Review article

Application of bioinformatics in tropical medicine

Wiwanitkit V

Wiwanitkit House, Bangkhae, Bangkok Thailand 10160

Abstract

Bioinformatics is a usage of information technology to help solve biological problems by designing novel and incisive algorithms and methods of analyses. Bioinformatics becomes a discipline vital in the era of post-genomics. In this review article, the application of bioinformatics in tropical medicine will be presented and discussed.

Keywords: Bioinformatics; Tropical medicine; Data mining technology; Genomics technology; Genomics technology

What is bioinformatics[1-8]?

Bioinformatics is the term which was first described in 1988 by Lim H with the original definition as "a collective term for data compilation, organisation, analysis and dissemination." It is a usage of information technology to help solve biological problems by designing novel and incisive algorithms and methods of analyses. It can help establish innovative software and create new/maintain existing databases of information, allowing open access to the records held within them. "Why this new science is helpful at present?" is an interesting query. The good answer is that present data is collected from a variety of sources and bioinformatics is a clue to allow the effortless transfer of information gathered and the interrogation of databases across the global interface. Therefore, bioinformatics becomes a discipline vital in the era of post-genomics.

A brief history of progress of bioinformatics returned back to genetics history which firstly began when Mendel proved his laws of hereditary with varieties of peas and flowers in 1865. The progress is listed in Table 1.

Table 1 Progress in history of bioinformatics

Period	Important events
19 th century	the invention of the compound microscope
1960	complete sequencing of an enzyme, ribonuclease at the first time[3-6]
1990	Human Genome Project (HGP) by the United States Department of Energy (DoE) and the National Institutes of Health[3-6]
1995	first complete genome (Haemophilus influenzae)[2]

An important jump is due to the HGP which is aimed at identification all chemical base pairs and all genes that make up the 23 chromosome pairs found in human DNA[3-6]. This helps develop the next generation of methods for simulating cellular behaviour and pathways and devise means to apply IT to the modelling of cellular functions as specified by the enormous datasets[3-6]. The HGP has brought to light the limitations of traditional work and the necessity

Correspondence to: Wiwanitkit V, Wiwanitkit House, Bangkhae, Bangkok Thailand 10160. Email: wviroj@ yahoo. com

to incorporate original techniques to allow greater understanding of protein function, protein-protein interactions and protein-DNA interactions. After publication of HGP, the "omic revolution" came[7-9]. Bioinformatics can help close the gap between the data stored and its biological significance. At present, bioinformatics has been split into various subjects, several omics.

Application of data mining technology in tropical medicine [8,9]

Data mining is the first step of bioinformatics work[8,9]. Passive techniques for data mining include various techniques in statistics such as min, max, SD, regression and correlation[8,9]. The scientists have a high workload to know and understand how to use those statistical techniques and how to interpret the result[8,9]. Active techniques are newly developed to serve this proposed. Artificial Intelligence (AI) is the hope for solving of data mining problem. At present, many online databases are available and help searching data for further manipulation. The most well-known database, PubMed, is the best example. With public available database and data mining technology, applied researches in tropical medicine can be performed. For example, the complex microscopic findings of parasites egg can be signalized, stored as AI and further manipulated for discrimination or diagnosis.

Application of genomics technology in tropical medicine

Genomics is an analysis of gene and transcript within the genome[10,11]. Nucleic acid characterization and recognizing coding sequences is the first step that can be done by several methods based on the basic principles that: a) all genes have certain regulatory signals positioned in or about them, b) all genes by definition contain specific code patterns, and c) and many genes have already been sequenced and recognized in other organisms[10-12], Therefore, locating the position of genes in DNA can be done by "searching by signal", "searching by content", and "homology inference", respectively[10-12]. At pres-

ent many servers have been established and can help with gene finding analyses. Most of those online servers, such as Gene Feature Search (http://searchlauncher. bcm. tmc. edu/seq-search/gene-search. html) and Gene Recognition (http://www. nslij-genetics. org/gene/), apply methods as well as consolidate the information and usually combine signal and content methods with homology inference in order to ascertain exon locations. Gene searching is applicable in studied on new identified pathogen in cases of emerging diseases.

Apart from gene searching, comparative genomics is also helpful in tropical medicine. Complete genomic sequences are now available for many model organisms[13-18]. Analyzing and comparing genetic material from different species to study evolution, gene function, and inherited disease can be easily done. Comparison can be on gene location, gene structure (exon number, exon length, intron length and sequence similarity) and gene characteristics (splice sites, codon usage and conserved synteny)[13-18]. Functional genomics, development and application of system-wide experimental approaches to assess gene function, can also be applicable[19-22]. High-throughput methodologies combined with statistical or computational analysis of the results is widely used. Phylogenetic profile, fusion method and Rosetta Stone model are the basic functional genomics techniques that are widely used in tropical medicine researches[19-22]. The phylogenetics analysis becomes a basic requirement test in study of new pathogens in emerging infectious diseases[21, 23]. This analysis help detect the origin of the outbreak. Network creation is the high level that help bridge the interrelationships in host-pathogen-environment system and this is considered as a novel applied functional genomics application.

Finally, structural genomics is the newest challenge in genomics technology[24,25]. Basically, proteins comprise of various combinations of different amino acids and the protein sequence assembled on the ribosome during protein synthesis is known as the primary structure. The sequence within primary structure is considered a code or signal with can be successfully manipulated by computational techniques. Prediction of secondary structure, a string of

amino acids with specific folds into various elements, and tertiary structure, a specific packing of secondary elements, can be done and is useful for further application. The three dimensional pattern of the protein can be directly linked to its function. Knowledge on structural genomics can help access the structural change in several conditions. For example, the study of the structural change due to mutation in tropical hemoglobinopathy was successfully performed[26-28]. The better understand on the pathogenesis can be derived.

Application of proteomics technology in tropical medicine [29,30]

Basically, proteins may be simple or conjugated. Proteomics help better access the protein. Similar to genomics, looking for homology by comparative proteomics can be done and homology is defined as the divergent evolution of two proteins from a common ancestor. This can help track the evolution pathway of focused species and help find new drugs. Solving protein structure by structural proteomics adding to current methods using NMR and X-ray crystallography can be very helpful in assessment of interaction between protein and new therapeutic agent design. In addition, functional proteomics help better understand protein function. Adding to classical techniques, protein isolation and gel electrophoresis, computational techniques can help in-depth study on protein function. Gene Ontology is the good way. Biological process, molecular function and cellular component prediction by gene ontology can be applied for study on interaction of molecules[31,32]. In silico mutation and hybridization can be tested[33,34].

Application of chemoinformatics technology in tropical medicine[35-38]

Chemoinformatics focus on the information resources needed to optimize the properties of a molecular ligand with computational means. It is useful for and vaccine drug discovery in medicine. Chemoinformatics covers the scope on the design, creation, organisation, management, retrieval, analysis, dissemination, visualization and use of chemical information.

Structural prediction of new chemical can be performed at presenting. In addition, substructure search can also be done. Similarity in chemicals led to successful adoption in pharmaceutical purpose. Quantum Mechanics on the three dimensional structure of chemical can also helpful in study of reaction. Calculate energy of a system as a function of nuclear positions by molecular mechanics principle "Energy = Stretching Energy + Bending Energy + Torsion Energy + Non-Bonded Interactions Energy" help understand the nature of reaction. Basically, reaction favorability usually fit with classical thermodynamics rule; reaction will proceed to thermodynamic equilibrium, maintaining a steady state ratio of products. "Total requirement energy = Energy getting in − Energy giving out" is the fundamental that can be applied in simulating of many reactions. For example, the study of energy favorability in drug and malarial interaction can help design new antimalarial drug[39].

REFERENCES

1 **Trgovcevic Z**. 50 years of molecular biology. *Lijec Vjesn.* 1994; 116:315-8.

2 **Fleischmann RD**, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR. et al. Wole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science.* 1995;269:496-512.

3 **Portin P**. Historical development of the concept of the gene. *J Med Philos.* 2002; 27:257-86.

4 **Csako G**. Present and future of rapid and/or high-throughput methods for nucleic acid testing. *Clin Chim Acta.* 2006; 363:6-31.

5 **Watson JD**, Cook-Deegan RM. Origins of the Human Genome Project. *FASEB J* . 1991; 5:8-11.

6 **Lewin R**. National Academy looks at human genome project, sees progress. *Science.* 1987; 235:747-8.

7 **Kanehisa M**, Bork P. Bioinformatics in the post-sequence era. *Nat Genet.* 2003; 33(Suppl): 305-10.

8 **Yang ZR**, Hamer R. Bio-basis function neural networks in protein data mining. *Curr Pharm Des.* 2007; 13:1403-13.

9 **Ahrens CH**, Wagner U, Rehrauer HK, Turker C, Schlapbach R. Current challenges and approaches for the synergistic use of systems biology data in the scientific community. *EXS.* 2007; 97:277-307.

10 **Koning DJ**, Haley CS. Genetical genomics in humans and model organisms. *Trends Genet.* 2005; 21:377-81.

11 **Claverie JM**, Abergel C, Audic S, Ogata H. Recent advances in computational genomics. *Pharmacogenomics.*

2001;2:361-72.

12 **Bentley DR**. Genomic sequence information should be released immediately and freely in the public domain. *Science*. 1996;274:533-4.

13 **Karlin S**, Mrazek J, Gentles AJ. Genome comparisons and analysis. *Curr Opin Struct Biol*. 2003;13:344-52.

14 **Haubold B**, Wiehe T. Comparative genomics: methods and applications. *Naturwissenschaften*. 2004;91:405-21

15 **Hardison RC**. Comparative genomics. *PLoS Biol*. 2003;1:E58.

16 **Kawagashira N**. Orthologue, paralogue and comparative genomics for gene network analysis. *Tanpakushitsu Kakusan Koso*. 2004;49(17 Suppl):2909.

17 **O'Brien SJ**, Wienberg J, Lyons LA. Comparative genomics: lessons from cats. *Trends Genet*. 1997;13:393-9.

18 **O'Brien SJ**, Menotti-Raymond M, Murphy WJ, Nash WG, Wienberg J, Stanyon R, et al. The promise of comparative genomics in mammals. *Science*. 1999;286:458-62, 479-81.

19 **Evans MJ**, Carlton MB, Russ AP. Gene trapping and functional genomics. *Trends Genet*. 1997;13:370-4.

20 **Rastan S**, Beeley LJ. Functional genomics: going forwards from the databases. *Curr Opin Genet Dev*. 1997;7:777-83.

21 **Eisen JA**, Fraser CM. Phylogenomics: intersection of evolution and genomics. *Science*. 2003;300:1706-7.

22 **Jordan BR**. "Genomics": buzzword or reality? *J Biomed Sci*. 1999;6:145-50.

23 **Relman DA**. Emerging infections and newly-recognised pathogens. *Neth J Med*. 1997;50:216-20.

24 **Baker EN**. Structural genomics as an approach towards understanding the biology of tuberculosis. *J Struct Funct Genomics*. 2007;8:57-65.

25 **Lundstrom K**. Structural genomics and drug discovery. *J Cell Mol Med*. 2007;11:224-38.

26 **Wiwanitkit V**. Structural analysis on the abnormal elongated hemoglobin "hemoglobin Geneva". *Nanomedicine*. 2005;1:216-8.

27 **Wiwanitkit V**. Secondary and tertiary structure aberration of alpha globin chain in haemoglobin Q-India disorder. *Indian J Pathol Microbiol*. 2006;49:491-4.

28 **Wiwanitkit V**. Modeling for tertiary structure of globin chain in Hemoglobin Suan-Dok disorder. *Hematology*. 2005;10:163-5.

29 **Cho WC**. Proteomics technologies and challenges. *Genomics Proteomics Bioinformatics*. 2007;5:77-85.

30 **Calvo KR**, Liotta LA, Petricoin EF. Clinical proteomics: from biomarker discovery and cell signaling profiles to individualized personal therapy. *Biosci Rep*. 2005;25:107-25.

31 **Harris MA**, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, et al. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004 1;32(Database issue):D258-61.

32 **Lewis SE**. Gene Ontology: looking backwards and forwards. Genome Biol. 2005;6:103.

33 Wiwanitkit V. A study on functional similarity between dengue non structural protein 1 and platelet integrin/adhesin protein, CD61. *J Ayub Med Coll Abbottabad*. 2006;18:13-6.

34 **Wiwanitkit V**. Interaction between BRCA1 and human papilloma virus E7: an ontology study. *Arch Gynecol Obstet*. 2006;274:146-9.

35 **Anders MW**, Yin H, Jones JP. Application of computational chemistry in the study of biologically reactive intermediates. *Adv Exp Med Biol*. 1996;387:347-53.

36 **Ghosh A**. Computational bioinorganic chemistry. Part III. The tools of the trade: from high-level ab initio calculations to structural bioinformatics. *Curr Opin Chem Biol*. 2003;7:110-2.

37 **Antoniou D**, Basner J, Núñez S, Schwartz SD. Computational and theoretical methods to explore the relation between enzyme dynamics and catalysis. *Chem Rev*. 2006;106:3170-87.

38 **Bruice TC**. Computational approaches: reaction trajectories, structures, and atomic motions. Enzyme reactions and proficiency. *Chem Rev*. 2006;106:3119-39.

39 **Wiwanitkit V**. Quantum chemical analysis of alternative pathways for iron activation step for artemisinin, a new antimalarial drug. *J Infect*. 2006;53:148-51.