DEVELOPMENT OF A WEB-BASED TOOL FOR VISUALIZING AND PROFILING HUMAN ENDOGENOUS RETROVIRUSES (HERVS) NEIGHBORING GENES

MR. THITIPONG KAWICHAI

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE
(BIOINFORMATICS AND SYSTEMS BIOLOYGY)
SCHOOL OF BIORESOURCES AND TECHNOLOGY AND
SCHOOL OF INFORMATION TECHNOLOGY
KING MONGKUT'S UNIVERSITY OF TECHNOLOGY THONBURI
2011

Development of a Web-based Tool for Visualizing and Profiling Human Endogenous Retroviruses (HERVs) Neighboring Genes

Mr. Thitipong Kawichai B.Sc. (Mathematics)

A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of Master of Science (Bioinformatics and Systems Biology)
School of Bioresources and Technology and School of Information Technology
King Mongkut's University of Technology Thonburi
2011

Thesis Committee

.... Chairman of Thesis Committee
(Asst. Prof. Jonathan Hoyin Chan, Ph.D.)

.... Member and Thesis Advisor
(Lecturer, Santitham Prom-on, Ph.D.)

.... Member and Thesis Co-advisor
(Asst. Prof. Asawin Meechai, Ph.D.)

.... Member
(Asst. Prof. Kanokwan Poomputsa, Ph.D.)

.... Member
(Researcher, Jittima Piriyapongsa, Ph.D.)

PREFACE

This thesis is a part of the requirements to accomplish my master degree in Bioinformatics and Systems Biology, King Mongkut's University of Technology Thonburi. The topic of the study is "Development of a web-based tool for visualizing and profiling human endogenous retroviruses (HERVs) neighboring genes" or "การพัฒนาเครื่องมือบนเว็บสำหรับการแสดงผลและวิเคราะห์เอนโดจีนัสรีโทรไวรัสที่อยู่รอบยีน" in Thai. This work has been done with an aim to develop a novel bioinformatics tool, so-called HERV Profiler, for facilitating a study of human endogenous retroviruses or HERVs neighboring human genes. Furthermore, the case studies illustrating the utilization of HERV Profiler were also included in this work.

The thesis report is composed of five chapters. The first chapter is about background and rationale, objectives, scope of works, and expected outputs. The second chapter is related to backgrounds and literature reviews. For the third chapter, it is about materials and research methods. The fourth chapter is the results and discussions. Lastly, the conclusions and suggestions would be written in the fifth chapter.

Thesis Title	Development of a web-based tool for visualizing and profiling human endogenous retroviruses (HERVs) neighboring genes
Thesis Credits	12
Candidate	Mr. Thitipong Kawichai
Advisor	Dr. Santitham Prom-on
Co-advisor	Asst. Prof. Dr. Asawin Meechai
Program	Master of Science
Field of Study	Bioinformatics and Systems Biology
Faculty	School of Bioresources and Technology and School of Information Technology
B.E.	2554

Abstract

Human endogenous retroviruses (HERVs) are the remnants of ancient retroviral infections dispersed throughout the human genome. In particular cases, HERV regulatory sequences still remain active, and this allows them to influence the transcriptions of neighboring genes. Although HERV resources have been available for more than a decade, there is still not a database or tool that really supports the studies on HERVs nearby genes. To address this concern, this thesis aims to develop a web-based tool, so-called HERV Profiler, for facilitating investigations on neighboring HERVs of human genes. Several HERV characteristics were compiled in the HERV Profiler database, such as location relative to genes, family, superfamily, type of truncation patterns, orientation, and intactness ratio. HERV Profiler provides two main features, including visualizing and profiling the neighboring HERVs. In the visualizing, there are two different views to display the neighboring HERVs, including displaying in tabular and graphical format. In the HERV profiling, users can construct HERV profiles from the input gene list, find over-represented HERV types among the gene list, and rank input genes based on their over-represented HERVs. In the case studies, HERV Profiler can discover potential HERV types and genes under the condition of Systemic Lupus Erythematosus (SLE). In conclusion, HERV Profiler is a powerful tool that can efficiently facilitate HERV studies. Currently, HERV Profiler is freely accessible at http://sigma.cpe.kmutt.ac.th/herv_profiler.

Keywords: HERVs / neighboring genes / web-based tool / HERV profiling

หัวข้อวิทยานิพนธ์	การพัฒนาเครื่องมือบนเว็บสำหรับการแสดงผลและวิเคราะห์เอนโดจีนัสรีโทรไวรัสที่อยู่รอบยีน
หน่วยกิต	12
ผู้เขียน	นายฐิติพงษ์ กาวิชัย
อาจารย์ที่ปรึกษา	ดร.สันติธรรม พรหมอ่อน
อาจารย์ที่ปรึกษาร่วม	ผศ.ดร.อัศวิน มีชัย
หลักสูตร	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	ชีวสารสนเทศและชีววิทยาระบบ
คณะ	ทรัพยากรชีวภาพและเทคโนโลยี และเทคโนโลยีสารสนเทศ
พ.ศ.	2554

บทคัดย่อ

เอนโดจีนัสรีโทรไวรัสในมนุษย์หรือเอชอีอาร์วี (Human Endogenous Retroviruses or HERVs) เป็นลำดับดีเอ็นเอ (DNA sequences) ที่หลงเหลือจากการติดเชื้อของไวรัสในอดีต ซึ่งสามารถพบได้ทั่วไปในจีโนมของมนุษย์ ในบางกรณีลำดับดีเอ็นเอของตัวควบคุมการแสดงออกของเอชอีอาร์วียังคงทำงานได้อยู่อย่างมีประสิทธิภาพ ส่งผลให้เอชอีอาร์วีสามารถส่งผลกระทบต่อการแสดงออกของยีนที่อยู่บริเวณข้างเคียงได้ ถึงแม้ว่าข้อมูลที่เกี่ยวข้องกับเอชอีอาร์วีจะมีอยู่มาเป็นระยะเวลานานกว่าหนึ่งทศวรรษแล้วก็ตาม แต่ก็ยังไม่มีฐานข้อมูลหรือเครื่องมือใดที่ช่วยในการศึกษาเกี่ยวกับเอชอีอาร์วีที่อยู่ใกล้ยีนได้อย่างแท้จริง ดังนั้นวิทยานิพนธ์นี้จึงมีวัตถุประสงค์เพื่อพัฒนาเครื่องมือบนเว็บ หรือเอชอีอาร์วีโปรไฟล์เลอร์ (HERV Profiler) เพื่อช่วยอำนวยความสะดวกในการศึกษาเอชอีอาร์วีที่อยู่ข้างเคียงยีนในมนุษย์ เอชอีอาร์วีโปรไฟล์เลอร์ได้เก็บรวบรวมลักษณะของเอชอีอาร์วีไว้มากมาย ได้แก่ บริเวณของยีนที่เอชอีอาร์วีแทรกตัวอยู่ แฟมิลี่ (family) ซุปเปอร์แฟมิลี่ (superfamily) รูปแบบการขาดหายของส่วนประกอบของเอชอีอาร์วี (type of truncation patterns) และสัดส่วนความสมบูรณ์ของเอชอีอาร์วี (intactness ratio) การทำงานของเอชอีอาร์วีโปรไฟล์เลอร์สามารถแบ่งออกได้เป็นสองส่วนหลักๆ คือ การแสดงผลและการทำโปรไฟล์ลิ่งของเอชอีอาร์วีที่อยู่ใกล้ยีน ผู้ใช้สามารถแสดงผลของเอชอีอาร์วีได้สองรูปแบบ คือ ในรูปแบบของตารางและรูปภาพ การทำโปรไฟล์ลิ่งจะช่วยให้ผู้ใช้สามารถสร้างโปรไฟล์ หาชนิดของเอชอีอาร์วีที่ปรากฏอย่างเด่นชัดในชุดของยีน และเรียงลำดับยีนตามความสำคัญของเอชอีอาร์ที่อยู่ข้างเคียงยีนนั้นได้ จากกรณีศึกษาของโรคเอสแอลอี (SLE or Systemic Lupus Erythematosus) พบว่าเครื่องมือนี้สามารถใช้วิเคราะห์หาชนิดของเอชอีอาร์วีและยีนที่น่าจะเกี่ยวข้องกันได้ ดังนั้นเครื่องมือนี้นับว่าเป็นเครื่องมือที่มีประสิทธิภาพอย่างยิ่งในการช่วยอำนวยความสะดวกแก่ผู้ใช้ที่ต้องการศึกษาเกี่ยวกับเอชอีอาร์ที่อยู่ข้างเคียงยีนในมนุษย์ เอชอีอาร์วีโปรไฟล์เลอร์เปิดให้บริการผ่านทางเว็บไซด์ http://sigma.cpe.kmutt.ac.th/herv_profiler

คำสำคัญ: เอชอีอาร์วี / ยีนที่อยู่ข้างเคียง / เครื่องมือบนเว็บ / เอชอีอาร์วีโปรไฟล์ลิ่ง

ACKNOWLEDGEMENTS

At first, I would like to express my sincere gratitude and deep appreciation to my advisor, Dr. Santitham Prom-on, for his kindness, invaluable guidance, and encouragement. He gave assistances and supported me everything along the time that the thesis has been done. Also, I would like to thank my co-advisor, Asst. Prof. Dr. Asawin Meechai, who gave me a lot of advices and useful suggestions. Frequently, he also assisted me to correct and improve my works. Furthermore, I would like to acknowledge my committees, Asst. Prof. Dr. Jonathan Hoyin Chan, Asst. Prof. Dr. Kanokwan Poomputsa, and Dr. Jittima Piriyapongsa for their valuable suggestions and comments, which were very useful in improving the works.

In addition, I would like to thank Mr. Pumipat Tongyoo for his suggestions, comments, and assistances. Those suggestions leaded me to solve several problems during the course of the thesis. Furthermore, I would like to thank my mother for her encouragement and taking care of me. Although there is no any technical suggestion from her, only the cheering words could inspire me to be able to overcome all of the problems occurred at that time. An important one in my family that I would like to specially acknowledge here is my father. Even though he is not here anymore, I believe that he is always beside me and give me his assistances until I have reached my goal now.

I am grateful to Bioinformatics and Systems Biology, King Mongkut's University of Technology Thonburi and National Center for Genetic Engineering and Biotechnology (BIOTEC) for supporting on scholarships and resources that facilitate me until the thesis have been completed. In addition, I would like to thank Development and Promotion of Science and Technology Talents Project (DPST) for financial support as well as a great opportunity to study and research.

Lastly, I would like to thank all of my friends here who have encouraged and supported me along the way until I finally reached my destination now.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF TECHNICAL VOCABULARIES AND ABBREVIATIONS

BaEV	Baboon Endogenous Virus
CCDS	Consensus Coding Sequence
DAVID	Database for Annotation, Visualization and Integrated Discovery
DNA	Deoxyribonucleic Acid
ERV	Endogenous Retroviruses
GEO	Gene Expression Omnibus database
GIRI	Genetic Information Research Institute
GO	Gene Ontology
HERV	Human Endogenous Retrovirus
HERVd	Human Endogenous Retroviruses Database
ID	Identifier
LTR	Long Terminal Repeat
MMTV	Mouse Mammary Tumor Virus
MaLR	Mammalian apparent LTR-retrotransposon
mRNA	messenger Ribonucleic Acid
MuLV	Murine Leukemia Virus
NCBI	National Center for Biotechnology Information
ORF	Open Reading Frame
RefSeq	Reference Sequence Database
RNA	Ribonucleic Acid
RU	Repbase Update
SLE	Systemic Lupus Erythematosus
TE	Transposable Element
tRNA	transfer Ribonucleic Acid

SWISS-PROT	Swiss Institute of Bioinformatics-Protein Database
UCSC	University of California, Santa Cruz
UniProt	Universal Protein Resource