 282.
[15] B. Gharizadeh, T. Nordstrom, A. Ahmadian, M. Ronaghi, P. Nyren, Anal. Biochem. 301 (2002) 82.
[16] E. Elahi, N. Pourmand, R. Cheung, A. Rofoogaran, J. Boisvert, K. Samimi-Rad, R.W. Davis, M. Ronaghi, (2002) submitted.
[17] B. Gharizadeh, M. Kalantari, C.A. Garcia, B. Johansson, P. Nyren, Lab. Invest. 81 (2001) 673.
[18] B. Pettersson, T. Leitner, M. Ronaghi, G. Bölske, M. Uhlén, K.E. Johansson, J. Bacteriol. (1996) 4131.
[19] H. Unnerstad, H. Ericsson, A. Alderborn, W. Tham, M.L. Danielsson-Tham, J.G. Mattsson, Appl. Environ. Microbiol. 67 (2001) 5339.
[20] H. Monstein, S. Nikpour-Badr, J. Jonasson, FEMS Microbiol. Lett. 15 (2001) 103.
[21] M. Nygren, E. Reizenstan, M. Ronaghi, J. Lundeberg, J. Clin. Microbiol. 38 (2000) 55.
[22] M. Ronaghi, M. Nygren, J. Lundeberg, P. Nyren, Anal. Biochem. 267 (1999) 65.
[23] B. Gharizadeh, J. Loffler, E. Norberg, J. Tolemar, S. Jalal, H. Einsele, L. Klingspor, P. Nyren, Mycologia (2002) submitted.
[24] M. Ronaghi, Genome Res. 11 (2001) 3.