บทความปริทัศน์
Review Article

Bioinformatics and Dentistry
ชีวสารสนเทศศาสตร์กับทันตแพทยศาสตร์

จิตจิโรจน์ อิทธิชัยเจริญ[1], ไมเคิล แมคคัลราฟ[2]
[1]ภาควิชาทันตวิทยา-พยาธิวิทยาช่องปาก คณะทันตแพทยศาสตร์ มหาวิทยาลัยเชียงใหม่
[2]โรงเรียนทันตแพทย์ คณะแพทย์ ทันตแพทย์ และวิทยาศาสตร์สุขภาพ มหาวิทยาลัยแห่งเมลเบิร์น
Jitjiroj Ittichaicharoen[1], McCullough Michael[2]
[1]Department of Odontology and Oral Pathology, Faculty of Dentistry, Chiang Mai University
[2]School of Dental Science, Faculty of Medicine, Dentistry and Heath Science, The University of Melbourne

ชม.ทันตสาร 2551; 29(2) : 7-14
CM Dent J 2008; 29(2) : 7-14

Abstract

Bioinformatics was established more than 40 years ago since the sequencing of the first protein by Margaret Dayhoff in 1965. Subsequently, many tools and technologies have been discovered to support Bioinformatics, and resulting in much scientific research in diverse fields such as microbiological, biomolecular and advanced medical explorations. Reciprocally, research efforts in these disciplines result in enhancing bioinformatics tools and techniques. Currently, bioinformatics has only had a superficial effect on dental research. Although many dental researchers undertake molecular research, they did not fully utilize bioinformatics. The aims of this paper is to outline how Bioinformatics can be utilized in many areas of dental research such as oral medicine, oral microbiology and oral cancer, to aid in the development of diagnostic methods, treatment planning, and potential preventive techniques.

บทคัดย่อ

ชีวสารสนเทศศาสตร์ (Bioinformatics) ได้รับการก่อตั้งมาเป็นเวลากว่า 40 ปี นับตั้งแต่โปรตีนตัวแรกถูกเรียงลำดับกรดอะมิโน โดย Margaret Dayhoff ในปี พ.ศ. 2508 หลังจากนั้นเป็นต้นมา วิทยาการ และเครื่องมือต่างๆ ก็ถูกคิดค้นขึ้นเพื่อสนับสนุนงานทางชีวสารสนเทศศาสตร์ และปรากฏในงานวิจัยหลากหลายสาขา เช่น ทางด้านจุลชีววิทยา อณูชีววิทยา และการค้นคว้าทางการแพทย์ ความพยายามในการวิจัยของสาขาวิชาเหล่านั้น ยังช่วยเพิ่มขีดความสามารถ และความแม่นยำให้กับวิทยาการ และเครื่องมือต่างๆ ในทางชีวสารสนเทศศาสตร์เองอีกด้วย แต่ปัจจุบันนี้ชีวสารสนเทศศาสตร์มีผลต่องานวิจัยทางทันตแพทย์เพียงเล็กน้อยเท่านั้น แม้ว่าจะมีงานวิจัยทางทันตแพทย์หลายชิ้นอาศัยวิธีการทางอณูชีววิทยา แต่การนำความรู้ทางชีวสารสนเทศศาสตร์มาใช้ก็ยังไม่เต็มที่เท่าที่ควร วัตถุประสงค์ของเอกสารนี้จึงพยายามชี้แนวทางการนำชีวสารสนเทศศาสตร์มาปรับใช้ในงานด้านต่างๆ ทางทันตแพทย์ศาสตร์ เช่น เวชศาสตร์ช่องปาก (oral medicine) เชื้อโรคในช่องปาก (oral microbiology) และการศึกษามะเร็งในช่องปาก (oral

Key Words: Bioinformatics, Dentistry, Oral biology, Genomics, Proteomics

cancer) เพื่อช่วยในการพัฒนาวิธีการวินิจฉัยโรค การวางแผนรักษา รวมทั้งการส่งเสริมป้องกันอย่างมีประสิทธิภาพ

คำไขรหัส: ชีวสารสนเทศศาสตร์ ทันตแพทยศาสตร์ ชีววิทยาในช่องปาก จีโนมิกส์ โปรตีโอมิกส์

General Feature of Bioinformatics

Bioinformatics is a relatively new discipline that had its roots in the 1960s(1) before the Human Genome Project commenced(2), at a time when computers were known only within limited areas of academic research(3). It could be argued that bioinformatic technologies were required before scientists could envision the possibility of mapping the complete human genome. It was almost 40 years ago that the first protein sequence study that is claimed to be the first step of bioinformatics was established by Margaret Dayhoff in a study titled 'Atlas of Protein Sequence and Structure'(1). Subsequently, the word 'bioinformatics' was first introduced in 1989 by Masys(4). The evolution of bioinformatics has become rapid as the price of computers dropped during the 80's, the Internet created global links during the 90's, and high productive computational technologies become common in the 2000's(5). Nowadays, the term 'Bioinformatics' can be described as a management system, applied by information technologies for use in molecular biology(6). The terms bioinformatics and computational biology are often incorrectly used interchangeably. The computational biology is actually a superset of bioinformatics because the computational biology involves all biological areas that use computation. According to the above definition, bioinformatics is limited to sequence, structural and functional analysis of products expressed by the computational biological methods(1). Thus, the development of bioinformatics requires the multi-disciplinary involvement of both biology science and computer science(6). As described previously(6) the 3 purposes of bioinformatics are

1. To organise biological data for easy access by scientists, biologists or any researchers.
2. To develop tools and techniques to analyse this data.
3. To translate this information into biological meaningful.

Bioinformatical techniques have been developed to serve the massive amount of information from the HGP, which is nearing completion of entire 3 billion base pair. Therefore, database management systems (DBMS) used to store these data provides security, integrity, multiple concurrent access and recovery control, and a user-accessible catalog(7).

Current bioinformatics tools in genomic research

At the beginning of the development of bioinformatics, when the supercomputer and communication network were not well known and low quality, scientists attempted to collected data by computer used for protein biochemistry(3). We now have ability to predict the result of proteins or genes by using the supercomputers that collect data from around the world via the global network, and analyzes the data by using powerful computer chips. We discuss this in terms of the available genomic databases, database mining tools and experimental analysis tools.

Genomic Database

There are three global bioinformatic databases. They are

1. NCBI, National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/Genbank/). Founded in November 1988, by the U.S. Senate who realized the need for computerized data processing to manage biomedical and biochemical information. NCBI conducts research analyzing gene sequence and its products, improving a better understanding, via collaboration, of the genes analyzed and predicting the potential structure of the molecules that these genes may express(8). Furthermore, NCBI provides some services and electronic tools to public, such as PubMed, BLAST, Entrez, BankIt, OMIM, Taxonomy and Structure.

2. EBI, European Bioinformatics Institute (www.ebi.ac.uk/embl/index.html) is a part of the European Molecular Biology Laboratory (EMBL) established in September 1994(8). EBI's main task is to carry out research and supply bioinformatics information by developing sophisticated algorithms and database ability. Some of services that are offered by EBI are database (SWISS-PROT), data submission, FASTA and BLITZ (query database and similarity search) and online applications.

3. GenomeNet, Japanese Bioinformatics Server (www.genome.ad.jp) is a Japanese online database system that was opened in September 1991. GenomeNet products research molecular and cellular biology. The fundamental database resource of GenomeNet is the Kyoto Encyclopedia of Genes and Genomes (KEGG), which contains a number of databases such as PATHWAY (for biological processes), BRITE (Functional hierarchies of biological systems), GENES (Gene catalogs and ortholog relations in complete genomes), and LIGAND (Chemical compounds, drugs, glycans, and reactions)(9).

NCBI, EBI and GenomeNet all work collaboratively and exchange their submitted information on a daily basis. Thus scientists receive well-organized up to date information. Moreover, each database acts as a repository for each other, if one crashes or is temporary offline for maintenance, the other two store identical data for updating.

Database Mining Tools

As an example of the type of database mining tools available, we will discuss Basic Local Alignment Search Tool (BLAST), searches for similarity among either the nucleotide sequence database or protein sequence database. To accomplish this task, BLAST works with the fundamentals of sequence alignment, can find potential homology for a portion of the compared sequences. Then scientist can subsequently predict function or create 3-D structure. The sequence alignment can be classified into two types. They are global, which tries to match overall alignment across the entire length of the given sequences, and local alignment, which finds the optimal alignment between local regions of the specified sequences. All sequence comparisons use scoring matrix to determine sequence identity. For comparing protein, the scoring matrixes of BLAST are Point-Accepted-Mutation (PAM) for global alignment and Blocks Substitution Matrix (BLOSSUM) for local alignment. Additionally, BLAST provides different programs for different purposes, such as BLASTp for searching protein sequence against the protein database and BLASTx used to search a translated nucleotide sequence against the protein database(1).

Analysis Tools

The experimental analysis tools that are useful for bioinformatics researches are PCR (Polymerase

Chain Reaction) and Microarray. PCR is used for a broad variety of experiments and analyses such as genetic fingerprinting, detection of hereditary diseases, gene cloning analysis of ancient DNA and comparison of gene expression. In contrast, Microarray allows scientists to observe thousands of gene expression concurrently(10). Thus researchers do not need to study gene expression one at a time by using a series complex laboratory protocol and time-consuming experimental methods.

Bioinformatics and Dentistry

Bioinformatics and its productive information methodologies and techniques affect all areas of biological science and medical knowledge. Currently, bioinformatics has had only a superficial influence on oral biology or dental knowledge. A simple search of the Pubmed scientific literature with specific keywords highlights the limited extent that oral and dental research has made of bioinformatics. A search in all articles in over 4,000 journals in the repository Pubmed(1), with no date limit and in all titles, keywords and abstracts resulting in 15,598 articles for the term [bioinformatics]; 60 for [bioinformatics and oral] and 15 for [bioinformatics and dental]. Moreover, analysis of these two later combinations shows many duplication and several that have no relationship to dentistry(11-13). Thus, this highlights the imbalance between the access dentistry has made with bioinformatics related to other areas of medical research. Three areas of dentistry were bioinformatics has been used are oral medicine, oral microbiology and oral cancers.

Oral Medicine

Oral medicine is an area of dentistry that studies and treats the relation between systemic diseases and oral diseases. The NATIONAL CENTER FOR HEALTH STATISTICS of the United States shows that fewer than 40% of American adults visited to a physician, while more than 60% attended a dental clinic(14). Thus, if dentists may be able to recognize early stages of a systemic disease, and thus refer the patient for early treatment. Moreover, oral fluids such as saliva may be useful as a marker for monitoring the progression of disease or the efficacy of drugs.

Xerostomia (dry mouth) can be caused by many systemic diseases, such as diabetes mellitus, Sjögren's syndrome or vitamin C deficiency(15, 16). In addition, the major proteins of saliva are amylases, statherins, histatins, glycosylated proline-rich proteins, peroxidase, cystatins, and mucin(17). Each of these major proteins can have several genetic polymorphisms therefore the alteration of the concentration and conformation of the proteins may reflect an underlining systemic disease. Moreover, scientists have shown stable mRNA in saliva(18). Consequently, saliva can be used as a marker for many diseases. For example, Reznick(19) found a significant correlation between the severity of diabetes as measured by glycosylated haemoglobin and the increase in both salivary and serum antioxidants.

This kind of data could be collected to build a database that is used as a criterion for normality against which at risk patient can be assessed. Furthermore, saliva collection is a non-invasive technique and may possibly not be time dependent as is blood sugar where patients must not eat overnight before blood sampling.

Oral Microbiology

Most dentists are aware of the variety of oral microbes that play a major role in many oral diseases such as dental caries, gingivitis, periodontitis and oral candidosis. Many researchers have published discoveries, new ideas and new methods for investigating and detecting these microbes and have used the techniques of bioinformatics to improve their experiments

Firstly, the advances and ease of PCR have been well utilized by dental researchers. For example, Hoshino(20) used PCR to distinguish strains of Streptococci in saliva by evaluating the glucosyltranfrase gene. Hoshino utilized GenBank database for DNA sequence and by multiple alignment PCR product sequences showed that streptococcal gtf genes have a 140 amino acids region that showed less than 51% identity among isolates. PCR was used to analyze laboratory and clinical isolates found that each strain of streptococci produces its specific product. Moreover, this result was used to detect gtf-specific product of streptococci from saliva. Then, the researcher could use this data to predict the risk of dental caries of patient by using a saliva sample(20).

The development of anti-caries vaccine has been improved by the development of efficient and inexpensive genome sequencing methods(21). As the first step in the development of an anti-caries vaccine, it is necessary to find the conserved genes that are expressed by all strains of the target pathogen. Bioinformatic tools are vital for this research.

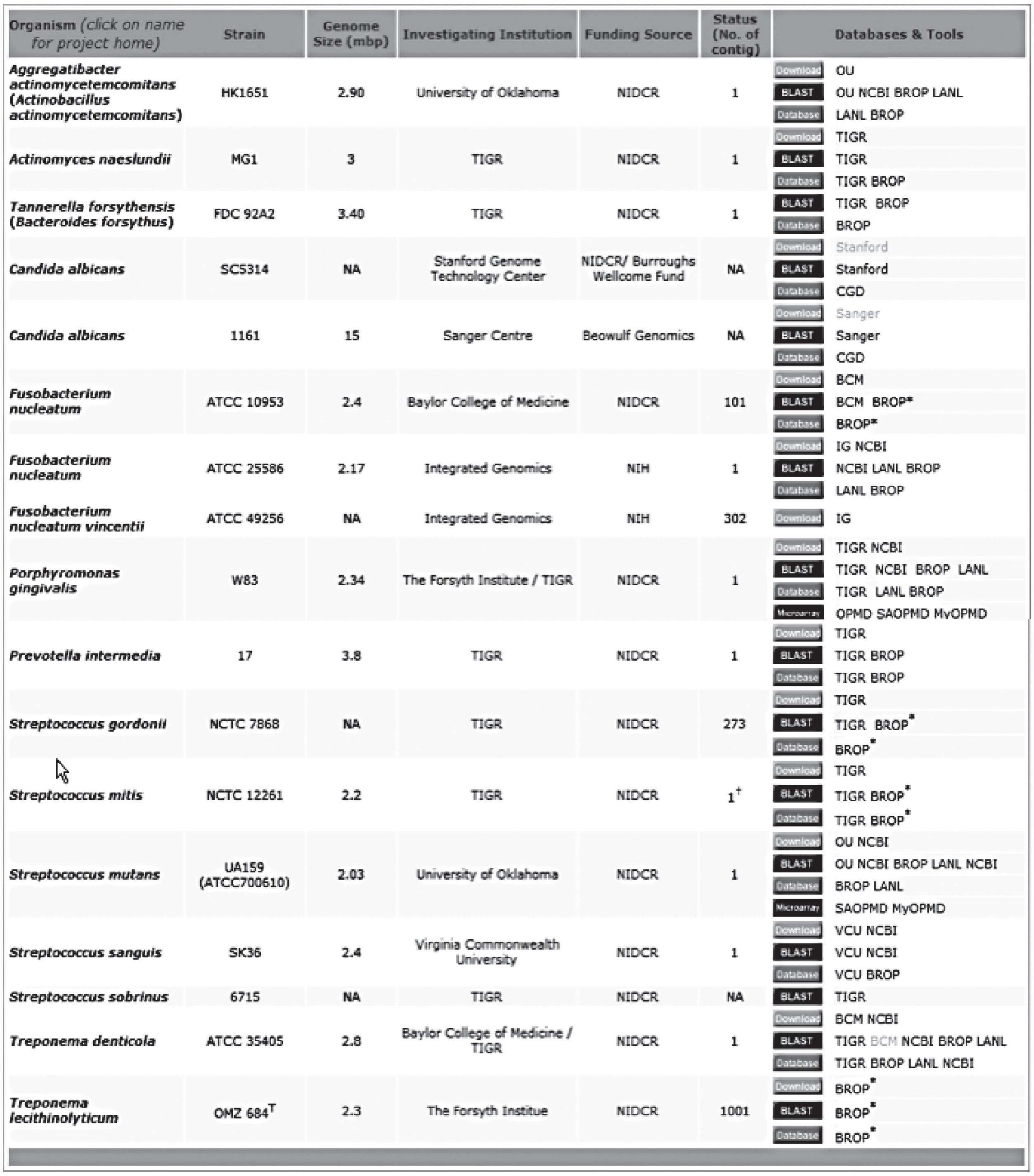
Bioinformatics Resource for Oral Pathogen (BROP) is an independent database website that provides information and offers bioinformatics tools for oral pathogens. BROP lists the oral pathogens that have been sequenced completely and partially decoded (Table 1). There is some duplication and inconsistency of information as different sources have used different criteria for gene identification, naming and functional description(22).

The mouth is a source of diverse pathogens whose analysis of function and virulence may allow for control and protection. Bioinformatics provides a variety of tools and techniques, however, the oral pathogen database needs to be developed continuously and standardized properly.

Oral Cancers

Oral mucosal change may result from several diverse causes such as normal variation of the mouth, autoimmune diseases, chronic inflammatory conditions of oral. The diagnosis of these lesions can require biopsy and histopathological examination and occasionally lesions can have the potential for future malignant development while not at the time of initial biopsy showing frank malignancy. Histopathological assessment alone fails to predict the malignant potential(10).

Early histopathological change in potentially malignant and normal lesions show varying degree of dysplasia and cellular atypia that may have little value in predicting future adverse outcome. Efforts have now been made to assess for variation in gene expression between leukoplakia and oral squamous cell carcinoma (OSCC). The salivary transcriptome in patients with oral cancer found that four salivary mRNAs (OAZ, SAT,IL8, and IL1b) predicatively revealed a 91% sensitivity and specificity for oral cancer detection(18). Furthermore, a variety of PCR techniques and microarray have been used to investigate gene expression of leukoplakia comparing with OSCC and this revealed marker genes significantly overexpressed in OSCC, while others were up-regulated in leukoplakia. This research outlines how the expression of 11 genes could be markers to differentiate leukoplakia from OSCC with 97% accuracy. This research has the potential as markers for cancerous transformation and to determine which lesions have a high risk for progression from precancerous to cancerous lesion and which malignant lesion may have a high metastasis rate.



Organism *(click on name for project home)*	**Strain**	**Genome Size (mbp)**	**Investigating Institution**	**Funding Source**	**Status (No. of contig)**	**Databases & Tools**
Aggregatibacter actinomycetemcomitans (Actinobacillus actinomycetemcomitans)	HK1651	2.90	University of Oklahoma	NIDCR	1	Download OU BLAST OU NCBI BROP LANL Database LANL BROP
Actinomyces naeslundii	MG1	3	TIGR	NIDCR	1	Download TIGR BLAST TIGR Database TIGR BROP
Tannerella forsythensis (Bacteroides forsythus)	FDC 92A2	3.40	TIGR	NIDCR	1	BLAST TIGR BROP Database BROP
Candida albicans	SC5314	NA	Stanford Genome Technology Center	NIDCR/ Burroughs Wellcome Fund	NA	Download Stanford BLAST Stanford Database CGD
Candida albicans	1161	15	Sanger Centre	Beowulf Genomics	NA	Download Sanger BLAST Sanger Database CGD
Fusobacterium nucleatum	ATCC 10953	2.4	Baylor College of Medicine	NIDCR	101	Download BCM BLAST BCM BROP* Database BROP*
Fusobacterium nucleatum	ATCC 25586	2.17	Integrated Genomics	NIH	1	Download IG NCBI BLAST NCBI LANL BROP Database LANL BROP
Fusobacterium nucleatum vincentii	ATCC 49256	NA	Integrated Genomics	NIH	302	Download IG
Porphyromonas gingivalis	W83	2.34	The Forsyth Institute / TIGR	NIDCR	1	Download TIGR NCBI BLAST TIGR NCBI BROP LANL Database TIGR LANL BROP Microarray OPMD SAOPMD MvOPMD
Prevotella intermedia	17	3.8	TIGR	NIDCR	1	Download TIGR BLAST TIGR BROP Database TIGR BROP
Streptococcus gordonii	NCTC 7868	NA	TIGR	NIDCR	273	Download TIGR BLAST TIGR BROP* Database BROP*
Streptococcus mitis	NCTC 12261	2.2	TIGR	NIDCR	1†	Download TIGR BLAST TIGR BROP* Database TIGR BROP*
Streptococcus mutans	UA159 (ATCC700610)	2.03	University of Oklahoma	NIDCR	1	Download OU NCBI BLAST OU NCBI BROP LANL NCBI Database BROP LANL Microarray SAOPMD MyOPMD
Streptococcus sanguis	SK36	2.4	Virginia Commonwealth University	NIDCR	1	Download VCU NCBI BLAST VCU NCBI Database VCU BROP
Streptococcus sobrinus	6715	NA	TIGR	NIDCR	NA	BLAST TIGR
Treponema denticola	ATCC 35405	2.8	Baylor College of Medicine / TIGR	NIDCR	1	Download BCM NCBI BLAST TIGR BCM NCBI BROP LANL Database TIGR BROP LANL NCBI
Treponema lecithinolyticum	OMZ 684^{T}	2.3	The Forsyth Institue	NIDCR	1001	Download BROP* BLAST BROP* Database BROP*

Table 1 *The list of oral pathogens that has been sequenced are publishes by BROP (the Bioinformatics of Resource of Oral Pathogens)*(24)

Conclusion

Bioinformatics has the potential to upgrade the diagnostic technique and to increase the accuracy of their diagnosis. Presently, few dental researches utilize bioinformatic tools to aid in their scientific endeavors. Furthermore, newer techniques that are being developed, such as microarray(23) and refined PCR techniques, have as their bases bioinformatics tools and require these tools for development and analysis. Such techniques have the potential to allow dental researchers not only to develop a new method of diagnosis, but potentially also to develop a new treatment.

Acknowledgements

We would like to thank Associate Professor Surawut Pongsiriwet for reviewing this manuscript and giving us a great comment

References

1. Jin X. *Essential Bioinformatics*. ed. New York: Cambridge University Press; 2006:3-27.
2. Kuo WP. Overview of bioinformatics and its application to oral genomics. *Adv Dent Res* 2003;17:89-94.
3. Hagen JB. THE ORIGINS OF BIOINFORMATICS. Nature Reviews Genetics 2000;1(3):231-236.
4. Bioinformatics.Net OILa. What is Bioin formatics? 2005 [cited 2007 26 April]; bioin formatics tutorial]. Available from: http://www.bioinformatics.vg/what-is-bioinformatics.shtml
5. Perez-Iratxeta C, Andrade-Navarro MA, Wren JD. Evolving research trends in bioinformatics. *Brief Bioinform* 2007;8(2):88-95. Epub 2006 Oct 2031.
6. Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. *Methods Inf Med* 2001;40(4):346-358.
7. Connolly T, Begg, C. *Database System a Practical Approach to Design, Implementation, and Management*. 4th ed. Essex: Pearson Education Limited; 2005:3-32.
8. Rashidi H, Buehler, LK Basics : *Applications in Biological Science and Medicine*. Boca Raton, Fla: CRC Press; 2000:Pages.
9. Kanehisa M, Goto S, Kawashima S, et al. The KEGG databases at GenomeNet. *Nucleic Acids Res* 2002;30(1):42-46.
10. Kuo WP, Whipple ME, Jenssen TK, et al. Microarrays and clinical dentistry. *J Am Dent Assoc* 2003;134(4):456-462.
11. Matthews D, Rocchi A, Wang EC, et al. Use of an interactive tool to assess patients' willingness-to-pay. *J Biomed Inform* 2001; 34(5):311-320.
12. Slavkin HC, Panagis JS, Kousvelari E. Future opportunities for bioengineering research at the National Institutes of Health. *Clin Orthop Relat Res* 1999(367 Suppl):S17-30.
13. Li ZR, Tian AJ, Yang YY. Preparing for the third millennium: the views of life informatics. *Medinfo* 1998;9(Pt 1):394-396.
14. Glick M. Molecular and protein markers of disease. *Ann N Y Acad Sci* 2007;1098:267-268.
15. van den Berg I, Pijpe J, Vissink A. Salivary gland parameters and clinical data related to the underlying disorder in patients with persisting xerostomia. *Eur J Oral Sci* 2007; 115(2):97-102.
16. Hu S, Loo JA, Wong DT. Human saliva proteome analysis. *Ann N Y Acad Sci* 2007; 1098:323-329.
17. Oppenheim FG, Salih E, Siqueira WL, et al. Salivary proteome and its genetic polymorphisms. *Ann N Y Acad Sci* 2007;1098:22-50. Epub 2007 Feb 2015.
18. Zimmermann BG, Park NJ, Wong DT. Genomic Targets in Saliva. *Annals of the New York Academy of Sciences* 2007;1098(1):184-191.

19. Reznick AZ, Shehadeh N, Shafir Y, et al. Free radicals related effects and antioxidants in saliva and serum of adolescents with Type 1 diabetes mellitus. *Arch Oral Biol* 2006; 51(8):640-648. Epub 2006 Apr 2018.
20. Hoshino T, Kawaguchi M, Shimizu N, et al. PCR detection and identification of oral streptococci in saliva samples using gtf genes. *Diagn Microbiol Infect Dis* 2004;48(3):195-199.
21. Tettelin H, Masignani V, Cieslewicz MJ, et al. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". *Proc Natl Acad Sci USA* 2005; 102(39):13950-13955. Epub 12005 Sep 13919.
22. Chen T, Abbey K, Deng WJ, et al. The bioinformatics resource for oral pathogens. *Nucleic Acids Res* 2005;33 (Web Server issue):W734-740.
23. Kuo WP, Whipple ME, Sonis ST, et al. Gene expression profiling by DNA microarrays and its application to dental research. *Oral Oncol* 2002;38(7):650-656.
24. BROP. ongoing oral pathogen genome sequencing projects 2004 Aug 21, 2004 [cited 2007 1 May]; a list of all the finished or ongoing oral pathogen genome sequencing projects and the available tools, databases and resource, including those provided here at BROP]. Available from: http://www.brop.org/index.php?name=News&file=article&sid=16

ขอสำเนาบทความที่:

อ.ทพ.จิตจิโรจน์ อิทธิชัยเจริญ ภาควิชาทันตวิทยา-พยาธิวิทยาช่องปาก คณะทันตแพทยศาสตร์ มหาวิทยาลัยเชียงใหม่ จ. เชียงใหม่ 50202

Reprint requests:

Dr. Jitjiroj Ittichaicharoen, Department of Odontology and Oral Pathology, Faculty of Dentistry, Chiang Mai University, Chiang Mai 50202