

## CONCEPTS OF BIOINFORMATICS AND ITS APPLICATION IN VETERINARY RESEARCH AND VACCINES DEVELOPMENT

KAIKABO<sup>1</sup>, A. A. and KALSHINGI<sup>2</sup>, H. A.

<sup>1</sup>Bacterial Research Department <sup>2</sup>Viral Vaccines Production Department  
National Veterinary Research Institute, P.M.B 01 Vom, Plateau State, Nigeria.

\*Correspondence: E-mail: [karkafa@yahoo.co.uk](mailto:karkafa@yahoo.co.uk) Tel: +2348036935243

### SUMMARY

Bioinformatics is the science of managing and analyzing biological information. Because of the rapidly growing sequence biological data, bioinformatics tools and algorithms have been developed. Using bioinformatics algorithms such as BLAST (basic local alignment sequence tool), FASTA and Clustal W solutions for sequence search and analysis, in combination with other techniques constitute a time saving and cost-effective way to obtain important data on gene and protein levels information not easily obtainable by other techniques. Bioinformatics has advanced the course of research and future veterinary vaccines development because it has provided new tools for identification of vaccine targets from sequenced biological data of organisms. In Nigeria, there is lack of bioinformatics training in the universities, expect for short training courses in which few people are train, to use such knowledge for their research work only. In order to rapidly move research and development in Nigeria in 21<sup>st</sup> century and beyond, bioinformatics training in Nigerian Universities is required. Given appropriate research and computer infrastructure, researchers and scientists in Nigeria may have relatively easy access to the products of bioinformatics. However, the future use of this technology hinges on the availability of bioinformatics knowledge in the public domain. This paper highlights on some tools use in bioinformatics and their application in research and development.

**KEYWORDS:** Bioinformatics, Veterinary research, Vaccines, BLAST, FASTA, Clustal W

### INTRODUCTION

With advent of large-scale gene sequencing, a new scientific discipline known as computational molecular biology or bioinformatics was conceived. The terms bioinformatics and computational biology are often used interchangeably. However, bioinformatics more properly refers to the creation and advancement of algorithms, computational and statistical techniques, and theory to solve formal and practical problems inspired from the management and analysis of biological data. Computational biology on the other hand, refers to hypothesis-driven investigation of a specific biological problem using computers, carried out with experimental or simulated data, with the primary goal of discovery and advancement of biological

knowledge. Put more simply, bioinformatics is concerned with the information while computational biology is concerned with the hypothesis (Wikipedia, 2005). Bioinformatics is one of the latest additions to scientific vocabulary which seems to suggest a bridge between the world of biology and that of information technology.

The origin of bioinformatics can be traced to the developments by Sanger (Sanger and Coulson, 1975) of a technique for the sequencing of nucleic acids. The original technique devised in 1975, was subsequently improved upon and automated (Sanger *et al.*, 1977 and Maxam and Gilbert, 1977). For this invention and for working out the first complete nucleotide sequence for an organism (Sanger *et al.*, 1977), Sanger won a second Nobel Prize in chemistry,

which was awarded in 1980. His first prize for inventing the technique for sequencing of protein was awarded in 1962 (Sanger and Tuppy, 1961).

The pioneering efforts of Sanger, paved the way for the sequencing of other genomes, starting with bacterial plasmid pBR322 (Sutcliffe, 1979). At present, over one billion DNA and protein sequences have been determined and deposited in computerized databases which can be accessed and retrieved for research use. These sequences contain wealth of information hidden within them; including protein structure, disease mechanisms and drug targets sites. Therefore, bioinformatics is concerned with the access and analysis of databases of published genes and protein sequences. The major problem facing the researchers in biomedical and pharmaceutical disciplines is how to extract biologically useful information from millions of sequences. So, bioinformatics is using multidisciplinary approach which combines computer science, information technology, molecular biology, biochemistry etc, to address this gap of knowledge.

In all, bioinformatics can be said to be a theoretical discipline which attempt to make predictions about biological functions from sequence data. The exploitation of the vast amount of information in various genome sequence databases is dependent on the ability of the researcher to assign functions to the sequences. Now, using bioinformatics algorithm (otherwise called bioinformatics tools) sequences could be analyzed to provide useful biological information on the sequenced material. Some commonly used bioinformatics tools include; basic local alignment search tool BLAST, FASTA, and Clustal W. The use of such algorithms with a particular reference to veterinary research and vaccines development is discussed in this paper.

### **BIOINFORMATICS: A DEFINITION**

**(Molecular) bio informatics:** Bioinformatics is conceptualizing biology in terms of molecules (in the sense of physical chemistry) and applying “information techniques” (derived from disciplines such as applied mathematics, computer science and statistics) to understand and organize the information associated with the

molecules on large scale. In short, bioinformatics is management information system for molecular biology and has many practical applications (Gerstein *et al.*, 2001). Thus, bioinformatics is application of computational techniques to understand and organize the information associated with biological macromolecules.

Similarly, Bewaji (2003) defined bioinformatics as application of information technology to the domain of biology. Thus, to synthesize information contained in sequenced data, the use of powerful computers and the bioinformatics algorithm is pertinent. Therefore, to use computational molecular biology in research and development knowledge of computer, programming languages, natural sciences (applied mathematics, statistics, physical chemistry and physics), biochemistry and molecular biology are very necessary.

### **THE AIMS OF BIOINFORMATICS**

#### ***The aims of bioinformatics are:***

Firstly, to organize data in away that allows researchers to access existing information and submit new entries as they are produced e.g. the protein data bank for 3D macromolecule structures (Bernstein *et al.*, 1977 and Berman *et al.*, 2000).

Secondly, to develop tools and resources that aid in the analysis of data e.g. having sequenced a particular protein it is of interest to compare it with previously characterized sequences. Such tools may include FASTA (Pearson and Lipman, 1988), PSI-BLAST (Altschul, 1997), BLAST and Clustal W (Altschul *et al.*, 1990 and Thompson *et al.*, 1994) among others.

Thirdly, to use tool(s) to analyze the data and interpret the results in a biologically meaningful manner.

### **BIOINFORMATICS ALGORITHMS**

Bioinformatics algorithms (tools) are used in the analysis of biological sequenced data; overviews of some of these tools are as follows.

### **BLAST**

The full name of this tool is Basic Local Alignment Search Tool (BLAST); it is employed for rapid searching of nucleotides and protein databases (Altschul *et al.*, 1990). It compares sequence data with the already characterized sequences to give meaningful biological information for the unknown sequence used for the search. It provides optimal local alignment to the query. This algorithm has the power to detect local and global alignment, region of similarity embedded in otherwise unrelated proteins. BLAST searches provide information on the function of uncharacterized proteins, this is obtained when the tool compare a query sequence to all sequence in a specified databases.

Comparison of sequence by basic local alignment search tool is done in a pair wise manner, each comparison is a given a score reflecting the degree of similarity between the query and the sequence being compared. At the end of the comparison, the higher the score the greater the degree of similarity. The similarity is measured and shown by aligning two sequences. Either as indicated, the alignment of sequence can be local or global which means local alignment is an optimal alignment that include **only** the most similar local regions on the sequence, while in global alignment it include all characters from each sequence. The graphical representation of the tool can be accessed at <http://www.ncbi.nlm.nih.gov/BLAST>.

### **Clustal W**

Unlike the above tool, Clustal W is a general-purpose algorithm used for multiple sequence analysis for deoxyribonucleic acids (DNA) or proteins. It produces biologically meaningful multiple sequence alignments (Thompson, 1994) of divergent sequences, it also computes best matches for the selected sequences and line them up so that identities, similarities and differences can be seen by viewing cladograms and phylograms both of which are parts of the Clustal W algorithm. The physical features of the tool can be accessed at web <http://www.bioweb.pasteur.fr/seqanal/interface/clustalW.html>.

### **FASTA**

FASTA tool allows for the comparison of a query sequence to a DNA sequence database. The tool use a fast search to initially identify sequences from database with a high degree of similarity to the query sequence, then it conduct a second comparison on the initially selected sequences. FASTA algorithm is slower and more sensitive than the BLAST; this is because FASTA tolerates gaps in the aligned sequences. Again, FASTA is employed in homology searching that is to say it search for homologous sequences in sequence databases of proteins and nucleotides for similar sequences (Pearson and Lipman, 1988). Graphical interface of this tool is on the web at <http://fasta.bioch.virginia.edu/>

## **BIOINFORMATICS IN VETERINARY RESEARCH**

In veterinary research, Bioinformatics tools could be used to:

### **(a) Generate biological data for Research**

Bioinformatics produces databanks such as collections of protein sequences for biology and biotechnology, as well as computer software for analysis. Users such as veterinary researchers, biotechnologists and academic biologists may choose to install these components on a local computer system, or access them over the internet, using the publicly available databanks. Bioinformatics currently deals with several main types of biological data as described (Pongor and Landsman, 1999):

- i- Sequences and structure of genes and proteins. Sequences are the simplest way to represent a macromolecule. The structure of genes that code for the sequence of amino acids in proteins is produced in this form by genome sequencing projects. Protein sequences are usually obtained via computer-based translation of genomic data.
- ii- 3-D molecular structures. These are obtained by physical measurement (X-ray, Nuclear Magnetic Resonance) combined with computer modeling.
- iii- Genome structures and functions. The genome of an organism is composed of its entire genetic material. Information on genome structure and function is a

basic description that is continuously updated with new information including links to other databases.

iv- Bibliographic data. Such as abstracts of scientific articles. The amount of data has increased exponentially, especially since the onset of genome projects, such as the human genome sequencing programmes. The data are currently organized into a small number of large public databanks available through the internet. Data management, including data processing and database maintenance, is the first and most fundamental task of bioinformatics. Another large sub-field, biomathematics or biocomputing, is concerned with developing specialized algorithms for accessing and analysing these data. The most frequent research tasks in this sub-field are sequence similarity searching, to find a protein or gene similar to a novel sequence, and database retrieval.

#### **(b) Mine, Retrieve and Analyze Biological Data**

The growing numbers of sequences in public domain, of pathogens, plants, animals that are useful resources for the researcher to exploit for research work could be mined, retrieved and analyzed by the use of bioinformatics tools to derive meaningful information. Another aspect of bioinformatics in sequence analysis is the automatic search for genes and regulatory sequences within a genome. One core aspect of research in computational biology focuses on database development: how to integrate and optimally query data from (for example) genomic DNA sequence, spatial and temporal patterns of mRNA expression, protein structure, immunological reactivity, clinical outcomes, publication records, and other sources. A second focus involves pattern recognition algorithms for such areas as nucleic acid or protein sequence assembly, sequence alignment for similarity comparisons or phylogeny reconstruction, motif recognition in linear sequences or higher-order structure, and common patterns of gene expression. Both database integration and pattern recognition depend absolutely on accessing data from diverse sources, and being able to integrate,

transform, and reproduce these data in new formats. For example, the widespread utility of BLAST for the identification of gene similarity (Altschul *et al.*, 1990) is attributable not only to the algorithm itself (and its implementation), but also to the availability of databases such as GenBank, the European Molecular Biology Laboratory (EMBL), and the DNA Data Bank of Japan (DDBJ), which pool genomic data from a variety of sources. BLAST would be of limited utility without a broad-based database to query. The "post genomic era" holds phenomenal promise for identifying the mechanistic bases of organismal development, metabolic processes, and disease, and we can confidently predict that bioinformatics research will have a dramatic impact on improving our understanding of such diverse areas as the regulation of gene expression, protein structure determination, comparative evolution, and drug discovery.

In veterinary research for instance, bioinformatics tools was use in the detection of Newcastle disease. Alfonso *et al.* (2006) use Bioinformatics to examine the genome of NDV to determine which sequence mismatches have the potential to produce false negative results. Ideally, molecular diagnostic tests are directed at highly conserved areas of the genome, but for tests which target a particular motif such as the fusion protein cleavage site (Fusion test), improvements are limited by the understanding of the genomic variability and the primer/probe binding constraints at the target region. This shows how bioinformatics may be used to identify diagnostic problems and to generate novel solutions for the continued improvement and development of molecular diagnostics. Similarly, Kumar (2003) used bioinformatics approach to identify antigenic epitopes from Rabies virus glycoprotein G, which could be use for the development of antirabies sub unit vaccine.

#### **(c) Predict and Identify Protein in a sequence**

Protein structure prediction is another important application of bioinformatics. The amino acid sequence of a protein, the so-called **primary structure**, can be easily determined from the sequence on the gene that codes for it.

In the vast majority of cases, this primary structure uniquely determines a structure in its native environment. (Of course, there are exceptions, such as the bovine spongiform encephalopathy (Mad Cow Disease - prion.) Knowledge of this structure is vital in understanding the function of the protein. For lack of better terms, structural information is usually classified as one of secondary, tertiary and quaternary structure.

One of the key ideas in bioinformatics is the notion of homology. In the genomic branch of bioinformatics, homology is used to predict the function of a gene: if the sequence of gene *A*, whose function is known, is homologous to the sequence of gene *B*, whose function is unknown, one could infer that *B* may share *A*'s function. In the structural branch of bioinformatics, homology is used to determine which parts of a protein are important in structure formation and interaction with other proteins. In a technique called homology modelling, this information is used to predict the structure of a protein once the structure of a homologous protein is known. This currently remains the only way to predict protein structures reliably. One example of this is the similar protein homology between haemoglobin in humans and the haemoglobin in legumes (leghemoglobin). Both serve the same purpose of transporting oxygen in the organism. Though both of these proteins have completely different amino acid sequences, their protein structures are virtually identical, which reflects their near identical purposes. Other techniques for predicting protein structure include protein threading and *de novo* (from scratch) physics-based modeling (wikipedia, 2005).

#### **(d) Laboratory application**

The use of computers for biology starts in the laboratory; for instance, to plan how a DNA molecule will be cut and tailored with the several hundreds of enzyme reagents available. In order to carry out the relatively simple task of cutting out a precise fragment of a DNA piece, it is necessary to find one or two enzymes that cut somewhere near the ends of the desired piece, but will not cut the fragment itself. One such

enzyme may cut a piece of DNA into a few, or into several hundred fragments, depending on the sequence of the DNA piece. A computer can enumerate all the possible fragments that can be obtained, and suggest enzyme combinations, and a protocol for the experiment. A more sophisticated task is the characterization of a gene sequence that is obtained from an experiment. To this end, the biologist performs a database search on several of the publicly accessible and frequently updated sequence databases available on the internet. The gene sequence is compared with the sequences in the DNA database, resulting in a ranked list of the 'hits' to the most similar sequences found in the database. Just a few sufficiently similar sequences are usually enough to predict the properties and hence the natural function of the new gene or protein with considerable probability. If no obviously similar sequences are found in the databank, then more sophisticated tools, such as pattern searching, could provide characteristics to predict properties of unknown genes or proteins. The majority of current molecular biology research relies on these techniques. Similarly, bioinformatics is applied in the laboratory for primer design, which is used to amplify sequence in techniques such as polymerase chain reaction and hybridization reactions.

### **VETERINARY VACCINES DEVELOPMENT**

DNA vaccines have many advantages over conventional vaccines, and offer an attractive approach to vaccination for veterinary species. Although DNA vaccination is still in its infancy even in the west, the biggest handicap to the development of animal vaccines is the lack of available data on immune responses to pathogens that exclusively infect veterinary species, although many studies in mice have shown that naked DNA vaccination induces a large spectrum of immune responses, so that this type of vaccine can readily be considered against pathogens that have not inspired large-scale immunological studies. DNA vaccination trials carried out in mammals (cattle, pig, sheep, dog, cat, horse, and rabbit), birds (duck, chicken, turkey) and fish (Salmonidae) against their target pathogens. Clearly, DNA vaccination has already inspired

great interest in the veterinary field. The use of bioinformatics in vaccine targets identification have greatly reduces the time and cost of Vaccinology. Thus, identification of target antigenic peptides in a genomic sequence of pathogens made development of naked DNA veterinary vaccines possible. This bioinformatics approach could also be use in the development of immunodiagnostic(s) and vaccine(s) such as *M. bovis* immunodiagnostic (Vordemeier *et al*, 2001) for the diagnosis of bovine tuberculosis, Heart water vaccine (Van kleef *et al*, 2006) for the control of *Ehrlichia ruminantium*, in heart water vaccine, bioinformatics tool aided the identification of antigens from the genomic sequence of *Cowdria ruminantium* which stimulate proliferation of lymphocytes in immunized cattle, and Foot and Mouth Disease vaccine (Taboga *et al* 1996), developed from bioinformatics identification of VP 1 fragment from foot and mouth disease virus genomic sequence which elicit antibody response. The developments of these biological was enhanced largely by genomics and proteomics studies using bioinformatics tools.

**(A)PROTEOMICS AND THE FUTURE OF VETERINARY VACCINE DEVELOPMENT**

A molecular understanding of all the information present in any given genome requires an interdisciplinary approach. Research in many different fields, such as structural biology, cell and molecular biology, biochemistry, genetics and informatics, will all need to contribute, as will the veterinary practitioner. Only if attention is focused on determining the functions and interactions of proteins will their biological significance be revealed. A combination of the different approaches in proteomics will provide a better understanding of physiological processes and how they regulate one another. In turn, this information will lead to a better understanding of pathological changes related to disease. The ultimate goal in veterinary medicine is to prevent or cure disease before permanent damage has occurred, or before side effects become evident. Treatment of diseases, if not curative, should provide palliative care and relief from the symptoms. The choice between a preventive course of action or a treatment relies on a clear recognition of the exact disease status of the patient. Proteomics will help

define and improve the process of disease diagnosis. For example, markers could be used to define certain tumours so that a diagnosis can be established, a prognosis defined, and a treatment decision made.. New proteomic approaches will also enable the identification of new therapeutic targets, which could be used to identify small chemical compounds in high-throughput screening procedures. They can also be further developed for new drugs and vaccines, which will allow the generation of new therapeutic concepts for the benefit of man and animals. Finally, new molecular understanding of how biological processes are regulated in cells will allow the development of diagnostic tools and drugs for diseases, which are yet unknown.

**BIOINFORMATICS IN THE DEVELOPING COUNTRIES**

Public bioinformatics resources, such as databanks and software tools that are crucial for biotechnology projects, are today available via the internet. Scientists need only a computer and an internet connection of a certain quality to use them. If these conditions exist, the situation of a developing country researchers and biologists are not different from that of an academic biologist and veterinary researcher in an industrialized country. However, the use of bioinformatics, to produce software from an algorithm, is not an easy one. It requires a supportive research and development climate that generates a local need for such research, and an appropriate computer science infrastructure. At present, bioinformatics has successfully been applied only in those developing countries where these requirements are met, such as Brazil, China, India, Mexico and South Africa (Pongor and Landsman, 1999). As at yet, there is no bioinformatics training in Nigerian universities, this is required to move research and development in 21<sup>st</sup> century and beyond.

**FUTURE PERSPECTIVES OF BIOINFORMATICS**

Bioinformatics cannot be disregarded by any country intending to remain up-to-date in the biomedical, biotechnological and agricultural

sectors. In addition to this general trend, developing countries may also want to manage their own specific data on indigenous biological species, on local epidemiology and biodiversity programmes. These tasks clearly require that statisticians and informatics experts become advanced users of bioinformatics software and develop a capability to solve problems locally. This process does not require large resources in it but will allow developing countries to further investigate their own biological resources. To facilitate this process biomathematics/biocomputing should be introduced to universities, and the establishment of small software groups and companies should be encouraged.

At present, advanced bioinformatics is concentrated in a few research centres and private companies around the world that have the capacity to employ personnel with highly specialized training. In spite of the fact that bioinformatics methods are freely accessible, there is clearly a gap between the developing and the industrialized world, which must be consciously narrowed. Bioinformatics is indeed the enabling technology for several fields of biomedical and agricultural research. The use of bioinformatics spreads freely through the internet and it helps developing countries to catch up with industrialized countries. All this is based, however, on the principle that information resources worldwide remain freely accessible. If this should change in the future, it might widen the North-South gap in biotechnology (Pongor and Landsman, 1999).

### ON-LINE BIOINFORMATICS TRAINING AND RESOURCES

Some useful bioinformatics sites are provided below; these will go along way in providing materials for self-training and research collaborations. However, researchers and scientists in Nigeria should be proactive in seeking collaboration with researchers/scientists in Africa and other parts of the world so that they can gainfully have bioinformatics competence to move research and development in the near future.

Bioinformatics networks, training and information sites in Africa and some parts of the world.

1. West African Biotechnology Workshop Series  
<http://www.wawb.org> ; [info@wawb.org](mailto:info@wawb.org)
  2. National Bioinformatics network (South Africa) Courses  
<http://www.nbn.ac.za/education/courses.html>
  3. South African National Bioinformatics Institute Training Information  
<http://www.sanbi.ac.za/mrc/training.html>
  4. Bioinformatics Training Course University of Sfax, Tunis.
  5. South African National Bioinformatics Institute University of the Western Cape, South Africa  
<http://www.sanbi.ac.za/tdrcourse>
  6. African Bioinformatics Network  
<http://www.abionet.org>
  7. National Bioinformatics Network (South Africa)  
<http://www.nbn.ac.za>
  8. Federation for African Societies for Biochemistry and Molecular Biology  
<http://www.fasmb.org.za>
  9. African Society for human Genetics  
<http://www.afshg.org>
- Other sites with direct training initiatives in Africa and globally are:
10. International Society for computational Biology (iscb) <http://www.iscb.org>(provide information on certificate and degree training on bioinformatics and computational biology worldwide)
  11. South-South initiative for Tropical Disease Research (ssi- tdr)  
<http://www.ssi-tdr.net>
  12. S-star Alliance Bioinformatics Education  
<http://www.s-star.org>

### REFERENCES

- ALTSCHUL, S.F., GISH, W., MILLER, W., MYERS E.W. and LIPMAN, D.J. (1990): Basic Local Alignment Search Tool. *Mol. Biol.*, 215:403-410.

- BEWAJI, C.O. (2003): BIOINFORMATICS and COMPUTATIONAL MOLECULAR BIOLOGY TECHNIQUES MANUAL. 2003 UNAAB Summer Course in Practical Biotechnology. University of Agriculture Abekouta, Nigeria. 28<sup>th</sup> July-1<sup>st</sup> August; 1-5.
- BERNSTEIN, F.C., KOETZLE, T.F., WILLIAMS, G.J., MEYER, E.F., JR. BRICE, M.D. and RODGER, J.R. (1997): The Protein Data Bank: A Computer-based Archival file Macromolecular Structures. *Eur.J.Biochem.*, **80** (2):319-24.
- BERMAN, H.M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T.N. and WEISSIG, H. (2000): The Protein Data Bank. *Nucleic Acids Res.*, **28** (1):235-42.
- GERSTEIN, M., GREENBAUM, D. and LUSCOMBE, N.M. (2001): What is Bioinformatics? An Introduction and Overview. *Year Book of Medical Informatics*; 83-99.
- KLEEF, M.V., GUNTER, N.J., HENRIETTE, M., ALLSOPP, B.A., SHKAP, V. and BROWN, W.C. (2000): Identification of Cowdria ruminantium antigen that stimulate proliferation of lymphocytes from cattle immunized by infection and treatment with in activated organism. *Immun. Infect.*, **68**(2):603-614.
- KUMAR, D. (2003): Identification of promiscuous MHC Class-I and MHC Class-II binding epitopes of rabies virus glycoprotein. Unpublished Ph.D Thesis submitted to Deemed University, Indian Veterinary Research Institute, Izatnagar (UP)-243122; 1-49.
- MAXAM, A.M. and GILBERT, W. (1977): A new method for sequencing DNA. *Proc.Natl.Acad.Sci. (USA)* **74**:560-564.
- MIAKIM, L., SUAREZ, D.L. and AFUNSO, C.L. (2006): Use of bioinformatics in improving detection of Newcastle disease virus. *J. Vet. Diag. Invest.* (in press).
- PONGOR, S. and LANDSMAN, D. (1999): Bioinformatics and the developing world *Biotechnology and Development Monitor*, **40**:10-13.
- PEARSON, W.R. and LIPMAN, D.J. (1988): Improved tools for biological sequence Comparison. *Proc.Natl.Acad.Sci. (USA)*, **85**(8):2444-2448.
- SANGER, F. and TUPPY, H. (1961): The amino acid sequence in the phenyl chain of insulin. *Biochem.J.* **49**:463-490.
- SANGER, F., AIR, G.M., BROWN, N.L., COULSON, A.R., FIDDES, J.C., HUTCHISON, C.A (111), SLOCOMBE, P.M. and SMITH, M. (1977): Nucleotide sequence of bacteriophage ØX174. *Nature*, **265**:687-695.
- SANGER, F. and COULSON, A.R (1975): A rapid method for determining sequences in DNA by primed DNA polymerase. *J.Mol.Biol.*, **94**:444-448.
- SUTCLIFFE, G. (1979): Nucleotide sequence of E.coli plamid pBR322. *Cold Spring Harbor Symp. Quant. Biol.*, **43**:77-90.
- TABOGA, O., CECILIA, T., ELISA, C. *et al.* (1997): A large scale evaluation of peptide vaccines against foot and mouth disease: Lack of solid protection in cattle and isolation of escape mutants. *J. Virol.*, **71**(4):2606-2614.
- THOMPSON, J.D., HIGGINS, D.G. and GIBSON, T.J. (1994): Clustal W: Improving sensitivity of Progress multiple sequence with affine gap penalties. *Nucleic Acids Res.*, **11**(22):4673-4680.
- VORDERMEIER, H.M., WHELAN, A., COKLE, P.C. *et al.* (2001): Use of synthetic peptides derived from antigens ESAT-6 & CFP-10 for differential diagnosis of bovine tuberculosis in cattle. *Clin. Diagn. Lab. Immun.*, **8**:571-578.
- WIKIPEDIA, (2005): Bioinformatics. In: <http://en.wikipedia.org/bioinformatics> Accessed 12/12/2006.