

การวิเคราะห์เวฟเลตของพอลิโนคลิกโอไทด์



นายจักรี อนันตศิริสมบัติ

ศูนย์วิทยทรัพยากร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาการคณนา ภาควิชาคณิตศาสตร์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2546

ISBN 974-17-5409-4

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

WAVELET ANALYSIS OF POLYNUCLEOTIDES



Mr.Jakree Anantasirisombat

ศูนย์วิทยทรัพยากร

จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computational Science

Department of Mathematics

Faculty of Science

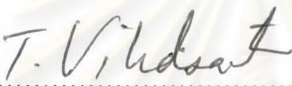
Chulalongkorn University

Academic Year 2003

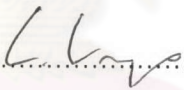
ISBN 974-17-5409-4

Thesis Title WAVELET ANALYSIS OF POLYNUCLEOTIDES
By Mr. Jakree Anantasirisombat
Field of Study Computaional Science
Thesis Advisor Assistant Professor Paisan Nakmahachalasint, Ph.D.
Thesis Co-advisor Chatchai Srinitiwarawong, Ph.D.


Accepted by the Faculty of Science, Chulalongkorn University in Partial
Fulfillment of the Requirements for the Master's Degree


..... Deputy Dean for Administrative Affairs,
Acting Dean, Faculty of Science
(Associate Professor Tharapong Vitidsant, Ph.D.)

THESIS COMMITTEE


..... Chairman
(Professor Chidchanok Lursinsap, Ph.D.)


..... Thesis Advisor
(Assistant Professor Paisan Nakmahachalasint, Ph.D.)


..... Thesis Co-advisor
(Chatchai Srinitiwarawong, Ph.D.)


..... Member
(Assistant Professor Wicharn Lewkeratiyutkul, Ph.D.)


..... Member
(Assistant Professor Anusorn Chonweeraayuth, Ph.D.)

จักรี อนันตศิริสมบัติ : การวิเคราะห์เวฟเลตของพอลินิวคลีโอไทด์ (WAVELET ANALYSIS OF POLYNUCLEOTIDES) อ. ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร.ไพศาล นาคมหาสาสินธุ์
อ. ที่ปรึกษาร่วม : อาจารย์ ดร.ฉัตรชัย ศรีนิตวิรวงศ์ 46 หน้า. ISBN 974-17-5409-4

การวิเคราะห์แบบเวฟเลตเป็นวิธีการหนึ่งทางการวิเคราะห์ และแสดงข้อมูลเชิงความถี่ วิธีการวิเคราะห์นี้จะนำมาประยุกต์ใช้กับข้อมูลทางชีววิทยา โดยเฉพาะกับสายดีเอ็นเอ วิธีการวิเคราะห์ประกอบด้วย 3 ขั้นตอนหลัก คือ การแปลงข้อมูลทางชีววิทยาให้อยู่ในรูปข้อมูลเชิงตัวเลข สำหรับการใช้คำนวณ, การหาค่าสูงสุดของความแตกต่างระหว่างข้อมูล เพื่อหาสัมประสิทธิ์ ขั้นตอนสุดท้ายคือการวิเคราะห์แบบเวฟเลต กับสายดีเอ็นเอที่ต้องการโดยใช้ชุดของสัมประสิทธิ์ที่ได้ เพื่อให้แสดงความแตกต่างของข้อมูลที่มีรูปแบบที่แน่นอนและไม่มีรูปแบบที่แน่นอน

วิทยานิพนธ์ฉบับนี้อธิบายถึงการประยุกต์ใช้ และขั้นตอนการทำนายบริเวณที่เป็นเอ็กซอนและอินทรอนบนสายดีเอ็นเอ ตัวอย่างที่ใช้ในการทดลองเป็นสิ่งมีชีวิตชั้นต่ำโดยเฉพาะกับ *Caenorhabditis elegans* และ *Saccharomyces cerevisiae* เอ็กซอนของ *C. elegans* สามารถระบุตำแหน่งได้จากผลของสเปกตรัม สำหรับเอ็กซอนของ *S. cerevisiae* สามารถระบุได้จากสเปกตรัมโดยใช้การอ่านตำแหน่งดีเอ็นเอไปข้างหน้า ส่วนสามตำแหน่งหลังได้จากการอ่านในทิศทางตรงข้าม โดยสเปกตรัมที่ได้จากการวิเคราะห์แบบเวฟเลตจะแสดงถึงบริเวณที่คาดว่าจะเป็ เอ็กซอนและบริเวณที่คาดว่าจะเป็อินทรอนในสายดีเอ็นเอนั้นได้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา คณิตศาสตร์
สาขาวิชา วิทยาการคณนา
ปีการศึกษา 2546

ลายมือชื่อนิสิต..... จักรี อนันตศิริสมบัติ
ลายมือชื่ออาจารย์ที่ปรึกษา..... Pain Namin
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

4472229823 : MAJOR COMPUTATIONAL SCIENCE

KEYWORDS: WAVELET ANALYSIS/ DNA SEQUENCE/ EXON/ INTRON

JAKREE ANANTASIRISOMBAT: WAVELET ANALYSIS OF POLYNUCLEOTIDES

THESIS ADVISOR: ASSISTANT PROFESSOR PAISAN NAKMAHACHALASINT, Ph.D.

THESIS CO-ADVISOR: CHATCHAI SRINITIWARAWONG, Ph.D.

46 pp. ISBN 974-17-5409-4

Some hidden information is revealed by means of frequency analysis. Wavelet analysis was introduced in this work to find correlation of interesting pattern in DNA sequences. Three main computational steps were applied. Conversion of bio-molecular sequences into numerical sequences was the first step in order to be mathematically analyzed. To obtain the set of coefficients that produces the maximum difference between exons and introns, the optimization was introduced. The last step was applying the maximization coefficients to the DNA sequences to distinguish the different patterns in that string.

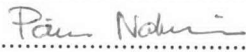
This thesis describes how the wavelet transform works, including its application and algorithms for prediction of exons and introns on a DNA sequence. Prokaryote is focused in this work, especially *Caenorhabditis elegans* and *Saccharomyces cerevisiae*. By applying the technique to *C. elegans*, locations of five exons on AF0099922 were accurately indicated by five peaks. In case of *S. cerevisiae*, locations of the first three exons on NC001135 were indicated by forward reading, while locations of the last three exons showed up by backward reading. Spectrum obtained by this method is able to predict the possibility of locations of exons and introns on a DNA sequences.

Department **Mathematics**

Field of study **Computational Science**

Academic year **2003**

Student's signature..... 

Advisor's signature..... 

Co-advisor's signature 

Acknowledgements

I would like to firstly express profound gratitude to my advisor, Assistant Professor Dr. Paisan Nakmahachalasint, for his help and fruitful suggestions. I am also grateful to my co-adviser, Dr. Chatchai Srinitiwarawong, for excellent support and noble guidance. Beyond, I express gratitude to Dr. Preprame Pattanamahakul and Dr. Rath Pichayangkura for acute information and enlighten me new visions in biology in all respects. Furthermore, I would like to thank Professor Dr. Chidchanok Luasinsap, the thesis chairman, and all members of this committee, Assistant Professor Dr. Wicharn Lewkeratitkul and Assistant Professor Anusorn Chonweeraayuth. This work will never be accomplished without them.

I would like to thank my best friend, Araya Wiwatwanich, who assists me including valuable collaboration and giving me aspiration. Moreover, I would like to thank all of my friends, especially Nucharee Ajarayamarayat, during my graduate program at Chulalongkorn University for their contribution in various kinds of helps. They all give me cheerfulness all through the program.

Most importantly, my family, I would love to give them a special thank especially my parents and Arjchanin Oonhawatt for warming encouragements. I want to thank all my colleagues at Advanced Info Service PLC.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Contents

	Page
Abstract (Thai).....	iv
Abstract (English).....	v
Acknowledgements	vi
List of Figures.....	viii
List of Tables	x
Chapter	
1 INTRODUCTION.....	1
1.1 Problem Identification and Objectives of Work.....	1
1.2 Scope of Work	1
2 LITERATURE REVIEW	3
3 BACKGROUND KNOWLEDGE	11
3.1 Wavelet Analysis	11
3.2 Basic Knowledge about DNA.....	17
4 EXPERIMENTAL RESULTS	23
4.1 DNA Spectra.....	23
4.2 Wavelet Analysis	27
4.2.1 Result of <i>Caenorhabditis elegans</i>	27
4.2.1 Result of <i>Saccharomyces cerevisiae</i>	32
5 CONCLUSIONS.....	42
REFERENCES.....	44
CURRICULUM VITAE.....	46

List of Figures

Figure	Page
2.1	Fourier transform of an exon.....5
2.2	The first half figure of Figure 2.1 of an exon.....6
2.3	2.3 A peak of Figure 2.2.....7
2.4	Spectrum plot of <i>Caenorhabditis elegans</i> by Fourier transform.....9
3.1	Morlet function with $f_c=1$ and $f_b=10$12
3.2	Plot of $s(t)$13
3.3	3D plot of wavelet transform of $s(t)$14
3.4	Fourier transform of $s(t)$15
3.5	Wavelet transform of $s(t)$16
3.6	Prokaryote and eukaryote.....17
3.7	Nucleic acid: RNA and DNA.....18
3.8	Bases and their bonding.....19
3.9	Transcription and translation.....20
3.10	Splicing.....21
3.11	Reading Frames.....22
4.1	Fourier transform of the exon of <i>Drosophila melanogaster</i>23
4.2	Wavelet transform on the exon of <i>Drosophila melanogaster</i>24
4.3	A peak of wavelet transform of the exon of <i>Drosophila melanogaster</i>25
4.4	The Fourier transform of the intron of <i>Drosophila melanogaster</i>26
4.5	The wavelet transform of the intron of <i>Drosophila melanogaster</i>26
4.6	Spectrum 3D plot by using wavelet transform.....28
4.7	Spectrum 3D plot by using wavelet transform.....28
4.8	Frequency analysis of <i>Caenorhabditis elegans</i> at $f_b=90$30
4.9	Frequency analysis of <i>Caenorhabditis elegans</i> at $f_b=560$30
4.10	Frequency analysis of <i>Caenorhabditis elegans</i> at $f_b=2130$31
4.11	Frequency analysis of <i>Caenorhabditis elegans</i> at $f_b=5032$31

Figure	Page
4.12 Spectrum 3D plot by using wavelet transform of <i>Saccharomyces cerevisiae</i>	34
4.13 Spectrum color plot by using wavelet transform of <i>Saccharomyces cerevisiae</i>	35
4.14 2D plot of spectrum of <i>Saccharomyces cerevisiae</i> with $f_b=44180$	36
4.15 2D plot of spectrum of <i>Saccharomyces cerevisiae</i> with $f_b=135310$	36
4.16 3D plot of frequency response of complement strand of <i>Saccharomyces cerevisiae</i>	38
4.17 Color plot of frequency response of complement strand of <i>Saccharomyces cerevisiae</i>	39
4.18 Frequency response of complement strand of <i>Saccharomyces cerevisiae</i> with $f_b=44180$	40
4.19 Frequency response of complement strand of <i>Saccharomyces cerevisiae</i> with $f_b=135310$	40

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

List of Tables

Table	Page
2.1 Locations of the five exons on the <i>Caenorhabditiselegans</i>	10
3.1 Codon Table	21
4.1 Locations of the five exons on the <i>Caenorhabditis elegans</i>	29
4.2 Locations of the six exons on the <i>Saccharomyces cerevisade</i>	33



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย