Forensic Science International: Genetics xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Next generation sequencing and its applications in forensic genetics

Claus Børsting *, Niels Morling

Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 24 October 2014
Received in revised form 12 January 2015
Accepted 11 February 2015

Keywords:
Next generation sequencing
Single-molecule sequencing
Forensic genetics
Review

ABSTRACT

It has been almost a decade since the first next generation sequencing (NGS) technologies emerged and quickly changed the way genetic research is conducted. Today, full genomes are mapped and published almost weekly and with ever increasing speed and decreasing costs. NGS methods and platforms have matured during the last 10 years, and the quality of the sequences has reached a level where NGS is used in clinical diagnostics of humans. Forensic genetic laboratories have also explored NGS technologies and especially in the last year, there has been a small explosion in the number of scientific articles and presentations at conferences with forensic aspects of NGS. These contributions have demonstrated that NGS offers new possibilities for forensic genetic case work. More information may be obtained from unique samples in a single experiment by analyzing combinations of markers (STRs, SNPs, insertion/deletions, mRNA) that cannot be analyzed simultaneously with the standard PCR-CE methods used today. The true variation in core forensic STR loci has been uncovered, and previously unknown STR alleles have been discovered. The detailed sequence information may aid mixture interpretation and will increase the statistical weight of the evidence. In this review, we will give an introduction to NGS and single-molecule sequencing, and we will discuss the possible applications of NGS in forensic genetics.

1. Introduction

DNA sequencing has a long history in forensic genetics. In the late 1980's and early 1990's, sequencing of mitochondrial DNA (mtDNA) was evaluated and used for case work at a time when restriction fragment length polymorphism (RFLP) analysis was the state of the art for human identification and years before the first short tandem repeat (STR) assays were developed. Successful RFLP analysis required micrograms of preferably intact DNA and that made the sensitive PCR-based mtDNA sequencing method the preferred tool for typing of low amounts of degraded sample materials, e.g., hair shafts and old bones [1–3]. Sequencing of the mtDNA control region was used extensively and the European DNA Profiling (EDNAP) Group's mitochondrial DNA population database project (EMPOP) was initiated in 1999 with the purpose of creating a common forensic standard for mtDNA sequencing and an on-line mtDNA database with high quality mtDNA population data [4,5]. Laboratories that qualified by successful participation in EMPOP collaborative exercises submitted mtDNA sequences into EMPOP and with release 11 (October 2013), the EMPOP database contained 34,617 mtDNA sequences from populations all over the world.

Sequencing was conducted with the Sanger dideoxynucleotide (ddNTP) chain terminating method [6], where the incorporation of a ddNTP to a growing DNA chain prevented further extension by the DNA polymerase (Fig. 1A). Early on, the synthesized DNA fragments were separated by slab gel electrophoresis and detected by either radioactively or fluorescently labeled deoxynucleotides (dNTPs) incorporated into the DNA fragments. Subsequent introduction of fluorescently labeled ddNTPs and capillary electrophoresis (CE) platforms [7] increased sensitivity and throughput, and decreased the cost of Sanger sequencing to a level where sequencing of complete genomes became possible. The improvements in CE technology and the development of highly sensitive PCR-based STR assays gradually reduced the need for mtDNA sequencing in forensic genetics during the 1990's. However, the Sanger sequencing method was used continuously for verification and identification of, e.g., STR alleles (see references in STRbase, http://www.cstl.nist.gov/strbase/). The ddNTP chain terminating method was also used for the so-called mini-sequencing or single base extension (SBE) reaction that was used for typing of single nucleotide polymorphisms (SNPs) [8]. SBE was a post-PCR cyclic reaction where SBE primers hybridized to the PCR products and were extended with a labeled ddNTP complimentary to the nucleotide in the SNP position (Fig. 1B). The SBE products

* Corresponding author at: Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Frederik V's Vej 11, DK-2100, Copenhagen, Denmark. Tel.: +45 3532 6225; fax: +45 3532 62891.
E-mail addresses: claus.boersting@sund.ku.dk, http://retsmedicin.ku.dk (C. Børsting).

http://dx.doi.org/10.1016/j.fsigen.2015.02.002
1872-4973/

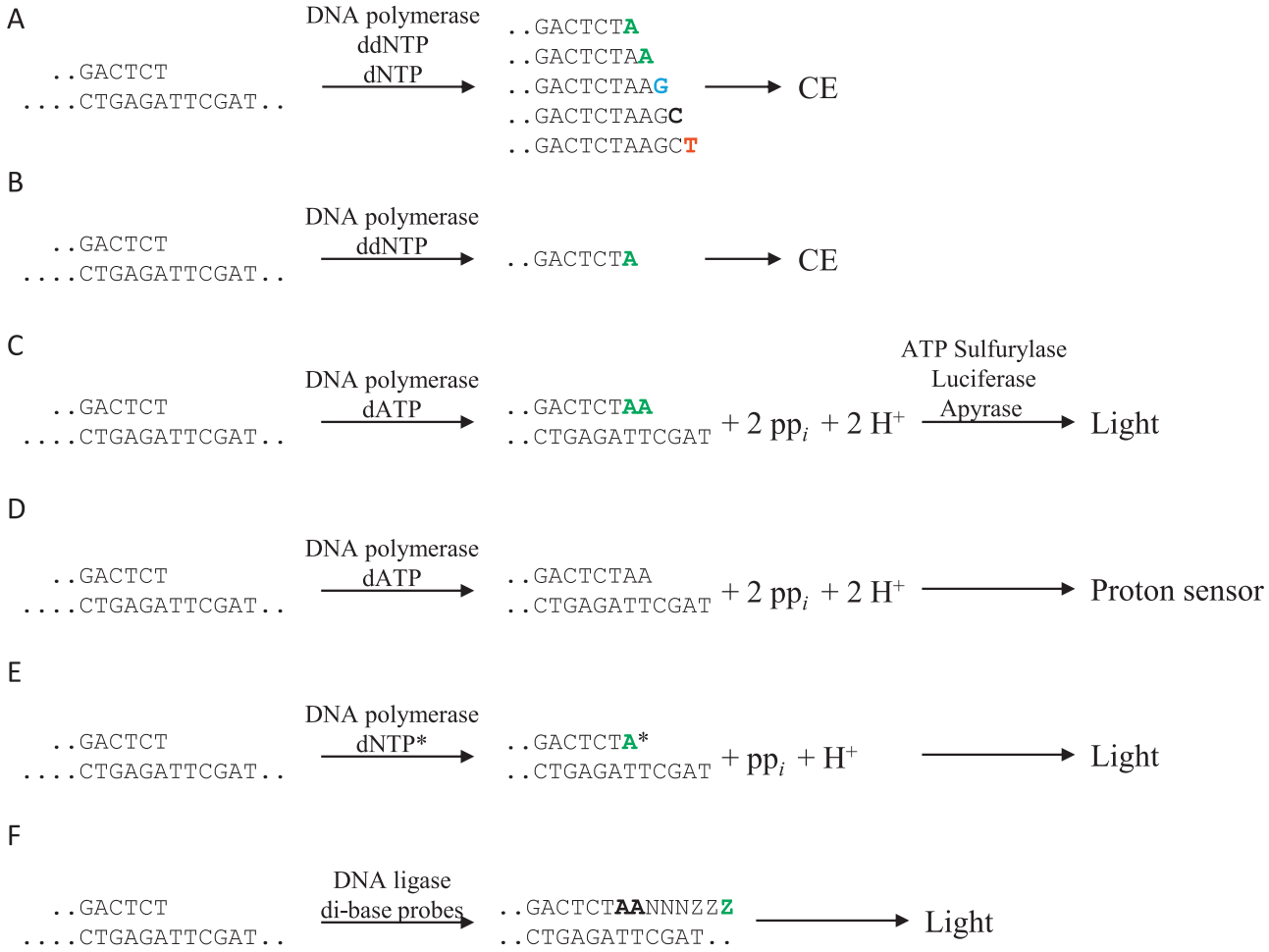

Fig. 1. Sequencing methods. (A) Sanger sequencing. DNA is synthesized in the presence of fluorescently labeled ddNTPs. The differently sized fragments are separated by CE and the sequence of fluorescently labeled nucleotides is detected by a camera. (B) Single base extention. The SBE primers are extended with a fluorescently labeled ddNTP complimentary to the nucleotide in the SNP locus. The extended SBE primers are detected by CE. (C) Pyrosequencing. Nucleotides are added sequentially to the sequencing reaction. Incorporation of one or more nucleotide(s) to the growing strand release one or more pyrophosphate(s) that are used in secondary enzymatic reactions to generate light. The light emission is detected by a camera. (D) Semi-conductor sequencing. Nucleotides are added sequentially to the sequencing reaction. Incorporation of one or more nucleotide(s) to the growing strand release one or more hydrogen ion(s) that are detected by an ion sensor. (E) Sequencing by synthesis. DNA synthesis is performed with fluorescently labeled dNTPs with reversible 3′ terminators (marked by an asterisk). Each addition of a nucleotide to the growing strand is detected by a camera. The terminator is chemically removed allowing for the next nucleotide to be incorporated. (F) Sequencing by ligation. The sequencing primer is hybridized to the target DNA and four sets of four fluorescently labeled di-base probes (all the 16 possible combinations) are added sequentially to the ligase reaction. Successful ligation of a probe to the sequencing primer is detected by a camera. The probes are cleaved (between the N and Z nucleotides) and another cycle of ligations can begin.

were detected by capillary electrophoresis, where the length of the extended SBE primer identified the SNP locus, and the ddNTP label identified the SNP allele. Panels of SNPs for human identification, pigmentary traits and ancestry information were identified, and SBE assays were validated and used in actual case work [9–14].

Pyrosequencing was presented as a real-time sequencing alternative to Sanger sequencing in 1996 [15]. Nucleotides were added sequentially to the DNA synthesis reaction, and the released pyrophosphate was used to generate light via a cascade of enzymatic reactions involving the three enzymes; ATP Sulfurylase, Luciferase and Apyrase (Fig. 1C). The light was detected in real-time by a CCD camera and thus, electrophoresis of the sequencing products was not necessary. Pyrosequencing was cheap and fast compared to Sanger sequencing, and the method was applied to mtDNA sequencing [16,17] and later also used for STR sequencing [18]. However, the short sequencing length and especially the limited multiplexing capability of the instruments were not compatible with the low amounts of DNA usually recovered from trace samples, and the method was never used in case work.

Even though the first pyrosequencing instruments never found a strong foothold in science, the pyrosequencing technology itself and the idea of real-time sequencing became the foundation on which the ongoing revolution in DNA sequencing was made. The first commercial high throughput sequencing platform, the Genome Sequencer 20 from 454 Life Sciences, used pyrosequencing [19], and it was possible to sequence the human genome in five months at a cost of $1.5 million with this technology [20]. In comparison, the first human genome was sequenced with Sanger sequencing technology during a period of 13 years and a cost of $2,700 million [21]. Several high throughput sequencing methods and platforms have since then been introduced. Most of them have been acquired by larger companies and sometimes the instruments have changed names, e.g., Solexa was changed to Illumina. Some have come and gone again, e.g., the HeliScope platform from Helicos BioSciences [22,23], and Roche has recently announced that the production of the highly successful 454 pyrosequencers will be terminated in 2015. Early on, these platforms were usually referred to as next generation sequencing or massively parallel sequencing platforms. However, with the introduction of single-molecule sequencing, some platforms were referred to as second generation sequencers and the single-molecule sequencer sometimes referred to as third generation sequencers or the next–next

Please cite this article in press as: C. Børsting, N. Morling, Next generation sequencing and its applications in forensic genetics, Forensic Sci. Int. Genet. (2015), http://dx.doi.org/10.1016/j.fsigen.2015.02.002