Advances in Biotechnology

Chapter 4

Advances in Biotechnology in the Post Genomics era

Amjad Ali*; Hamza Arshad Dar[1]; Tahreem Zaheer[1]

[1]Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences & Technology (NUST), H-12, Islamabad, Pakistan 44000.

****Corresponding to: Amjad Ali,** Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences & Technology (NUST), H-12, Islamabad, Pakistan 44000.*

Email: amjaduni@gmail.com

1. Introduction

Pregenomic era comprised of efforts to sequence genome and now in the post genomic era where we have greater than 1000 genomes available, science is heading toward extracting valuable information from them. Sequencing has helped in revealing the hidden meaning of nucleotide and protein sequencing. Shifting from the trends of the pregenomic era to post genomic era resulted in enormous data. In this chapter, we have explored the impact of advancements in genomics on organisms ranging from viruses to plants with focus on their applications in Biotechnology. In particular, we have discussed the influence of rapidly available sequencing data in exploiting the viruses for our benefit, especially in vaccine development. In this regard, some Bioinformatics-based tools and software have been discussed. The Human Genome Project and its importance as an example and a motivation for other similar organism-specific large-scale sequencing projects has been highlighted. Finally, some aspects related to genomics-based Biotechnological aspects of plant sciences had been explored. We conclude that recent progress in genomics has brought about major breakthroughs in terms of applications of Biotechnology in different sectors such as vaccinology, proteomics, personalized medicine, as seen in **Figure 1**.

The journey began in 1976, when RNA of E.coli infecting bacteriophage, MS2, was sequenced completely [1]. Following this discovery, a DNA containing bacteriophage, PhiX174, was sequenced by Sanger and his team [2]. It was the first DNA based genome that was se-

quenced. PhiX174 was later used as a model organism in the ushering era of synthetic biology [3]. Sanger shot gun sequencing provided a platform to sequence genome with greater ease, but cost was a major constraint of this technique. In 1981, Cauliflower mosaic virus was sequenced and variation within the different strains were analyzed by using comparative genomics [4]. In 2004, complete genome sequencing of mimivirus blurred the distinction between bacteria and viruses [5]. Unlike bacteria, viruses do not contain rDNA to study phylogenetic relationship, so a clone based sequencing strategy was used to sequence and classify unculturable marine viruses [6]. The sequencing of these marine viruses gave insight into their role in biogeochemical cycles [7].

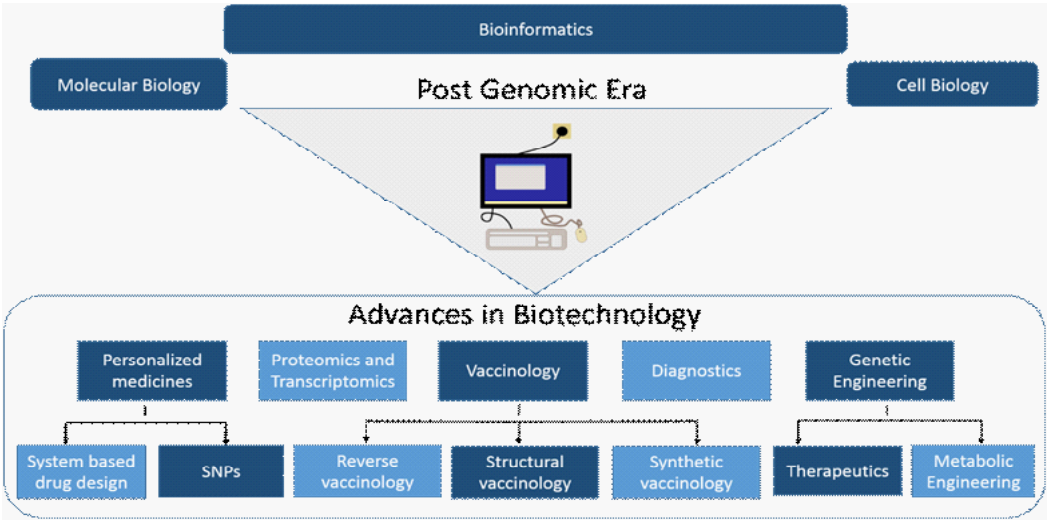

Figure 1: Advancements of Biotechnology in the postgenomic era in different sectors is illustrated.

2. Impact of Genomics in Virology

The development of Next Generation sequencing has brought about a revolution in the field of virology. Viral genomes, though rather small size, maintain their intellectual curiosity amongst scientists [8]. The emergence of pandemic viral infections such as H5N1 and H1N1 also necessitated the availability of whole genome sequence to gain an insight into the evolution and molecular epidemiology of these viruses [9]. This was particularly true since earlier phylogenetic analysis based on partial sequence had failed to comprehend the complex historical recombination events that are potentially responsible for pandemic emergence. NGS along with partitioning and barcoding has enabled the efficient sequencing of complete viral genomes leading to better understanding of the transmission and emergence of clinically important viruses [8].

The opportunity to sequence and compare multiple whole genomes has highlighted the crucial genetic differences between different viral isolates [9]. The current knowledge about sequencing has enabled researchers to analyze drug resistance in DNA and RNA viruses (Cytomegalovirus and Haemophilus influenza virus). High coverage sequencing (also termed as deep sequencing) helped to identify lesser drug resistant variants. However, whole genome sequencing of viruses can help us in understanding of better and potential drug resistant vari-

www.openaccessebooks.com

ants. Other than research purposes, sequencing analysis is equally important in clinical studies. For instance, highly active antiretroviral therapy in case of HIV has significantly improved the survival rate of HIV patients [10]. Apart from this, metagenomics analysis is also extensively used as a diagnostic tool. Herpes simplex virus was identified in the cerebrospinal fluid (CSF) of patients who were suspected to have viral meningoencephalitis. Pan viral screening is believed to aid in diagnostics of Central nervous system infections [11]. However, there is a need to develop a more rapid sequencing technology to share real time sequence information to guide healthcare sector for the control of outbreaks [8].

The rapid growth of viral genome sequences and their Bioinformatics analysis has brought about a revolution in viral genomics. The development has challenged the conventional classification and nomenclature of these organisms [12]. Genomics and Bioinformatics-based software and tools need to be developed to utilize the genome attributes such as phylogenomics and unique features in the strain's biology and also about the viral families. Therefore, the information derived from primary sequence data can be useful compared to the previous use of immunochemical methods that probed limited and often murky epitopes that are actually an indirect interpretation of the primary sequence data in the form of a tertiary sequence.

Viral sequencing data is being used in Forensic studies. Sexually transmitted viruses such as HIV (Human Immunodeficiency virus)were used to generate phylogenetic profiles of disease and link victim and assailant [13]. Some viruses such as HCV (Hepatitis C virus) [14], EBV (Epstein Bar virus) [15], and BKV (BK virus) [16] can prove to be significant in determining place of birth and locality of suspicious individuals.

Advances in Bioinformatics has enabled scientists to acquire a better understanding of the biology of pathogenic viruses. For example, viruses belonging to the Poxviridae family infect a variety of hosts and cause small pox disease in humans. Moreover, their natural occurrence and potential bioterrorism concerns has aroused an interest in the scientific community [12]. Ebola virus is also suspected to be a bioweapon [17]. A collection of genomes through recent advancements in genome sequencing has permitted the understanding of core genes (orthologous genes) that are present in all the members of the Poxviridae family. Faced with the challenges of analyzing simple and smaller genomes of viruses, a poxvirus-specific computational tool was developed by Hendrickson et al. to predict accurate gene sets [18]. This comparative approach highlighted the concept of reductive evolution in which loss of particular genes is thought to play an essential role in the speciation and restriction of emergent viruses to operate in particular environments. Eaton et al., explored the idea of core genes in the Iridoviridae family [19]. They concluded that genomes contain groups of repetitive sequences.A similar study was conducted in Nucleo-Cytoplasmic Large DNA Viruses (NCLDV) and orthologous genes were determined in 6 families using Comparative phylogenetics [20]. Thus, in the postgenomic era, numerous Bioinformatics tools have been developed for comparative

genome analyses which of course was dependent on the availability of genome sequences.

In pregenomic era Edward Jenner used a cow pox virus to induce immunity against smallpox viruses in the human, but understanding of mechanism of vaccines was limited at the time [21]. On the basis of further innovation and advancements in the field of Vaccinology, vaccines were categorized into first generation vaccines (having inactivated/killed lysate of pathogens), second generation vaccines (pure antigenic determinants of pathogen)and third generation vaccines or modern Vaccinology (that use genomics, transcriptomics and genome analysis to construct vaccine candidates) [22]. The approaches used in classical Vaccinology (1^{st} generation and 2^{nd} generation vaccines) were unable to fully overcome infections due to the diversity and complexity of microbial genomes. Poorly activated pathogen lysates may cause adverse effects, so there was a need to introduce novel strategy to develop universally applicable and safe vaccines.

3. Reverse Vaccinology

With the accessibility of complete genomic data of pathogenic microorganisms, an innovative approach known as “reverse vaccinology” has been designed for vaccine development. Computer-aided analyses can be conducted utilizing the genome sequence of a particular pathogen to predict the antigenic components for the development of a potential vaccine [23]. The advantages are multi-fold. There exists no requirement to grow and cultivate the microorganism. The entire procedure is done using computers without the requirement of laboratory apparatus such as pipettes, fermenters and so on. Pathogens requiring strict handling can be studied without any safety concerns. The framework takes into consideration all the proteins that are expressed (*invivo* or *invitro*) by a pathogen at a given time. Antigens used in conventional wet laboratory experiments are identified; moreover, novel antigens are discovered based on a completely different framework. In case of viruses, the mutation rate is higher so reverse vaccinology approach can provide data regarding putative antigenic vaccines that are conserved across all the strains in viral species. In case of Dengue virus 9000 viral sequences were analyzed to determine potential vaccine constructs that can elicit immunity against nearly all the strains of dengue virus [24]. Similar studies are conducted in Zika virus [25,26], human papilloma virus [27], Congo virus etc. [28,29]. More than 9500 reference sequences of viral genomes are available on NCBI. The reverse vaccinology approach provides new and yet unexplored insights into the mechanisms of immune intervention.

This top down strategy of the post genomic era has reduced the time and cost required for making vaccines. However, testing these vaccines in rodents and then in mammals is required before clinical trials. In contrast to classical vaccinology era the labor-intensive efforts are reduced [30,31]. Now whole genomes can be analyzed and only antigenic immunogenic, non-homologous to human and surface exposed vaccine constructs can be designed that can

elicit immune response in human body without any risk of allergy or autoimmunity [32]. The approach is illustrated in **Figure 2**.

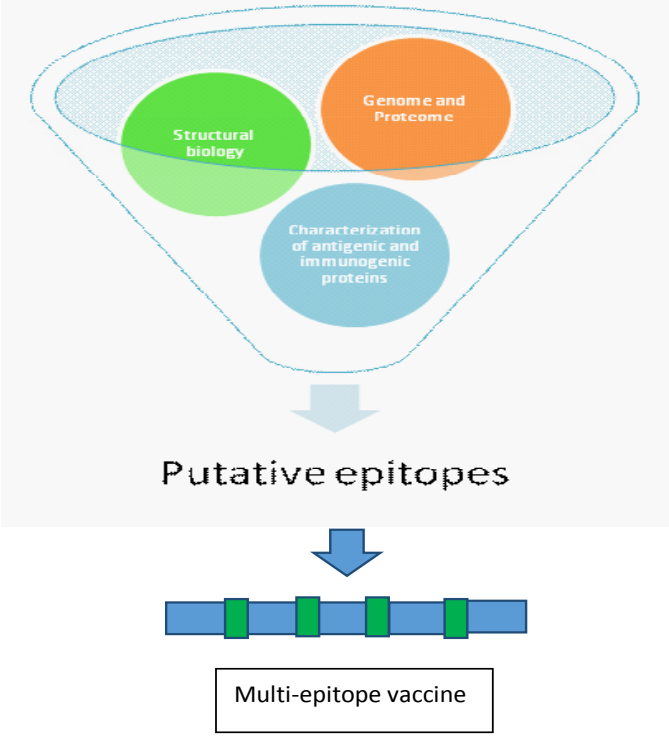

Figure 2: Reverse vaccinology. Genome and Proteome analysis can be used to predict epitopes. These epitopes are then characterized based on their antigenicity and immunogenicity [28]. Individual epitopes can be linked to make a multi-epitope vaccine [33].

The tools used in viral reversevaccinology and other in silico analysis are mentioned here-with:

Tool name	Use	Links
NetMHC(34,35) CTLPred(36) nHLAPred(37) Propred 1(38) Propred(39) RankPrep(40–42)	Prediction of MHC binding epitopes	http://www.cbs.dtu.dk/services/NetMHC/ http://crdd.osdd.net/raghava/ctlpred/ http://crdd.osdd.net/raghava/nhlapred/ http://crdd.osdd.net/raghava/propred1/ http://osddlinux.osdd.net/raghava/propred/ http://imed.med.ucm.es/Tools/rankpep
AVPpred(43)	Antiviral peptide prediction algorithm	http://crdd.osdd.net/servers/avppred/
HBVdb(44) GATU(45)	Annotation of viral genomes	http://hbvdb.ibcp.fr https://virology.uvic.ca/virology-ca-tools/gatu/
VaxiJen(46) ANTIGENpro (47)	Antigen prediction	http://www.jenner.ac.uk/VaxiJen
Allergen FP(48) ALGPred(49)	Allergenicity prediction	http://ddg-pharmfac.net/Allergen FP http://www.imtech.res.in/raghava/algpred/
SOLpro (50)	Solubility upon overexpression in E.coli	scratch.proteomics.ics.uci.edu/
siVirus(51)	Antiviral siRNA design software	http://sivirus.rnai.jp/
ViReMa(52)	Algorithm for detection of recombination junctions in viral Genomes	https://omictools.com/viral-recombination-mapper-tool

However, due to lack of adequate knowledge on aspects of immunological aspects of vaccine, good correlates of protective immunity are uncommon and is the major limitation of reverse vaccinology. Moreover, the approach is entirely protein-specific; other non-protein antigens such as polysaccharides and glycolipids are not covered in this method. Another drawback of reverse vaccinology is the genetic instability of some viruses. To circumvent this limitation, structure based reverse vaccinology and synthetic genomics can be applied for rational vaccine design. Structural vaccinology integrates data from structural biology, human immunology and bioinformatics to predict immunogenic and antigenic residues [53]. Crystal structure of Respiratory Synctial virus conjugated with fusion glycoprotein showed high neutralizing antibody titres using Structural vaccinology approaches [54]. However, in case of synthetic genomics, genomes can be artificially synthesized using genetic material. One of the recent examples of synthetic genomics was vaccine against avian influenza virus [55]. These vaccines can be manufactured rapidly and mimic natural viruses in their mode of action. Synthesized genomes and engineered antigens have improved the efficacy of vaccines, but understanding the pathogenesis of viruses is still of primary interest.

Some viruses have genome integration capabilities hence that they are actively used as viral vectors in gene therapy. These viral vectors are safe and effective [56]. Nevertheless, viral integration at certain sites may cause malignant transformation and altered gene expression. With the help of bioinformatics, a pipeline was recently designed to determine integration sites in NGS based viral vectors that could be used in gene therapy data. The tool is efficient and performs analysis by Agilent Sure Select through rapidly evolving targeted sequencing and PCR based linear amplification strategies. It is available at https://github.com/G100DKFZ/gene-is [57]. Some other tools that also determine viral integration sites are ViralFusionSeq [58] and Virus-Clip [59]. At the time of writing, GENE-IS is the first tool that gives information based on two sequencing strategies and has no specific constraints regarding input data.

4. The Human Genome Project and its Impact on Biotechnology

Work on the ENCODE (Encyclopedia of DNA Elements) project was made possible after the completion of the Human Genome Project [60]. The scientists working in the ENCODE project channelized their efforts to develop an understanding of the functional components of the human genome [61]. These efforts proved fruitful as they resulted in a huge amount of data regarding the regulatory networks that control the expression of human genes [62]. Computer aided pathway analysis has been used to locate protein and enzymes in their pathways and bioreactors, respectively. In 2005, computational analysis led to allocation of 622 enzymes in biological pathways and 2709 enzymes to bioreactors [63]. Nevertheless, more research is required to decipher the functions of low annotated human genes and large non-coding genomic regions that are transcribed [60].

The HGP has directly influenced advancements in the field of proteomics. Proteins as structural components, molecular machines, or signaling devices dictate the cell-specific functionality of the transcribed genome. The HGP has greatly aided the utilization of mass spectrometry, a crucial proteomics tool, by giving reference sequences and ultimately the predictions regarding the masses of all the tryptic peptides in the human proteome [64]. This is required for the mass-spectrometry based proteomics analysis. The Mass spectrometry (MS) has in turn been the driving force of novel applications like targeted proteomics [65]. Several servers like mascot [66], sequest [67], SQID [68] are used for the analysis of data obtained from MS. This data can also be used to identify Post translational modifications(PTMs) in proteins/peptide that may help in the understanding of the role in biological pathways; SIMS server is also available to identify PTMs in MS data [69].

The HGP has also contributed significantly to our understanding of evolution. The successful completion of this project jump-started the whole genome sequencing of other eukaryotic organisms and bacterial species [70]. The resulting collection of whole genome sequencing data from a variety of living organisms ranging from microbes to human has led to the genealogical tree of life that strongly supports the notion that all species that exist nowadays arise from a common ancestor (14,71). Especially, genome analysis of Neanderthal is likely to provide more insightful results into the evolutionary aspects of human beings especially [72].

The accessibility of all the diseases genes in human, along with genes from the human pathogens that are the causative agents of infectious diseases, will have a direct influence on drug development efforts. The human genome contains nearly 30,000 genes and it is expected that most, if not all of them, would be targets of therapeutic interventions. Functional and structural analyses of these genes and their encoded proteins respectively is likely to increase the number of drugs being developed in the coming years. Pharmaceutical sector is actively engaged to exploit the yet unexplored potential of recent advancements in genomics [73]. Due to complexity of biological system, system based drug discovery is also an effective approach to design drugs [74].

The variation in the human population can be analyzed by the power of genomics which will contribute to the science of medicine. DNA sequences are already in use for diagnostic purposes to identify the association of unique sequence variants or Single Nucleotide Polymorphisms (SNPs) with a particular disease. Distinct from point mutations, SNPs are sequence variants that are frequently found in the human population. These genetic variants do not in itself cause disease; rather they contribute to disease susceptibility in an additive manner. More than 10 million SNPs in human population were identified till 2011. This data was used to study the impact of SNPs on pharmacogenomics [75]. Moreover, these SNPs are also linked with complicated responses such as personalized responses to drug therapy. Hence, it may be possible to elucidate the variants that makes humans more prone to develop diseases such as

diabetes and asthma. Moreover, SNPs can be identified that influence individual response to drugs, thus ultimately increasing the likelihood of developing personalized therapies to target the unique genetic make-up of particular patients. SNPs are present in elite controllers of HIV and restrict the binding of virus with co-receptor CCR5 to block viral entry. The survival rate in elite controllers is comparatively higher than progressors [76]. However, the associated social, ethical, legal and moral issues need to be recognized and addressed to protect privacy and to prevent discrimination.

5. Post Genomics era in Plant Biotechnology

One of the many factors that limit crop production is salinity. Plants respond to saline stress in a complex way and the response is mediated by many genes which are the components of different signaling pathways in which cross-talk has also been reported [77]. Hence, it is difficult to fully understand how plants respond to salinity. Advancements in the field of genomics has provided the much-needed knowledge for crop improvement. Genes responsive against salinity induced stress have been identified and characterized, signaling pathways have been mapped, thus ultimately providing the basis for enhancing the salinity stress response of existing plants [78]. The information is crucial in the development of stress tolerant crops through tools like gene pyramiding that has been applied in marker assisted breeding and genetic engineering [79]. The advent of Genome editing by CRISPR/Cas9, TALENs, etc. has enabled plant biologists to produce desired genetically engineered crops with improved productivity, yield, etc. Recent progress in genomics has led to increased understanding of plant responses against environmental stresses such as salinity stress and drought conditions [78]. This has in turn increased prospects for generating stress tolerant plant varieties such as wheat, rice etc.

The genome of potato had been sequenced firstly using homozygous DM1-3 518 R44 or DM and later on with a heterozygous diploid line RH89-039-16 or RH [80-82]. The availability of the whole genome sequence as well as associated annotation of almost 39000 potato genes has enabled the identification of candidate genes in those regions that are concerned with specific traits [83]. Genome sequence assisted in the identification of StCDF1 gene that is responsible for plant maturity as well as StSP6A gene that is required for tuber initiation in potato [84,85]. The study of genome also generated a collection of candidate resistance genes, thus significantly improving our ability for robust discovery along with the prospects of introgressive hybridization of R-genes in potato [86,87]. The integrated approach of biotechnology and genomics is a positive step to solve global food security concern. Oleic acid cultivars were genetically modified to enhance vegetable oil production. More than 40% increase in consumption of this oil is expected to be achieved by 2020 in the US population [88].

6. Artificial Chromosomes

To incorporate larger segment of DNA, Yeast artificial chromosome (YAC) was introduced. The system proved helpful in studying genes with the normal promoter [89–94]. The advancement in scientific knowledge and Human genome project has led to the synthesis of BAC (Bacterial Artificial Chromosome) that are used for functional analysis of proteins [95]. MAC (Mammalian Artificial Chromosome) was constructed a year after generation of YAC. In 1997, Human artificial chromosome was introduced [96], refined in 2010, and was later used in inserting HSV(Herpes simplex virus) into cancer cells making them susceptible to ganciclovir antiviral drug. The virus infected cells were cleaved afterwards [97]. Post genomic era has provided us numerous opportunities to deeply understand antiviral mechanisms, expression profiling, and pathway construction using NGS and single cell sequencing.

7. Conclusion

Genome sequencing and associated huge amount of data has transformed the World of Biotechnology. Nowadays, sequencing cost has reduced considerably enabling robust whole genome sequencing of living organisms. This recent progress has triggered the development of different Bioinformatics tools and software to analyze the huge biological data. This has aided in the better characterization of different viruses and facilitated vaccine development using sequencing data in reverse vaccinology. Moreover, these analytical tools have facilitated drug development and gene therapy using viruses. The Human Genome Project has greatly facilitated the understanding of the human genome; variations in the human genome associated with particular disease were able to be identified and a better understanding of the human evolution has been achieved by comparative genomes and phylogenomics. The ENCODE project, in itself dependent on human genome, aims to elucidate the functions/s of the non-coding regions in the human genome. Advancements in genomics has led to the identification and characterization of genes contributing to tolerance against salinity stress and drought conditions in plants thus providing an opportunity to generate genetically modified crop varieties with improved resistance against these abiotic stress factors. Biotechnology along with genomics can also be used to solve global food crisis. Finally, Yeast Artificial Chromosome and Bacteria Artificial Chromosome can be used to incorporate large DNA fragments. Hence, further advancements in genomics will no doubt have a significant impact in shaping the Biotechnology of tomorrow.

8. References

1. Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, et al. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature [Internet]. 1976; 260(5551): 500–507.

2. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, et al. Nucleotide sequence of bacteriophage phi X174 DNA. Nature [Internet]. 1977; 265(5596): 687–95.

3. Goulian M, Kornberg A, Sinsheimer RL. Enzymatic synthesis of DNA, XXIV. Synthesis of infectious phage phi-

X174 DNA. Proc Natl Acad Sci U S A [Internet]. 1967; 58(6): 2321–2328.

4. Gardner RC, Howarth AJ, Hahn P, Brown-Luedi M, Shepherd RJ, Messing J. The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shotgun sequencing. Nucleic Acids Res [Internet]. 1981; 9(12): 2871–2888.

5. D'Elios MM, Appelmelk BJ, Amedei A, Bergman MP, Del Prete G. Gastric autoimmunity: the role of Helicobacter pylori and molecular mimicry. Trends Mol Med. 2004; 10(7): 316–323.

6. Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, et al. Genomic analysis of uncultured marine viral communities. Proc Natl Acad Sci U S A [Internet]. 2002; 99(22): 14250–14255.

7. Jover LF, Effler TC, Buchan A, Wilhelm SW, Weitz JS. The elemental composition of virus particles: implications for marine biogeochemical cycles. Nat Rev Microbiol [Internet]. 2014; 12(7): 519–28.

8. Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. Application of next-generation sequencing technologies in virology. J Gen Virol [Internet]. 2012; 93(Pt_9): 1853–1868.

9. Hoper D, Hoffmann B, Beer M. Simple, Sensitive, and Swift Sequencing of Complete H5N1 Avian Influenza Virus Genomes. J Clin Microbiol [Internet]. 2009; 47(3): 674–9.

10. Paredes R, Clotet B. Clinical management of HIV-1 resistance. Antiviral Res [Internet]. 2010; 85(1): 245–265.

11. Guan H, Shen A, Lv X, Yang X, Ren H, Zhao Y, et al. Detection of virus in CSF from the cases with meningoencephalitis by next-generation sequencing. J Neurovirol [Internet]. 2016; 22(2): 240–245.

12. Seto D. Viral genomics and bioinformatics. Viruses [Internet]. 2010 [cited 2017 Nov 7];2(12):2587–93.

13. Metzker ML, Mindell DP, Liu X-M, Ptak RG, Gibbs RA, Hillis DM. Molecular evidence of HIV-1 transmission in a criminal case. Proc Natl Acad Sci. 2002; 99(22): 14292–14297.

14. Kato H, Maeno Y, Seko-Nakamura Y, Monma-Ohtaki J, Sugiura S, Takahashi K, et al. Identification and phylogenetic analysis of hepatitis C virus in forensic blood samples obtained from injecting drug users. Forensic Sci Int. 2007; 168(1): 27–33.

15. Ikegaya H, Motani H, Sakurada K, Sato K, Akutsu T, Yoshino M. Forensic application of Epstein-Barr virus genotype: Correlation between viral genotype and geographical area. J Virol Methods. 2008; 147(1): 78–85.

16. Ikegaya H, Motani H, Saukko P, Sato K, Akutsu T, Sakurada K. BK virus genotype distribution offers information of tracing the geographical origins of unidentified cadaver. Forensic Sci Int. 2007; 173(1): 41–46.

17. Cenciarelli O, Gabbarini V, Pietropaoli S, Malizia A, Tamburrini A, Ludovici GM, et al. Viral bioterrorism: Learning the lesson of Ebola virus in West Africa 2013–2015. Virus Res [Internet]. 2015; 210: 318–326.

18. Hendrickson RC, Wang C, Hatcher EL, Lefkowitz EJ. Orthopoxvirus genome evolution: the role of gene loss. Viruses [Internet]. 2010; 2(9): 1933–1967.

19. Eaton HE, Metcalf J, Penny E, Tcherepanov V, Upton C, Brunetti CR. Comparative genomic analysis of the family Iridoviridae: re-annotating and defining the core set of iridovirus genes. Virol J. 2007; 4(1): 11.

20. Yutin N, Wolf YI, Raoult D, Koonin E V. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. Virol J. 2009; 6(1): 223.

21. Baxby D. Edward Jenner's inquiry after 200 years. BMJ [Internet]. 1999; 318(7180): 390.

22. Pulendran B, Ahmed R. Immunological mechanisms of vaccination. Nat Immunol [Internet]. 2011 Jun;12(6):509–17.

23. Rappuoli R. Reverse vaccinology, a genome-based approach to vaccine development. Vaccine. 2001; 19(17): 2688–2691.

24. Khan AM, Heiny AT, Lee KX, Srinivasan KN, Tan TW, August JT, et al. Large-scale analysis of antigenic diversity of T-cell epitopes in dengue virus. BMC Bioinformatics. 2006; 7(5): S4.

25. Dar H, Zaheer T, Rehman MT, Ali A, Javed A, Khan GA, et al. Prediction of promiscuous T-cell epitopes in the Zika virus polyprotein: An in silico approach. Asian Pac J Trop Med [Internet]. 2016; 9(9): 844–850.

26. Badawi MM, Osman MM, Alla AAF, Ahmedani AM, hamed Abdalla M, Gasemelseed MM, et al. Highly Conserved Epitopes of ZIKA Envelope Glycoprotein May Act as a Novel Peptide Vaccine with High Coverage: Immunoinformatics Approach. Am J Biomed Res. 2016; 4(3): 46–60.

27. Baidya S, Rasel Das M, Kabir G, Arifuzzaman M. Epitope design of L1 protein for vaccine production against Human Papilloma Virus types 16 and 18. Bioinformation. 2017; 13(3): 86.

28. Tipu HN. Immunoinformatic Analysis of Crimean Congo Hemorrhagic Fever Virus Glycoproteins and Epitope Prediction for Synthetic Peptide Vaccine. J Coll Physicians Surg Pakistan. 2016; 26(2): 108–112.

29. Shekhar C, Dev K, Verma SK, Kumar A. In-silico: screening and modeling of CTL binding epitopes of Crimean congo hemorrhagic fever virus. Trends Bioinform. 2012; 5: 14–24.

30. Soria-Guerra RE, Nieto-Gomez R, Govea-Alonso DO, Rosales-Mendoza S. An overview of bioinformatics tools for epitope prediction: Implications on vaccine development. J Biomed Inform. 2015; 53: 405–14.

31. Oberg AL, Kennedy RB, Li P, Ovsyannikova IG, Poland GA. Systems biology approaches to new vaccine development. Curr Opin Immunol [Internet]. 2011; 23(3): 436–443.

32. Liljeroos L, Malito E, Ferlenghi I, Bottomley MJ. Structural and Computational Biology in the Design of Immunogenic Vaccine Antigens. J Immunol Res [Internet]. 2015; 2015: 1–17.

33. Ali M, Pandey RK, Khatoon N, Narula A, Mishra A, Prajapati VK. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. Sci Rep. 2017; 7(1): 9232.

34. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics. 2015; 32(4): 511–517.

35. Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci. 2003; 12(5): 1007–1017.

36. Bhasin M, Raghava GPS. Prediction of CTL epitopes using QM, SVM and ANN techniques. Vaccine. 2004; 22(23): 3195–3204.

37. Bhasin M, Raghava G. A hybrid approach for predicting promiscuous MHC class I restricted T cell epitopes. J Biosci [Internet]. 2007; 32(1): 31–42.

38. Singh H, Raghava GPS. ProPred1: prediction of promiscuous MHC Class-I binding sites. Bioinformatics. 2003; 19(8):1009–1014.

39. Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. Bioinformatics. 2001; 17(12): 1236–1237.

40. Reche PA, Glutting J-P, Reinherz EL. Prediction of MHC class I binding peptides using profile motifs. Hum Immunol. 2002; 63(9): 701–709.

41. Reche PA, Glutting J-P, Zhang H, Reinherz EL. Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. Immunogenetics. 2004; 56(6): 405–419.

42. Reche PA, Reinherz EL. Prediction of peptide-MHC binding using profiles. Immunoinformatics Predict Immunogenicity Silico. 2007; 185–200.

43. Thakur N, Qureshi A, Kumar M. AVPpred: collection and prediction of highly effective antiviral peptides. Nucleic Acids Res. 2012; 40(W1): W199--W204.

44. Hayer J, Jadeau F, Deléage G, Kay A, Zoulim F, Combet C. HBVdb: a knowledge database for Hepatitis B Virus. Nucleic Acids Res. 2012; 41(D1): D566--D570.

45. Tcherepanov V, Ehlers A, Upton C. Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. BMC Genomics. 2006; 7(1): 150.

46. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007; 8(1): 4.

47. Magnan CN, Zeller M, Kayala MA, Vigil A, Randall A, Felgner PL, et al. High-throughput prediction of protein antigenicity using protein microarray data. Bioinformatics. 2010; 26(23): 2936–2943.

48. Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP: allergenicity prediction by descriptor fingerprints. Bioinformatics. 2013; 30(6) :846–851.

49. Saha S, Raghava GPS. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res. 2006; 34(suppl_2): W202--W209.

50. Magnan CN, Randall A, Baldi P. SOLpro: accurate sequence-based prediction of protein solubility. Bioinformatics. 2009; 25(17): 2200–2207.

51. Naito Y, Ui-Tei K, Nishikawa T, Takebe Y, Saigo K. siVirus: web-based antiviral siRNA design software for highly divergent viral sequences. Nucleic Acids Res. 2006; 34(suppl_2): W448--W450.

52. Routh A, Johnson JE. Discovery of functional genomic motifs in viruses with ViReMa--a Virus Recombination Mapper--for analysis of next-generation sequencing data. Nucleic Acids Res. 2013; 42(2): e11--e11.

53. Bruno L, Cortese M, Rappuoli R, Merola M. Lessons from Reverse Vaccinology for viral vaccine design. Curr Opin Virol [Internet]. 2015; 11: 89–97.

54. McLellan JS, Yang Y, Graham BS, Kwong PD. Structure of respiratory syncytial virus fusion glycoprotein in the postfusion conformation reveals preservation of neutralizing epitopes. J Virol. 2011; 85(15): 7788–7796.

55. Dormitzer PR, Suphaphiphat P, Gibson DG, Wentworth DE, Stockwell TB, Algire MA, et al. Synthetic generation of influenza vaccine viruses for rapid response to pandemics. Sci Transl Med. 2013; 5(185): 185ra68--185ra68.

56. Bainbridge JWB, Smith AJ, Barker SS, Robbie S, Henderson R, Balaggan K, et al. Effect of gene therapy on visual function in Leber's congenital amaurosis. N Engl J Med. 2008; 358(21): 2231–2239.

57. Afzal S, Wilkening S, von Kalle C, Schmidt M, Fronza R. GENE-IS: Time-Efficient and Accurate Analysis of Viral Integration Events in Large-Scale Gene Therapy Data. Mol Ther Acids. 2017; 6: 133–139.

58. Li J-W, Wan R, Yu C-S, Co NN, Wong N, Chan T-F. ViralFusionSeq: accurately discover viral integration events and reconstruct fusion transcripts at single-base resolution. Bioinformatics. 2013; 29(5): 649–651.

59. Ho DWH, Sze KMF, Ng IOL. Virus-Clip: a fast and memory-efficient viral integration site detection tool at single-base resolution with annotation capability. Oncotarget. 2015; 6(25): 20959.

60. Hood L, Rowen L. The human genome project: big science transforms biology and medicine. Genome Med. 2013; 5(9): 79.

61. Consortium TEP. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science (80-) [Internet]. 2004;

306(5696): 636–640.

62. Consortium EP, others. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489(7414): 57–74.

63. Romero P, Wagg J, Green ML, Kaiser D, Krummenacker M, Karp PD. Computational prediction of human metabolic pathways from the complete human genome. Genome Biol. 2004; 6(1): R2.

64. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003; 422(6928): 198–207.

65. Picotti P, Aebersold R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Methods. 2012; 9(6): 555–566.

66. Koenig T, Menze BH, Kirchner M, Monigatti F, Parker KC, Patterson T, et al. Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics. J Proteome Res. 2008; 7(9): 3708–3717.

67. Tabb DL, Eng JK, Yates III JR. Protein identification by SEQUEST. In: Proteome Research: Mass Spectrometry. Springer; 2001. p. 125–42.

68. Li W, Ji L, Goya J, Tan G, Wysocki VH. SQID: an intensity-incorporated protein identification algorithm for tandem mass spectrometry. J Proteome Res. 2011; 10(4): 1593–1602.

69. Liu J, Erassov A, Halina P, Canete M, Vo ND, Chung C, et al. Sequential interval motif search: unrestricted database surveys of global MS/MS data sets for detection of putative post-translational modifications. Anal Chem. 2008; 80(20): 7846–7854.

70. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Verezemska O, Isbandi M, et al. Genomes OnLine Database (GOLD) v. 6: data updates and feature enhancements. Nucleic Acids Res. 2016; gkw992.

71. Theobald DL. A formal test of the theory of universal common ancestry. Nature. 2010; 465(7295): 219–222.

72. Stoneking M, Krause J. Learning about human population history from ancient and modern genomes. Nat Rev Genet. 2011; 12(9): 603–614.

73. Primose SB, Twyman RM. Biotechnology and Genomics in Medicine. In: Genomics [Internet]. Malden, MA, USA: Blackwell Science Ltd; 2007. p. 1–19.

74. Waldman SA, Terzic A. Systems-Based Discovery Advances Drug Development. Clin Pharmacol Ther. 2013; 93(4): 285–287.

75. E Laing R, Hess P, Shen Y, Wang J, X Hu S. The role and impact of SNPs in pharmacogenomics and personalized medicine. Curr Drug Metab. 2011; 12(5): 460–486.

76. Picton ACP, Paximadis M, Tiemessen CT. Genetic variation within the gene encoding the HIV-1 CCR5 coreceptor in two South African populations. Infect Genet Evol. 2010; 10(4): 487–494.

77. Cabello J V, Lodeyro AF, Zurbriggen MD. Novel perspectives for the engineering of abiotic stress tolerance in plants. Curr Opin Biotechnol. 2014; 26: 62–70.

78. Chantre Nongpiur R, Lata Singla-Pareek S, Pareek A. Genomics approaches for improving salinity stress tolerance in crop plants. Curr Genomics. 2016; 17(4): 343–357.

79. Reguera M, Peleg Z, Blumwald E. Targeting metabolic pathways for genetic engineering abiotic stress-tolerance in crops. Biochim Biophys Acta (BBA)-Gene Regul Mech. 2012; 1819(2): 186–194.

80. Van Os H, Andrzejewski S, Bakker E, Barrena I, Bryan GJ, Caromel B, et al. Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. Genetics. 2006;173(2):1075–87.

81. Consortium PGS, Others. Genome sequence and analysis of the tuber crop potato. Nature. 2011; 475(7355): 189–195.

82. Paz MM, Veilleux RE. Influence of culture medium and in vitro conditions on shoot regeneration in Solanum phureja monoploids and fertility of regenerated doubled monoploids. Plant Breed. 1999; 118(1): 53–57.

83. Barrell PJ, Meiyalaghan S, Jacobs JME, Conner AJ. Applications of biotechnology and genomics in potato improvement. Plant Biotechnol J. 2013; 11(8): 907–20.

84. Navarro C, Abelenda JA, Cruz-Oró E, Cuéllar CA, Tamaki S, Silva J, et al. Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. Nature. 2011; 478(7367): 119–122.

85. Kloosterman B, Abelenda JA, Gomez M del MC, Oortwijn M, de Boer JM, Kowitwanich K, et al. Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature. 2013; 495(7440): 246–250.

86. Jupe F, Pritchard L, Etherington GJ, MacKenzie K, Cock PJA, Wright F, et al. Identification and localisation of the NB-LRR gene family within the potato genome. BMC Genomics. 2012; 13(1): 75.

87. Lozano R, Ponce O, Ramirez M, Mostajo N, Orjeda G. Genome-wide identification and mapping of NBS-encoding resistance genes in Solanum tuberosum group phureja. PLoS One. 2012; 7(4): e34775.

88. Wilson RF. The role of genomics and biotechnology in achieving global food security for high-oleic vegetable oil. J Oleo Sci. 2012; 61(7): 357–367.

89. Huxley C. Exploring gene function: use of yeast artificial chromosome transgenesis. Methods. 1998;14(2): 199–210.

90. Anand R. Yeast artificial chromosomes (YACs) and the analysis of complex genomes. Trends Biotechnol [Internet]. 1992; 10(1–2): 35–40.

91. Giraldo P, Montoliu L. Size matters: use of YACs, BACs and PACs in transgenic animals. Transgenic Res [Internet]. 2001; 10(2): 83–103.

92. Brem G, Besenfelder U, Aigner B, Müller M, Liebl I, Schütz G, et al. YAC transgenesis in farm animals: Rescue of albinism in rabbits. Mol Reprod Dev [Internet]. 1996; 44(1): 56–62.

93. Moreira PN, Pozueta J, Pérez-Crespo M, Valdivieso F, Gutiérrez-Adán A, Montoliu L. Improving the generation of genomic-type transgenic mice by ICSI. Transgenic Res [Internet]. 2007; 16(2):163–168.

94. Pook MA, Al-Mahdawi S, Carroll CJ, Cossée M, Puccio H, Lawrence L, et al. Rescue of the Friedreich's ataxia knockout mouse by human YAC transgenesis. Neurogenetics [Internet]. 2001; 3(4): 185–193.

95.Poser I, Sarov M, Hutchins JRA, Hériché J-K, Toyoda Y, Pozniakovsky A, et al. BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat Methods [Internet]. 2008; 5(5): 409–415.

96. Harrington JJ, Bokkelen G Van, Mays RW, Gustashaw K, Willard HF. Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. Nat Genet [Internet]. 1997; 15(4): 345–355.

97. Kazuki Y, Hoshiya H, Takiguchi M, Abe S, Iida Y, Osaki M, et al. Refined human artificial chromosome vectors for gene therapy and animal transgenesis. Gene Ther [Internet]. 2011; 18(4): 384–393.