International Journal of Agricultural Sciences ISSN 2167-0447 Vol. 11 (2), pp. 001-004, February, 2021. Available online at www.internationalscholarsjournals.org © International Scholars Journals

Review

Advances in DNA sequencing: Challenges and limitations of personal sequencing

K. Mehla[1], S. Chaudhary[2], A. Kumar[3], V. Kumar[3], P. Chauhan[4], S. Gupta[5], J. Singh[6], P. Kumar[7], V. Kumar[8], N. Kumar[3], A. Jindal[3], S.Kumar[3], V. Sharma[3], S. Chand[3], N. Mahajan[3], A. Singh[3], B. Ramesh[9]and D. Singh[3]*

[1]Institute of Genomics and Integrated Biology, Mall Road, New Delhi, India
[2]Banasthali Vidyapith, Rajsthan, India
[3]Molecular Biology Laboratory, Department of Genetics and Plant Breeding, SVP University of Agriculture and Technology, Meerut -250110, India.
[4]Department of Cell Biology, SVP University of Agriculture and Technology, Meerut -250110, India.
[5]Department of Molecular Biology and Genetic Engineering, SVP University of Agriculture and Technology, Meerut -250110, India.
[6]Department of Immunology and Defense Mechanism, SVP University of Agriculture and Technology, Meerut -250110, India.
[7]Department of Physiology and Biochemistry, SVP University of Agriculture and Technology, Meerut -250110, India.
[8]Department of Ag. Biotechnology, AAU, Anand, Gujarat, India.
[9]Department of Genetics and Plant Breeding, CCS University, Meerut, UP, India.

Accepted 04 January, 2021

The advent of second and third generation DNA sequencing techniqnologies have revolutionized the genomics research. The Single Molecule Sequencing technologies are adequate to minimize occurrence of errors in sequencing. The third generation sequencing technologies could overcome limitations of first and second-generation sequencing technologies. The almost complete human genome sequencing was performed by using single molecular sequencing technologies. Benefits, risks and safeguards associated with the disclosure of information of very intimate kind contained in human genome, were also discussed. The power of whole sequencing for human welfare arises only from associations with medical histories, behavioral characteristics, physical descriptions and genome-environment interactions not from mere knowledge of base pairs.

Key words: DNA sequencing, medical histories, behavioral characteristics, physical descriptions.

INTRODUCTION

Frederick Sanger's method of DNA sequencing has revolutionized the genomics research. Development of high throughput DNA sequencing protocols and advanced computational analysis methods made sequencing a routine procedure. Presently, commercially popular non-Sanger ultra-high-throughput second-generation sequencing technologies became available in 2007. Another non-Sanger much faster and cheaper single molecular sequencing (SMS), the third generation sequencing technologies became available in 2008 for commercial purposes and that could overcome limitations of first and second-generation sequencing technologies. The almost complete human genome sequencing was performed by using Single Molecular Sequencing technologies. Benefits, risks and safeguards associated with the disclosure of information of very intimate kind contained in human genome, were also discussed.

During the last more than three decades since Frederick Sanger developed dideoxy termination method of DNA sequencing, sequencing technologies received major advancements. Frederick Sanger' method, synergistically involving enzymology and chemistry has transformed DNA sequencing in few years into a routine and so simple

*Corresponding author. E-mail: devisingh11@gmail.com. Tel: +1212888541(O), +919358918146. Fax: +1212888505.

technique that genomes of a variety of the important plants and animals including human has been completely sequenced. Since the onset of genomics research in the mid-1990s, polymorphisms at the DNA level assumed importance and could be studied by numerous approaches. Certainly, the most direct strategy is the determination of the nucleotide sequence of a defined region (Maxam and Gilbert, 1977; Sanger et al., 1997), and the alignment of this sequence to a corresponding region in the genome of another related organism. Advent of the PCR speeds up progress in researches towards DNA sequencing (Mullis et al., 1994; Saiki et al., 1988). Consequently, it became possible to isolate DNA regions of interest from DNA of other organism with enhanced speed. Universal primer pairs were designed on the basis of sequence information for conserved parts of the DNA, and the PCR-amplified target regions were either sequenced directly or sequenced after cloning (Hillis et al., 1988). The popularity of DNA sequencing was further enhanced by the development of fluorescence-labeled primers and nucleotides that could be used for the automated detection of DNA molecules in gel- or capillary-based sequencing instruments (Smith et al., 1996). With readings of up to 1200 base pairs, fluorescence sequencing provided much higher resolution than the traditional approach using radioisotopes. Moreover, it became easier to perform sequencing by transferring the data directly to a computer. The fact, that the technical equipment is more expensive than traditional sequencing facilities is not a real problem because custom sequencing services have become commercialized at cheaper rate. The extent of homology between various sequences can be deduced from the alignment, and phylogenies can be reconstructed by a variety of approaches (Felsenstein, 2004; Hall, 2001; Huelsenbeck and Crandall, 1997; Huelsenbeck et al., 2001; Page and Holmes, 1998; Swofford et al., 1996). DNA sequencing provides highly robust, reproducible, and informative data sets, and can be adapted to different levels of discriminatory potential by choosing appropriate genomic target regions. However, certain limitations are observed. Firstly, DNA sequencing is tedious and expensive when very large numbers of individuals are required to be assayed. Secondly, the highly specific sampling at least for certain areas of research is used which could represent only a small part of the genome. For example, phylogeny reconstructions based on DNA sequence data generally result in gene trees, which do not necessarily reflect the species tree. Over the past decade, with the development of high throughput DNA sequencing protocols and advanced computational analysis methods, it has been possible to generate assemblies of sequences encompassing the majority of the human genome (Shendure et al., 2004; Chen, 2005). To meet the increased sequencing demands, several non-Sanger ultra-high-throughput sequencing systems became commercially available in 2007.

The commercially popular non-Sanger ultra-high-throughput sequencing systems such as “Genome Analyzer”, “Genome Sequencer 20/FLX’ ‘SOLiDTM system” etc. were described as second generation or next generation sequencing systems. Two versions of the human genome were developed by Human Genome Sequencing Consortium (Chen, 2005) and Celera Genomics (Shendure et al., 2004; Gupta, 2008) by using the second generation or next generation sequencing systems. The versions were derived from clone-based and random whole genome shotgun sequencing strategies, respectively. The Human Genome Sequencing Consortium assembly is a composite derived from haploids of numerous donors, whereas the Celera version of the genome is a consensus sequence derived from five individuals. Both versions almost exclusively report DNA variation in the form of single nucleotide polymorphisms (SNPs). However smaller-scale (100 bp) insertion/deletion sequences or large-scale structural variants (Mitchelson, 2007; Harding and Keller, 1992) also contribute to human biology and disease (Brakmann et al., 2002; Crut et al., 2005). It, therefore, warrants a comprehensive analyses and review applications of such technologies. Keller Harding (1992) proposed single molecule sequence system (SMS) in 1989 that could be realized in the laboratory through several approaches, such as scanning probe microscopy, exonuclease sequencing, and sequencing by synthesis (Brakmann et al., 2002; Bayley, 2006). It is another non-Sanger DNA sequencing approach which became available in 2008 for commercial purposes. This approach has been described as a third generation or next-next generation sequencing technology. It has been observed that SMS is much faster and cheaper technology. The following limitations/problems were observed while working with second generation sequencing systems (Jett et al., 1989; Harding and Keller, 1992). The amplification of the target DNA by polymerase chain reaction (PCR) creates several problems such as introduction of a bias in template representation, and the introduction of errors during amplification. The second major problem was ‘‘phasing’’ of the DNA strands introduced by (Gupta, 2009) diploid nature of human genome (determining on which of the two chromosomes a variant is located). Two variants that lie in a gene on the same chromosome ‘‘phased variants’’ are not the same as two variants (that is, alleles) located in genes on separate chromosomes. These limitations could be largely overcome by SMS technique. However, efficiency of SMS is still debatable and warrants continuous efforts to scrutinize and improve these technologies to solve the common major problems of second generation or/and third generation sequencing technologies such as weak automation, short read-lengths, higher error-rates, and data management. Major emphasis for improvement of SMS has recently been given due to emerging demands for personal sequence.

The SMS technologies for the first time were applied to sequence a human genome of Stanford professor Stephen Quake. Despite the fact, professor Quake knew the risks in publishing identifiable personal information of the most intimate kind. He could dare to tell the world that it was his DNA that had been sequenced. Also, to date, only a few personal genomes have been fully sequenced. The genome sequences of few people like Craig Venter and James Watson have been published (Levy et al., 2007; Wheeler et al., 2008). In view of the benefits from open sharing of genomic data, urgency of the need to circumscribe the use of personal genomic information has been emphasized. It is anticipated that proper intelligence and knowledge will counteract the inefficiencies of healthcare systems that still spend most of their money treating patients who are near to death. The promise of genomic medicine is preventative and predictive healthcare. Nevertheless, genomic medicine only works if the research community has databases combining genomic data and information about corresponding phenotypic expression. The mere knowledge of the base sequence in a human genome is hardly of little value. Its power for human welfare arises only from associations with medical histories, behavioral characteristics, physical descriptions and genome-environment interactions. On the other hand, consequences of putting the personal genome into public domain seem to have capacity to promote all social illnesses already prevalent in our societies such as discrimination, defamation, sexual orientation and socially defined 'race' discrimination etc. Consequently, it may reflect its impact to deterioration/ restructuring human society with respect to physical and mental health, aptitude or suitability for athleticism or employment, and eligibility for health, disability and life insurance. Also, it may result into unduly tenacious adherence to beliefs in genetic determinism that might prompt discomfort and/or harassment, with personal genetic information being available over the internet or in any widely available medium, that is, print media etc.

Governments have introduced legislation in order to reduce such risks of disclosure of personal information of very intimate kind. The US has the Genetic Information Non discrimination Act (GINA). The EU has its Directive 95/46/EC on data protection. The UK amended its 2004 Human Tissue Act to make it illegal to sample someone's DNA without their consent. Whereas Germany passed legislation in April specifically prohibiting anonymous paternity tests and outlawing genetic testing for predisposition to illnesses of later life. One step ahead, apex court of India passed an order not to disclose or interfere in privacy of a individual and therefore, issued directives to Indian Government to stop even tests such as Norco tests which are conducted by investigating agencies on criminals to know information of intimate kind. Such measures may actually reduce the risk but they also increase the perception of risk: people worry more about disclosing their genomes.

Such considerations made it clear that other side of the picture is not as bright as was anticipated. Ten years after former president Bill Clinton announced that the first draft of the human genome was complete, medicine sector has yet to see any large part of the promised benefits (Anonymous, 2010). Paynter et al. (2010) reported that old fashioned method of taking a family history was a bettor guide. One area of potential improvement has been the discovery of genetic markers for cardiovascular disease as well as intermediate phenotypes such as cholesterol and blood pressure.

Recent efforts using genome-wide association studies have greatly expanded the discovery of genetic markers associated with cardiovascular disease. To date, however, the utility of single genetic markers based on genome analysis, to improve cardiovascular risk prediction has shown mixed results, even for the most promising marker, located in the 9p21 region (Talmud et al., 2008; Brautbar et al., 2009). A genetic risk score comprising 101 single nucleotide polymorphisms was not significantly associated with the incidence of total cardiovascular disease (Paynter et al., 2010).). To combine the relatively small effects of individual genes and to better capture the complex relationship between genetics and cardiovascular disease, the use of amultilocus genetic risk score has been proposed (Morrison et al., 2007). One such score developed by Kathiresan et al. (2008) including 9 genetic markers associated with increased lipid levels but showed no improvement in discrimination and, only a slight improvement in reclassification was observed. In large part, however, the predictive abilities of recently discovered genetic markers have not been tested 6 for biologists, the genome has yielded one insightful surprise after another. But the primary goal of the 3 Billion US Dollars. Human Genome Project- to ferret out the genetic roots of common fatal diseases like cancer and Alzheimer’s and then generate treatments-remains largely elusive. Indeed, after10 years of effort, geneticists are almost back to square one in knowing where to look for the roots of the common diseases.

REFERENCES

Anonymous (2010).Decade on, gene map just a promise. The Times of India, New Delhi, Monday, June, pp 14-15.

Bayley H (2006). Sequencing single molecules of DNA. Curr. Opin. Chem. Biol., 10: 628-637.

Brakmann S, Löbermann S (2002). A further step towards single-molecule sequencing: Escherichia coli exonuclease III degrades DNA that is fluorescently labeled at each base pair. Angew. Chem. Int. Ed. Engl., 41: 3215-3217.

Brautbar A, Ballantyne CM, Lawson K (2009). Impact of adding a single allele in the 9p21 locus to traditional risk factors on reclassification of coronary heart disease risk and implications for lipid-modifying therapy in the Atherosclerosis Risk in Communities (ARIC) study. Circ. Cardiovasc Genet., 2(3): 279-285.

Chan EY (2005). Advances in sequencing technology. Mutant. Res., 573: 13-40.

Crut A, Géron-Landre B, Bonnet I, Bonneau S, Desbiolles P, Escudé C (2005). Detection of single DNA molecules by multicolor quantum-dot end-labeling. Nucleic Acids Res., 33: e98.
Felsenstein J (2004). Inferring Phylogenies, Sinauer Associates, Sunderland, MA.
Gupta PK (2008). Ultra fast and low cost sequencing methods for applied genomics research. Proc. Natl. Acad. Sci. India, 78: 91-102.
Gupta PK (2009).Single-molecule DNA sequencing technologies for future genomics research. Trends Biotechnol., 26(11): 602-611.
Hall BG (2001). Phylogenetic Trees Made Easy, Sinauer Associates, Sunderland, MA.
Harding JD, Keller RA (1992). Single-molecule detection as an approach to rapid DNA sequencing. Trends Biotechnol. 10: 55–57.
Hillis DM, Moritz C, Mable BK (1996). Molecular Systematics, 2nd Edition. Sinauer Associates, Sunderland, MA.
Huelsenbeck JP, Crandall KA (1997). Phylogeny estimation and hypothesis testing using maximum likelihood, Annu Rev. Ecol. Syst., 28: 437-466.
Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. Science, 294: 2310-2314.
Jett JH, Keller RA, Martin JC, Marrone BL, Moyzis RK, Ratliff RL, Seitzinger NK, Shera EB, Stewart CC (1989). High-speed DNA sequencing: an approach based upon fluorescence detection of single molecules. J. Biomol. Struct. Dyn., 7: 301-309.
Kathiresan S, Melander O, Anevski D (2008). Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J. Med., 358(12): 1240-1249.
Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness E F, Denisov G, Lin Y, MacDonald JR, Pang AWC, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz S A, Busam DA, Beeson KY, McIntosh T C, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer S W, Strausberg RL, Venter JC (2007). The diploid genome sequence of an individual human. PloS Biol., 5: e254.
Maxam AM, Gilbert W (1977). A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A., 74: 560-564.
Mitchelson K (2007). New High Throughput Technologies for DNA Sequencing and Genomics (Vol. 2), Elsevier.
Morrison AC, Bare LA, Chambless LE (2007). Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. Am. J. Epidemiol., 166(1): 28-35.
Mullis KB, Ferré F, Gibbs RA (1994). The Polymerase Chain Reaction,Birkhäuser, Basel, Switzerland.
Page RDM, Holmes EC (1998). Molecular Evolution: A Phylogenetic Approach, Blackwell Science Ltd., Oxford, U.K.
Paynter, Nina P, Chasman, Daniel I, Paré, Guillaume, Buring, Julie E, Cook, Nancy R, Miletich, Joseph P, Ridker, Paul M (2010).Association between a literature-based genetic risk score and cardiovascular events in women. JAMA, 303(7): 631-637.
Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA, (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science, 239: 487-491.
Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chainterminating inhibitors, Proc. Natl. Acad. Sci. U.S.A., 74: 5463-5467.
Shendure J, Mitra RD, Varma C , Church GM (2004). Advanced sequencing technologies. Nat. Rev. Genet., 5: 335-344.
Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SBH, Hood LE, (1986). Fluorescence detection in automated DNA sequence analysis. Nature, 321: 674-679.
Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996). Phylogenetic inference. In Molecular Systematics, 2nd Edition, Hillis, D.M., Moritz, C., and Mable, B.K., Eds., Sinauer Associates, Sunderland, MA, pp. 407-514.
Talmud PJ, Cooper JA, Palmen J (2008). Chromosome 9p21.3 coronary heart disease locus genotype and prospective risk of CHD in healthy middle-aged men. Clin. Chem., 54(3): 467-474.
Wheeler DA (2008). The complete genome of an individual by massively parallel DNA sequencing. Nature, 452: 872-876.