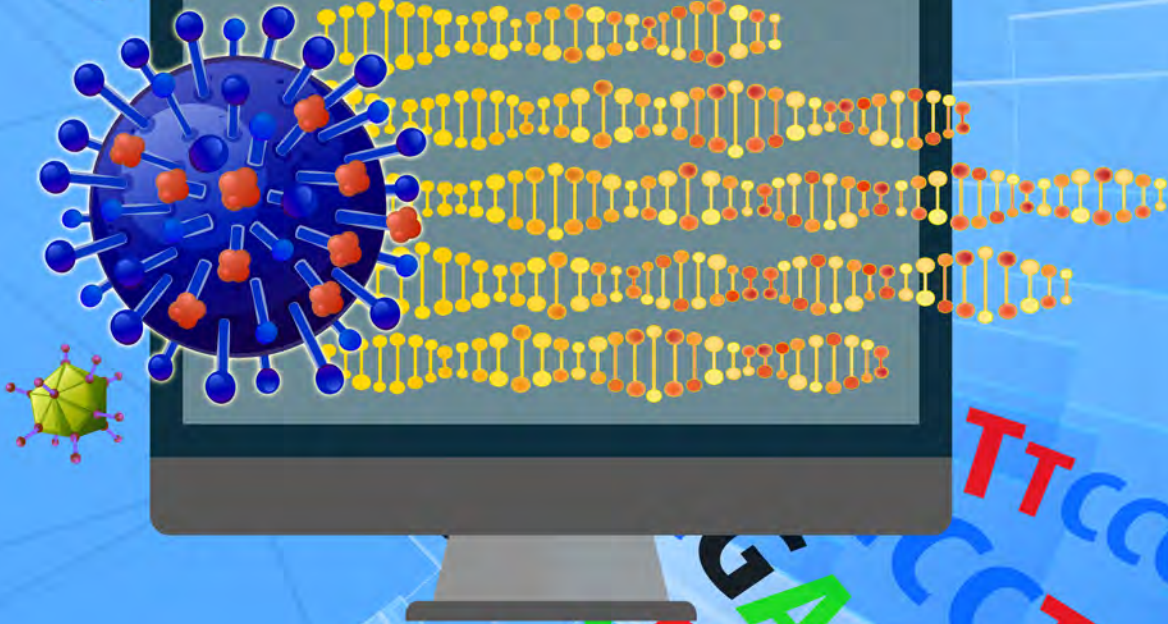
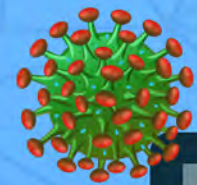




# อณูไวรัสวิทยาเชิงคอมพิวเตอร์

## Computational Molecular Virology



รศ.ดร.สัญญาชัย พยุงภร  
ภาควิชาชีวเคมี คณะแพทยศาสตร์  
จุฬาลงกรณ์มหาวิทยาลัย

TTCCCCCTAT  
GACCTATCTC  
TCGACTACGT

# อณูไวรัสวิทยาเชิงคอมพิวเตอร์

## Computational Molecular Virology

รองศาสตราจารย์ ดร.สัณชัย พยุงภร

Associate Professor Sunchai Payungporn, Ph.D.

ภาควิชาชีวเคมี คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

Department of Biochemistry, Faculty of Medicine

Chulalongkorn University

อณูไวรัสวิทยาเชิงคอมพิวเตอร์ (Computational Molecular Virology)

เลขมาตรฐานสากลประจำหนังสือ ISBN 978-616-468-907-7

ผู้นิพนธ์: รองศาสตราจารย์ ดร.สัญญาชัย พยุงภร

ภาควิชาชีวเคมี คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

พิมพ์ครั้งที่ 1 จำนวน 100 เล่ม พ.ศ. 2561

พิมพ์ที่: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย กรุงเทพฯ โทร 02-218-3562

พิสูจน์อักษร: นางสาวอริยา ขำวุฒิ

จัดรูปเล่ม: นายวรรณ สวัสดิ์วงศ์

ออกแบบปก: นายวรรณ สวัสดิ์วงศ์

The graphics on the cover of this book were designed by brgfx and Haryarts available on Freepik website.

### ข้อมูลทางบรรณานุกรมของสำนักหอสมุดแห่งชาติ

รองศาสตราจารย์ ดร.สัญญาชัย พยุงภร

อณูไวรัสวิทยาเชิงคอมพิวเตอร์ (Computational Molecular Virology)

โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย 2561, 245 หน้า

1. ฐานข้อมูลทางด้านชีวภาพ
2. ฐานข้อมูลทางด้านไวรัสวิทยา
3. การวิเคราะห์ข้อมูลลำดับนิวคลีโอไทด์
4. เครื่องมือในการค้นหาเปรียบเทียบความคล้ายคลึงกันทางชีวภาพ
5. การวิเคราะห์ความสัมพันธ์เชิงวิวัฒนาการ
6. เทคนิคที่เกี่ยวข้องกับพีซีอาร์ในด้านอณูไวรัสวิทยา
7. การออกแบบไพรเมอร์ด้วยคอมพิวเตอร์
8. เทคนิคอาร์เอ็นเอไอในด้านอณูไวรัสวิทยา

ISBN 978-616-468-907-7



<< ดาวโหลดหนังสืออิเล็กทรอนิกส์ฟรี

<< Free E-book is available for download

หนังสือเล่มนี้จัดพิมพ์เพื่อประโยชน์ทางการศึกษา ห้ามจำหน่าย

This book is published for academic purpose. Not for sale.

## คำนำ

หนังสือเล่มนี้ถูกจัดทำขึ้น โดยมีวัตถุประสงค์ในการเผยแพร่ความรู้ทางวิชาการ และการแนะแนวทางการฝึกทักษะทางด้านคอมพิวเตอร์ในการศึกษา และการทำวิจัยที่เกี่ยวข้องกับอนุชีววิทยาของเชื้อไวรัส โดยประกอบด้วยรายละเอียดของหัวข้อต่างๆ ได้แก่ ฐานข้อมูลทางด้านชีวภาพ ฐานข้อมูลทางด้านไวรัสวิทยา การวิเคราะห์ข้อมูลลำดับนิวคลีโอไทด์ การวิเคราะห์ความสัมพันธ์เชิงวิวัฒนาการ เทคนิคด้านอนุไวรัสวิทยาที่เกี่ยวข้องกับ PCR และ RNAi รวมทั้งการออกแบบไพรเมอร์ด้วยคอมพิวเตอร์ เป็นต้น ดังนั้นหนังสือเล่มนี้จึงเหมาะสำหรับนักศึกษาในระดับปริญญาตรี โท และเอก รวมทั้งนักวิจัยในสาขาวิทยาศาสตร์สุขภาพ วิทยาศาสตร์การแพทย์ เทคโนโลยีชีวภาพ ชีวเคมี อนุชีววิทยา จุลชีววิทยา และไวรัสวิทยา โดยหวังเป็นอย่างยิ่งว่าผู้อ่านจะนำความรู้ และทักษะในการวิเคราะห์ข้อมูลทางด้านอนุชีววิทยาของเชื้อไวรัสด้วยการใช้คอมพิวเตอร์ไปใช้ในการต่อยอดงานวิจัยและสามารถแนะนำผู้ที่เกี่ยวข้องได้

ผู้นิพนธ์ขอขอบคุณ ศูนย์เชี่ยวชาญเฉพาะทางด้านไวรัสวิทยาคลินิก ศูนย์เชี่ยวชาญเฉพาะทางด้านโรคตับอักเสบและมะเร็งตับ และ ศูนย์เชี่ยวชาญเฉพาะทางด้านชีววิทยาเชิงระบบ คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ให้โอกาสในการศึกษาวิจัย ทำให้สามารถรวบรวมความรู้และประสบการณ์ตลอดระยะเวลา 15 ปี ในงานวิจัยที่เกี่ยวข้องกับอนุชีววิทยาของเชื้อไวรัสอุบัติใหม่และอุบัติซ้ำในประเทศไทย เพื่อเรียบเรียงขึ้นเป็นหนังสือเล่มนี้ สำหรับเนื้อหาในส่วนที่เป็นเว็บไซต์ โปรแกรม และฐานข้อมูล ซึ่งมีการอัปเดตอย่างต่อเนื่อง อาจทำให้เนื้อหาและรูปภาพแตกต่างไปจากข้อมูลที่แสดงในหนังสือ ผู้นิพนธ์จึงขอภัยในความคลาดเคลื่อนมา ณ ที่นี้ และหากมีโอกาสจะดำเนินการปรับปรุงข้อมูลในการพิมพ์ครั้งต่อไป อย่างไรก็ตามรูปภาพส่วนใหญ่ที่ใช้ประกอบในหนังสือ นำมาจากเว็บไซต์ โปรแกรม และฐานข้อมูลที่เป็นข้อมูลสาธารณะเพื่อประโยชน์ในการเรียนรู้และส่งเสริมทางวิชาการเท่านั้น มิได้มีผลประโยชน์เชิงพาณิชย์ โดยผู้นิพนธ์ไม่มีความประสงค์ในการจำหน่ายหนังสือเล่มนี้ แต่จะดำเนินการจัดพิมพ์จำนวนจำกัดและแจกจ่ายไปยังสถาบันทางการศึกษา ตลอดจนหน่วยงานวิจัยต่างๆ พร้อมทั้งจัดทำเป็นหนังสืออิเล็กทรอนิกส์เผยแพร่แก่ผู้อ่านทั่วไป เพื่อให้เกิดประโยชน์สูงสุดทางด้านวิชาการอย่างแท้จริง

รองศาสตราจารย์ ดร.สัณชัย พยุงกร



## คำนิยม

ความก้าวหน้าทางวิทยาศาสตร์และเทคโนโลยีเป็นไปอย่างรวดเร็ว ประกอบกับข้อมูลสารสนเทศเป็นจำนวนมาก จึงจำเป็นอย่างยิ่งที่นักวิทยาศาสตร์ นักวิจัย จำเป็นที่จะต้องรู้วิธีการเข้าถึงข้อมูลจำนวนมากได้อย่างถูกต้อง รวดเร็ว และสามารถวิเคราะห์ดังกล่าว เพื่อที่จะใช้ต่อยอดในการสร้างเป็นองค์ความรู้ใหม่ หรือประยุกต์ใช้ให้เกิดประโยชน์ต่อไป

หนังสือเรื่องอนุไวรัสวิทยาเชิงคอมพิวเตอร์ เป็นหนังสือที่ได้รวบรวมเนื้อหาทางด้านอนุชีววิทยาของเชื้อไวรัสไว้อย่างสมบูรณ์และครบถ้วน ได้แก่ ฐานข้อมูลทางด้านชีวภาพ ฐานข้อมูลทางด้านไวรัสวิทยา การวิเคราะห์ข้อมูลลำดับนิวคลีโอไทด์ การวิเคราะห์ความสัมพันธ์เชิงวิวัฒนาการ เทคนิคด้านอนุไวรัสวิทยาที่เกี่ยวข้องกับ PCR และ RNAi รวมไปถึงการออกแบบไพรเมอร์ด้วยคอมพิวเตอร์ โดยเนื้อหาในหนังสือเล่มนี้ เกิดจากการสั่งสมประสบการณ์หลายปีในการสอนและการศึกษาวิจัยของ รองศาสตราจารย์ ดร.สัญญา พยุงภร ซึ่งเป็นผู้ที่มีความมุ่งมั่น ตั้งใจในการทำงานอย่างต่อเนื่อง สุดท้ายนี้จึงหวังเป็นอย่างยิ่งว่าหนังสือเล่มนี้จะเป็นประโยชน์สูงสุดทางด้านวิชาการแก่ นิสิต นักศึกษา นักวิทยาศาสตร์ และนักวิจัย ในสถาบันการศึกษาและสถาบันวิจัยต่างๆ

ศาสตราจารย์ นายแพทย์ ยง ภู่วรวรรณ

หัวหน้าศูนย์เชี่ยวชาญเฉพาะทางด้านไวรัสวิทยาคลินิก

คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

## สารบัญ

	หน้า
<b>บทที่ 1</b>	
<b>ฐานข้อมูลทางด้านชีวภาพ (Biological Databases)</b>	1
ฐานข้อมูล NCBI	2
ฐานข้อมูล PubMed	5
ฐานข้อมูล Genome	12
ฐานข้อมูล Nucleotide	14
ฐานข้อมูล Protein	18
ฐานข้อมูล Structure	21
ฐานข้อมูล Taxonomy	26
ฐานข้อมูล OMIM	27
<b>บทที่ 2</b>	
<b>ฐานข้อมูลทางด้านไวรัสวิทยา (Virological Databases)</b>	31
ฐานข้อมูล ICTV	32
ฐานข้อมูล ViralZone	33
ฐานข้อมูล ViPR	34
ฐานข้อมูล VIPERdb	34
ฐานข้อมูล IEDB	36
ฐานข้อมูล VirHostnet	37
ฐานข้อมูล NCBI-VVR	38
ฐานข้อมูลสำหรับเชื้อไวรัสตับอักเสบบี (Databases for hepatitis B virus)	42
ฐานข้อมูลสำหรับเชื้อไวรัสตับอักเสบซี (Databases for hepatitis C virus)	45
ฐานข้อมูลสำหรับเชื้อไวรัสเอชไอวี (Databases for HIV)	47
ฐานข้อมูลสำหรับเชื้อไวรัสปาปิลโลมา (Databases for papillomavirus)	49
ฐานข้อมูลสำหรับเชื้อไวรัสไข้หวัดใหญ่ (Databases for influenza virus)	51

สารบัญ (ต่อ)

	หน้า
<b>บทที่ 3</b>	
<b>การวิเคราะห์ข้อมูลลำดับนิวคลีโอไทด์ (Nucleotide sequence analysis)</b>	57
เทคนิคการหาลำดับสารพันธุกรรม (DNA sequencing)	58
การนำข้อมูลลำดับนิวคลีโอไทด์มาเชื่อมต่อกัน (Sequence assembly)	62
การจัดเรียงลำดับนิวคลีโอไทด์ (Nucleotide sequences alignment)	63
โปรแกรมในการวิเคราะห์ลำดับนิวคลีโอไทด์ และลำดับกรดอะมิโน	68
<b>บทที่ 4</b>	
<b>เครื่องมือในการค้นหาเปรียบเทียบความคล้ายคลึงกันทางชีวภาพ (BLAST)</b>	91
Nucleotide BLAST	94
Protein BLAST	99
Translated BLAST: BLASTx	103
Translated BLAST: tBLASTn	106
Translated BLAST: tBLASTx	109
การ BLAST แบบพิเศษ (Specialized BLAST)	112
<b>บทที่ 5</b>	
<b>การวิเคราะห์ความสัมพันธ์เชิงวิวัฒนาการ (Phylogenetic analysis)</b>	115
การวิเคราะห์ความสัมพันธ์เชิงวิวัฒนาการ (Phylogenetic analysis)	117
ประเภทของแผนภาพต้นไม้เชิงวิวัฒนาการ (Types of phylogenetic tree)	118
วิธีการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ	119
การใช้โปรแกรมทางคอมพิวเตอร์ในการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ	128

## สารบัญ (ต่อ)

	หน้า
<b>บทที่ 6</b>	
<b>เทคนิคพีซีอาร์ในด้านอณูไวรัสวิทยา (PCR techniques in molecular virology)</b>	<b>141</b>
การเพิ่มจำนวนสารพันธุกรรมด้วยเทคนิค PCR	142
เทคนิคที่มีพื้นฐานมาจาก PCR (PCR-based assay)	145
Reverse Transcription PCR (RT-PCR)	146
Nested PCR	147
Multiplex PCR	148
PCR-Restriction Fragment Length Polymorphism (PCR-RFLP)	149
Allele specific PCR	150
Universal PCR	151
Real-time PCR	153
Melting curve analysis	161
Multiplex real-time PCR	165
Allelic discrimination assay	168
<b>บทที่ 7</b>	
<b>การออกแบบไพรเมอร์ด้วยคอมพิวเตอร์ (Computational based primers design)</b>	<b>173</b>
การออกแบบ thermal profiles สำหรับ PCR	174
หลักการออกแบบ primers	175
โปรแกรมคอมพิวเตอร์สำหรับการออกแบบ primers	179
การออกแบบ primers / probes จาก nucleotide sequences alignment	194
ฐานข้อมูลสำหรับ primers และ probes	204

สารบัญ (ต่อ)

	หน้า
<b>บทที่ 8</b>	
<b>เทคนิคอาร์เอ็นเอไอในด้านอณูไวรัสวิทยา (RNAi technique in molecular virology)</b>	<b>209</b>
ประวัติการค้นพบกระบวนการ RNA interference (RNAi)	210
กลไกของกระบวนการ RNA interference (RNAi)	210
วิวัฒนาการร่วมของกระบวนการ RNAi ใน eukaryotes	213
เทคโนโลยีในการสร้าง siRNA	213
การเปรียบเทียบเทคโนโลยีการสร้าง siRNA	216
การออกแบบ siRNA	217
การสร้าง shRNA expression vector	219
โปรแกรมคอมพิวเตอร์สำหรับการออกแบบ siRNA และ shRNA	220
โปรแกรมการออกแบบ siRNA/shRNA สำหรับเชื้อไวรัส	230
วิธีการขนส่ง siRNA เข้าสู่เซลล์	234
การประยุกต์ใช้เทคนิค RNAi ในการยับยั้งการเพิ่มจำนวนของเชื้อไวรัส	235
<b>ดัชนี (Index)</b>	<b>239</b>
<b>ประวัติผู้นิพนธ์ (Biography)</b>	<b>243</b>



# 1

ฐานข้อมูลทางด้านชีวภาพ

Biological Databases

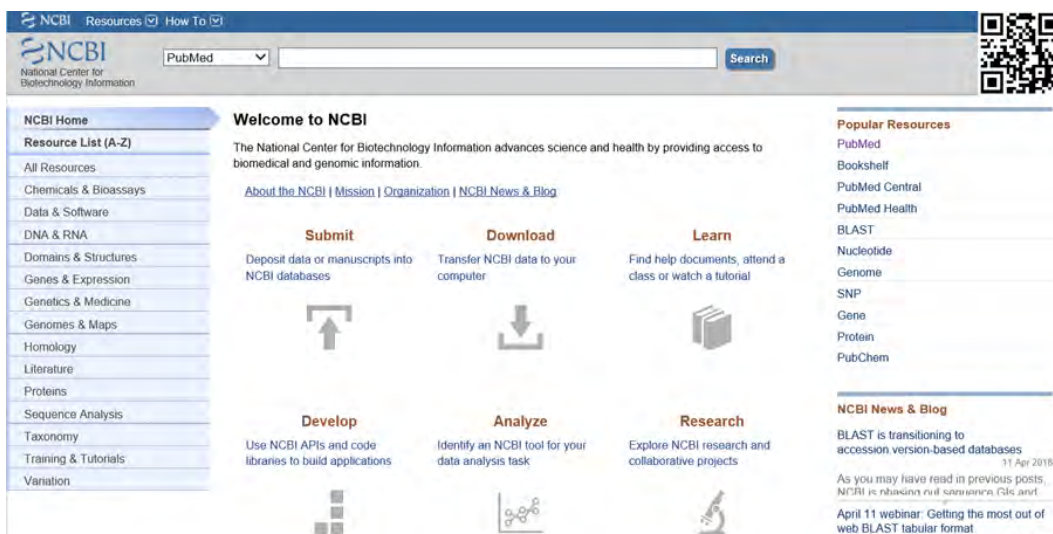
## ฐานข้อมูลทางด้านชีวภาพ

### Biological Databases

ในปัจจุบันฐานข้อมูลทางด้านชีวภาพมีความสำคัญเป็นอย่างมาก เนื่องจากเป็นแหล่งรวบรวมข้อมูลที่ได้จากการศึกษาวิจัย และการทดลองทางวิทยาศาสตร์สะสมมาตั้งแต่อดีตถึงปัจจุบัน ซึ่งการเก็บข้อมูลมหาศาลเหล่านี้ จะมีประโยชน์ในการเผยแพร่ให้นักวิจัย และนักวิทยาศาสตร์ทั่วโลกได้นำไปประยุกต์ใช้ต่อไป ในที่นี้จะกล่าวถึงเฉพาะฐานข้อมูลของ National Center for Biotechnology Information (NCBI) เนื่องจากมีข้อมูลที่มีประโยชน์มากมาย และเป็นที่ยอมรับมากที่สุด

**ฐานข้อมูล NCBI** [<https://www.ncbi.nlm.nih.gov>]

ฐานข้อมูล NCBI ก่อตั้งขึ้นในปี ค.ศ. 1988 จัดเป็นแผนกหนึ่งใน National Library of Medicine (NLM) ของ National Institutes of Health (NIH) ประเทศสหรัฐอเมริกา ซึ่งเป็นผู้นำทางด้านชีวสารสนเทศ วิทยาการคอมพิวเตอร์ และเทคโนโลยีชีวภาพ เมื่อเข้าสู่หน้าแรกของฐานข้อมูล NCBI (รูปที่ 1.1) จะประกอบไปด้วยส่วนต่างๆ ได้แก่



รูปที่ 1.1 หน้าแรกของฐานข้อมูล NCBI [<https://www.ncbi.nlm.nih.gov>]

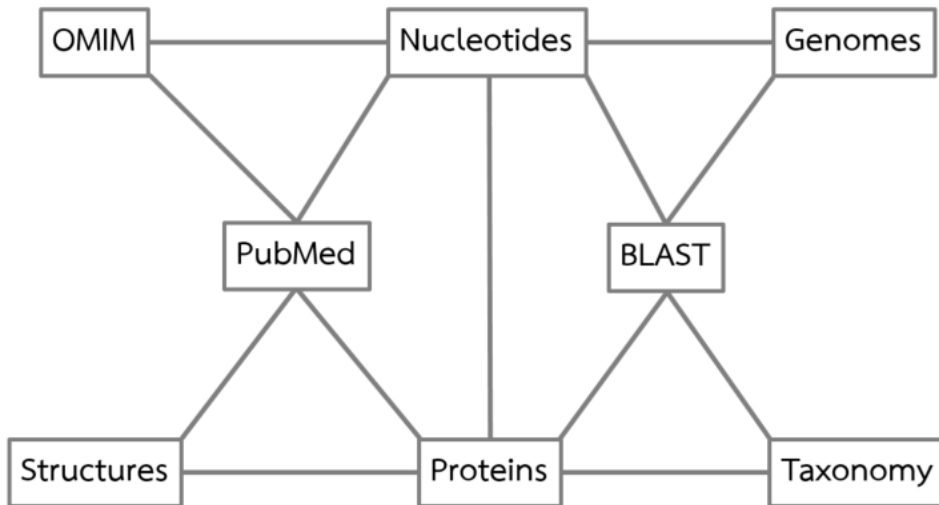
- Search เป็นการค้นหาในฐานข้อมูลโดยใส่คำสำคัญ (key words) ที่เราต้องการค้นหา
- Welcome to NCBI เป็นการแบ่งหมวดหมู่ของฐานข้อมูลตามลักษณะการใช้ ได้แก่
  - Submit ใช้ในกรณีที่ต้องการส่งข้อมูลเข้าสู่ NCBI database
  - Download ใช้ในกรณีที่ต้องการดาวน์โหลดข้อมูลจาก NCBI database
  - Learn ใช้เพื่อค้นหาสื่อการสอนสำหรับการใช้ NCBI database
  - Develop ใช้เพื่อค้นหาโปรแกรมหรือ code ต่างๆ สำหรับการพัฒนา software
  - Analyze ใช้เพื่อค้นหาโปรแกรมหรือเครื่องมือ ในการวิเคราะห์ข้อมูล
  - Research ใช้เพื่อค้นหางานวิจัยของ NCBI และโครงการวิจัยร่วมต่างๆ
- Resource เป็นการแบ่งหมวดหมู่ตามกลุ่มของฐานข้อมูล ได้แก่ Chemicals & Bioassays, Data & Software, DNA & RNA, Domains & Structures, Genes & Expression, Genetics & Medicine, Genomes & Maps, Homology, Literature, Proteins, Sequence Analysis, Taxonomy, Training & Tutorials และ Variation เป็นต้น
- Popular Resources เป็นการแสดงฐานข้อมูลที่ได้รับความนิยมอย่างแพร่หลาย ได้แก่ PubMed, Bookshelf, PubMed Central, PubMed Health, BLAST, Nucleotide, Genome, SNP, Gene, Protein และ PubChem เป็นต้น
- NCBI News & Blog เป็นการแจ้งข่าวสารหรือการปรับปรุงใหม่ๆ จากทาง NCBI

เนื่องจาก NCBI database เป็นฐานข้อมูลที่มีขนาดใหญ่มาก และมีการรวบรวมข้อมูลทางด้านชีวภาพ และการแพทย์ จึงเหมาะสำหรับนักวิจัย และนักวิทยาศาสตร์ในแขนงต่างๆ ได้แก่ เคมี ชีวเคมี พันธุศาสตร์ อณูชีววิทยา จุลชีววิทยา วิทยาศาสตร์การแพทย์ เทคโนโลยีชีวภาพ และ พันธุวิศวกรรม เป็นต้น สำหรับในหนังสือเล่มนี้จะมุ่งเน้นเฉพาะฐานข้อมูลที่เกี่ยวข้องในการศึกษา อณูชีววิทยาของไวรัสดังนี้

- PubMed ฐานข้อมูลการอ้างอิงและบทคัดย่อสำหรับบทความวิจัยทางด้านวิทยาศาสตร์และการแพทย์ ซึ่งสามารถเชื่อมโยงไปสู่บทความฉบับเต็มในเว็บไซต์ของวารสารทางวิชาการระดับนานาชาติ
- Genomes ฐานข้อมูลที่ประกอบด้วยรหัสพันธุกรรมทั้งจีโนมของสิ่งมีชีวิตมากกว่า 1000 ชนิด
- Nucleotide ฐานข้อมูลที่รวบรวมข้อมูลลำดับนิวคลีโอไทด์ ของสิ่งมีชีวิตจากหลายแหล่งข้อมูล ได้แก่ GenBank, RefSeq, TPA และ PDB
- Protein ฐานข้อมูลที่รวบรวมข้อมูลโปรตีนจากหลายแหล่งข้อมูล ได้แก่ GenPept, RefSeq, Swiss-Prot, PIR, PRF และ PDB
- Structure ฐานข้อมูลโครงสร้างสามมิติของสารชีวโมเลกุล ที่รวบรวมจาก Protein Data Bank (PDB) และประกอบด้วยเครื่องมือสำหรับการจำลองภาพสามมิติและการวิเคราะห์เปรียบเทียบโครงสร้าง
- Taxonomy ฐานข้อมูลชื่อและอนุกรมวิธานของสิ่งมีชีวิตมากกว่า 160,000 ชนิดที่มีข้อมูลโมเลกุลในฐานข้อมูล NCBI
- OMIM ฐานข้อมูลของยีนและความผิดปกติทางพันธุกรรมของมนุษย์
- BLAST (Basic Local Alignment Search Tool) เป็นเครื่องมือที่ใช้ในการค้นหาส่วนของยีนหรือโปรตีนที่มีความคล้ายคลึงกัน โดยโปรแกรมจะเปรียบเทียบลำดับนิวคลีโอไทด์หรือกรดอะมิโนกับฐานข้อมูลและคำนวณนัยสำคัญทางสถิติของความเหมือน จึงสามารถใช้เพื่อหาความสัมพันธ์ระหว่างบทบาทหน้าที่และวิวัฒนาการของยีนในสิ่งมีชีวิต

ในการใช้งานแต่ละฐานข้อมูลย่อยที่อยู่ใน NCBI database จะมีความเชื่อมโยงกัน (รูปที่ 1.2) กล่าวคือ เมื่อเราค้นหาบทความวิชาการในฐานข้อมูล PubMed และในบทความนั้นๆ มีการรายงานข้อมูลเกี่ยวกับลำดับนิวคลีโอไทด์ของเชื้อไวรัส ก็จะมี link ให้กดเมื่อเข้าไปดูข้อมูลลำดับนิวคลีโอไทด์ที่เกี่ยวข้องได้

ในทางกลับกันถ้าเราค้นหาข้อมูลลำดับนิวคลีโอไทด์ของเชื้อไวรัส ก็จะมี link ให้กดเข้าไปดูว่าลำดับนิวคลีโอไทด์ดังกล่าวได้มาจากบทความวิจัยเรื่องใด ดังนั้นจึงทำให้ผู้ใช้ฐานข้อมูลมีความสะดวกสบายในการค้นคว้าข้อมูลอย่างรวดเร็ว และมีการอ้างอิงแหล่งที่มาได้อย่างเหมาะสม ในลำดับต่อไปจะเป็นรายละเอียดการใช้งานฐานข้อมูลต่างๆ ใน NCBI database



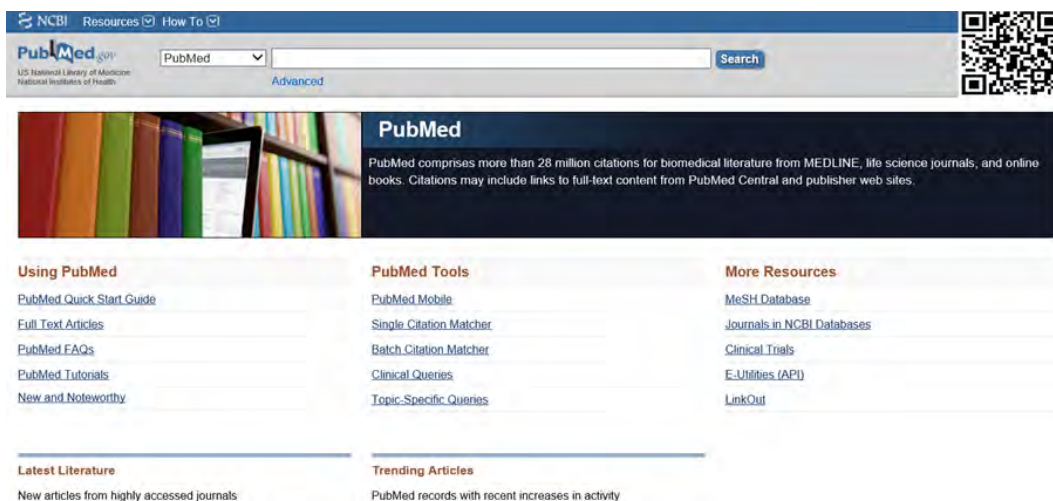
รูปที่ 1.2 แผนผังความเชื่อมโยงระหว่างฐานข้อมูลต่างๆ ใน NCBI ที่นิยมใช้

### ฐานข้อมูล PubMed [<https://www.ncbi.nlm.nih.gov/pubmed>]

PubMed เป็นฐานข้อมูลบทความวิจัยทางด้านวิทยาศาสตร์และการแพทย์ ซึ่งสามารถเชื่อมโยงไปสู่บทความฉบับเต็มในเว็บไซต์ของวารสารทางวิชาการระดับนานาชาติได้ จากหน้าแรกของ NCBI website (รูปที่ 1.1) เมื่อคลิกที่คำว่า PubMed (บริเวณด้านขวาภายใต้หัวข้อ Popular Resources) จะเชื่อมโยงไปยังฐานข้อมูล PubMed (รูปที่ 1.3) ทั้งนี้สามารถพิมพ์คำสำคัญที่ต้องการค้นหาใส่ลงไปในส่วนช่องว่างด้านบน จากนั้นคลิกที่คำว่า Search ก็จะเป็นการค้นหาบทความที่เกี่ยวข้องได้ ยกตัวอย่างเช่น หากพิมพ์คำว่า hepatitis B virus แล้วคลิกที่คำว่า Search ก็จะได้ผลดังรูปที่ 1.4



## ฐานข้อมูลทางด้านชีวภาพ



### รูปที่ 1.3 ฐานข้อมูล PubMed [https://www.ncbi.nlm.nih.gov/pubmed]

#### Search results

Items: 1 to 20 of 46486

<< First < Prev Page 1 of 2325 Next > Last >>

- [Antiviral prophylaxis for preventing reactivation of hepatitis B virus in rheumatic patients: a systematic review and meta-analysis.](#)
  1. [systematic review and meta-analysis.](#)  
Su J, Long L, Zou K.  
Clin Rheumatol. 2018 Apr 10. doi: 10.1007/s10067-018-4096-0. [Epub ahead of print]  
PMID: 29637482
- [Perinatal management of hepatitis B virus: Clinical implementation of updated Australasian management guidelines.](#)
  2. [management guidelines.](#)  
Bergin H, Wood G, Walker SP, Hui L.  
Obstet Med. 2018 Mar;11(1):23-27. doi: 10.1177/1753495X17731990. Epub 2017 Oct 25.  
PMID: 29636810
- [HIV and viral hepatitis coinfection analysis using surveillance data from 15 US states and two cities.](#)
  3. [HIV and viral hepatitis coinfection analysis using surveillance data from 15 US states and two cities.](#)  
Bosh KA, Coyle JR, Hansen V, Kim EM, Speers S, Comer M, Maddox LM, Khuwaja S, Zhou W, Jatta A, Mayer R, Brantley AD, Muriithi NW, Bhattacharjee R, Flynn C, Bouton L, John B, Keusch J, Barber CA, Sweet K, Ramaswamy C, Westheimer EF, VanderBusch L, Nishimura A, Vu A, Hoffman-Arriaga L, Rowlinson E, Carter AO, Yerkes LE, Li W, Reuer JR, Stockman LJ, Tang T, Brooks JT, Teshale EH, Hall HI.  
Epidemiol Infect. 2018 Apr 11:1-11. doi: 10.1017/S0950268818000766. [Epub ahead of print]  
PMID: 29636119
- [A LC/MS/MS method for determination of tenofovir in human plasma and its application to toxicity monitoring.](#)
  4. [monitoring.](#)  
Wiryakosol N, Puangpetch A, Manosuthi W, Tomongkon S, Sukasem C, Pinthong D.  
J Chromatogr B Analyt Technol Biomed Life Sci. 2018 Mar 29;1085:89-95. doi: 10.1016/j.jchromb.2018.03.045. [Epub ahead of print]  
PMID: 29635209  
[Similar articles](#)

### รูปที่ 1.4 ตัวอย่างผลการค้นหาคำใน PubMed โดยใช้คำว่า hepatitis B virus

จากรูปที่ 1.4 เป็นผลที่ได้จากการค้นหาบทความที่เกี่ยวข้องกับคำว่า hepatitis B virus ในฐานข้อมูล PubMed จะพบว่ามีความเกี่ยวข้องจำนวนมาก (46,486 เรื่อง) หากดูรายละเอียดเพิ่มเติมในรูปที่ 1.4 จะพบว่าบทความเรื่องที่ 1 และ 2 นั้นมีคำว่า hepatitis B virus ปรากฏอยู่ในชื่อเรื่อง ในขณะที่เรื่องที่ 3 พบคำว่า viral hepatitis ไม่ใช่ hepatitis B virus (อาจจะเป็น hepatitis virus ชนิดอื่น) หรือเรื่องที่ 4 จะไม่พบคำว่า hepatitis B virus ในชื่อเรื่องเลย แสดงให้เห็นว่าการใส่คำสำคัญ (key word) เพียงคำเดียวเพื่อค้นหาบทความใน PubMed นั้นคำสำคัญดังกล่าวไม่จำเป็นต้องปรากฏอยู่ในชื่อเรื่องของบทความเท่านั้น อาจจะปรากฏอยู่ในส่วนอื่นๆ ของบทความก็ได้ เช่น บทคัดย่อ บทนำ เอกสารอ้างอิง เป็นต้น ซึ่งจะทำให้ได้จำนวนบทความที่เกี่ยวข้องมากเกินไป ไม่จำเพาะเจาะจง ดังนั้นในทางปฏิบัติจึงควรใช้คำสำคัญหลายๆ คำพร้อมกันเพื่อให้การค้นหามีความจำเพาะและตรงประเด็นกับเรื่องที่เราต้องการ นอกจากนี้เรายังสามารถระบุได้ว่าคำสำคัญแต่ละคำที่ใส่ลงไปในนั้นต้องการให้ค้นหาในส่วใดของบทความ ซึ่งจะเรียกว่าการค้นหาขั้นสูง (advanced search) จากรูปที่ 1.3 เมื่อคลิกที่คำว่า Advanced (อยู่ใต้ช่องว่างที่ให้ใส่คำสำคัญ) จะได้ผลดังรูปที่ 1.5

### ➤ การค้นหาบทความในฐานข้อมูล PubMed ด้วยวิธี Advanced search

การค้นหาบทความในฐานข้อมูล PubMed ด้วยวิธี Advanced search นั้นจำเป็นที่จะต้องระบุขอบเขตในการค้นหาคำสำคัญว่าจะต้องการหาคำนั้นๆ ในส่วใดของบทความ เช่น ค้นหาเฉพาะในส่วนชื่อเรื่อง ชื่อวารสาร ชื่อผู้พิมพ์ ปีที่ตีพิมพ์ ประเภทของบทความ เป็นต้น ซึ่งขอบเขตของการค้นหาที่สำคัญและนิยมใช้ในการค้นหาบทความสรุป ดังตารางที่ 1

นอกจากการกำหนดขอบเขตของการค้นหาแล้ว การค้นหาบทความด้วยวิธี Advanced search จำเป็นต้องใส่คำสำคัญหลายๆ คำเพื่อใช้ในการค้นหาร่วมกัน จึงจำเป็นต้องใช้ตัวเชื่อมโยงเพื่อสร้างเงื่อนไขในการค้นหา ได้แก่

- AND (และ) เป็นการระบุเงื่อนไขการค้นหาว่าจะต้องพบทั้ง 2 คำจึงจะแสดงผล เช่น ใส่คำว่า Payungporn AND hepatitis virus ระบบการค้นหาก็จะแสดงเฉพาะบทความที่มีคำทั้ง 2 คำปรากฏอยู่ในบทความ

- OR (หรือ) เป็นการระบุเงื่อนไขการค้นหาว่าถ้าพบคำใดคำหนึ่ง ก็จะแสดงผล เช่น ใส่คำว่า Payungporn OR hepatitis virus ระบบการค้นหาก็จะแสดงบทความที่มีคำใดคำหนึ่งปรากฏอยู่ในบทความก็ได้
- NOT (ไม่) เป็นการระบุเงื่อนไขการค้นหาว่าให้พบคำหนึ่งเท่านั้น แต่ต้องไม่พบอีกคำหนึ่ง จึงจะแสดงผล เช่น ใส่คำว่า Payungporn NOT hepatitis virus ระบบการค้นหาก็จะแสดงบทความที่มีเฉพาะคำว่า Payungporn แต่จะต้องไม่มีคำว่า hepatitis virus ปรากฏอยู่ในบทความ

**ตารางที่ 1** การกำหนดขอบเขตของการค้นหาบทความด้วยวิธี advanced search

ขอบเขตของการค้นหา	ความหมาย	ตัวอย่างคำสำคัญที่ใช้
Affiliation	ชื่อสถาบันของทีมนักวิจัย	Chulalongkorn university
Author	ชื่อสกุลของผู้นิพนธ์	Payungporn
Author-First*	ชื่อสกุลของผู้นิพนธ์ลำดับแรก	Payungporn
Author-Last*	ชื่อสกุลของผู้นิพนธ์ลำดับสุดท้าย	Payungporn
Date-Publication	วัน/เดือน/ปี ที่ตีพิมพ์บทความ	2015-2018
Journal	ชื่อวารสาร	J Virol
Language	ภาษาที่ใช้ในบทความ	English
Publication Type	ประเภทของบทความ	Review
Title	ชื่อเรื่องบทความวิจัย	Hepatitis
Title/Abstract	ชื่อเรื่องหรือบทคัดย่อ	Hepatitis

\* หมายเหตุ: Author-First ผู้นิพนธ์ลำดับแรก จัดเป็นผู้นิพนธ์หลักที่มีส่วนร่วมในการทำวิจัยมากที่สุด ส่วน Author-Last ผู้นิพนธ์ลำดับสุดท้าย มักจะเป็น corresponding author ที่ควบคุมและกำกับงานวิจัย

NCBI Resources How To  
PubMed Home More Resources Help

PubMed Advanced Search Builder

((hepatitis B virus[Title]) AND "Nature"[Journal]) AND ("2013"[Date - Publication] : "2018"[Date - Publication])

Edit

Builder

Title hepatitis B virus Show index list

AND Journal Nature Show index list

AND Date - Publication 2013 to 2018 Show index list

AND All Fields Show index list

Search or Add to history

### รูปที่ 1.5 การค้นหาบทความวิจัยในฐานข้อมูล PubMed ด้วยวิธี Advanced search

[<https://www.ncbi.nlm.nih.gov/pubmed/advanced>]

#### ➤ ตัวอย่างการค้นหาบทความในฐานข้อมูล PubMed ด้วยวิธี Advanced search

ตัวอย่างที่ 1 ถ้าต้องการค้นหาบทความเกี่ยวกับ hepatitis B virus ที่ตีพิมพ์ในวารสารวิชาการ Nature ในช่วง 5 ปีที่ผ่านมา สามารถค้นหาได้ (รูปที่ 1.5) ดังนี้

- คำสำคัญที่ 1 ใช้คำว่า hepatitis B virus โดยระบุให้ค้นหาเฉพาะชื่อเรื่อง (Title)
- คำสำคัญที่ 2 ใช้คำว่า Nature โดยระบุให้ค้นหาเฉพาะชื่อวารสาร (Journal)
- คำสำคัญที่ 3 ใช้คำว่า 2013 to 2018 โดยระบุให้ค้นหาเฉพาะปีที่ตีพิมพ์ (Date-Publication)
- โดยใช้เงื่อนไข AND ในการเชื่อมคำสำคัญทั้ง 3 คำ เข้าด้วยกัน เพื่อสร้างเงื่อนไขในการค้นหาให้มีความจำเพาะ จากผลจากการค้นหาจะพบว่ามีความเพียง 1 เรื่องเท่านั้นที่ตรงประเด็นตามเงื่อนไข

ตัวอย่างที่ 2 ถ้าต้องการค้นหาบทความเกี่ยวกับ hepatitis B virus ที่ตีพิมพ์จากสถาบันต่างๆ ในประเทศไทย ยกเว้นจุฬาลงกรณ์มหาวิทยาลัย สามารถค้นหาได้ดังนี้

- คำสำคัญที่ 1 ใช้คำว่า hepatitis B virus โดยระบุให้ค้นหาเฉพาะชื่อเรื่อง (Title)
- คำสำคัญที่ 2 ใช้คำว่า Thailand โดยระบุให้ค้นหาเฉพาะชื่อสถาบัน (Affiliation)

- คำสำคัญที่ 3 ใช้คำว่า Chulalongkorn university โดยระบุให้ค้นหาเฉพาะชื่อสถาบัน (Affiliation)
- โดยใช้เงื่อนไข AND ในการเชื่อมคำสำคัญที่ 1 และ 2 เข้าด้วยกัน และใช้เงื่อนไข NOT ในการเชื่อมคำสำคัญที่ 2 และ 3 เพื่อสร้างเงื่อนไขในการค้นหาให้มีความจำเพาะ จากผลจากการค้นหาจะพบว่ามียุทธศาสตร์จำนวน 46 เรื่องที่ตรงประเด็นตามเงื่อนไข [ผลการค้นหาเมื่อวันที่ 12 เม.ย. 61]

ตัวอย่างที่ 3 ถ้าต้องการค้นหายุทธศาสตร์เกี่ยวกับ เชื้อไวรัสไข้หวัดใหญ่ สายพันธุ์ H5N1 หรือ H1N1 ที่ตีพิมพ์เป็นบทความประเภท review โดยมีผู้พิมพ์คือ Payungporn สามารถค้นหาได้ดังนี้

- คำสำคัญที่ 1 ใช้คำว่า H5N1 โดยระบุให้ค้นหาเฉพาะชื่อเรื่อง (Title)
- คำสำคัญที่ 2 ใช้คำว่า H1N1 โดยระบุให้ค้นหาเฉพาะชื่อเรื่อง (Title)
- คำสำคัญที่ 3 ใช้คำว่า review โดยระบุให้ค้นหาเฉพาะประเภทของบทความ (Publication Type)
- คำสำคัญที่ 4 ใช้คำว่า Payungporn โดยระบุให้ค้นหาเฉพาะชื่อผู้พิมพ์ (Author)
- โดยใช้เงื่อนไข OR ในการเชื่อมคำสำคัญที่ 1 และ 2 เข้าด้วยกัน และใช้เงื่อนไข AND ในการเชื่อมคำสำคัญที่ 2 และ 3 จากนั้นใช้เงื่อนไข AND ในการเชื่อมคำสำคัญที่ 3 และ 4 เพื่อสร้างเงื่อนไขในการค้นหาให้มีความจำเพาะมากขึ้น จากผลจากการค้นหาจะพบว่ามียุทธศาสตร์จำนวน 2 เรื่องที่ตรงประเด็นตามเงื่อนไข

### ➤ รูปแบบผลการค้นหายุทธศาสตร์จากฐานข้อมูล PubMed

การแสดงผล (format) จากการค้นหายุทธศาสตร์ที่ได้จากฐานข้อมูล PubMed มีหลายรูปแบบที่สำคัญ (รูปที่ 1.6) ได้แก่

- Summary เป็นรูปแบบสรุปที่ประกอบด้วย ชื่อเรื่อง ชื่อผู้พิมพ์ ชื่อวารสาร ปี/เล่ม/เลขหน้า และประเภทของบทความ โดยสามารถคลิก link เชื่อมโยงให้แสดงเป็นรูปแบบบทคัดย่อได้



- Summary (text) เป็นรูปแบบสรุปเฉพาะข้อความที่ประกอบด้วย ชื่อเรื่อง ชื่อผู้นิพนธ์ ชื่อวารสาร ปี/เล่ม/เลขหน้า และประเภทของบทความ แต่ไม่สามารถคลิก link เชื่อมโยงให้แสดงเป็นรูปแบบอื่นได้
- Abstract เป็นรูปแบบที่ประกอบด้วย ชื่อเรื่อง ชื่อผู้นิพนธ์ ชื่อสถาบัน ชื่อวารสาร ปี/เล่ม/เลขหน้า และบทคัดย่อ โดยสามารถคลิก link เชื่อมโยงไปหาบทความฉบับเต็มในเว็บไซต์ของวารสารทางวิชาการระดับนานาชาติได้
- Abstract (text) เป็นรูปแบบเฉพาะข้อความที่ประกอบด้วย ชื่อเรื่อง ชื่อผู้นิพนธ์ ชื่อสถาบัน ชื่อวารสาร ปี/เล่ม/เลขหน้า และบทคัดย่อ แต่ไม่สามารถคลิก link เชื่อมโยงให้แสดงเป็นรูปแบบอื่นได้

A) Format: Summary - Send to -

[Detection of oseltamivir sensitive/resistant strains of pandemic influenza A virus \(H1N1\) from patients admitted to hospitals in Thailand.](#)

Payungporn S, Poomipak W, Makkoch J, Rianthavorn P, Theamboonlers A, Poovorawan Y.  
*J Virol Methods.* 2011 Nov;177(2):133-9. doi: 10.1016/j.jviromet.2011.07.008. Epub 2011 Jul 28.  
 PMID: 21820468 [Free PMC Article](#)  
[Similar articles](#)

Similar articles  
 Cited by 2 PubMed Central articles  
 Related information

B) Format: Abstract - Send to -

[J Virol Methods.](#) 2011 Nov;177(2):133-9. doi: 10.1016/j.jviromet.2011.07.008. Epub 2011 Jul 28.

**Detection of oseltamivir sensitive/resistant strains of pandemic influenza A virus (H1N1) from patients admitted to hospitals in Thailand.**

[Payungporn S<sup>1</sup>, Poomipak W, Makkoch J, Rianthavorn P, Theamboonlers A, Poovorawan Y.](#)

**Author information**

**Abstract**  
 Oseltamivir has been used widely for prophylaxis or treatment during outbreaks of the pandemic influenza virus (H1N1) in several countries. The aim of this study was to develop a real-time RT-PCR (reverse transcription-polymerase chain reaction) to be applied for detection and monitoring of the oseltamivir resistant strains of this virus during three outbreaks (May 2009 to October 2010) in Thailand. The real-time RT-PCR assay for detecting H275Y proved highly specific for the pandemic influenza virus (H1N1) as no cross-amplification was detected with other respiratory viruses or human total RNA. The assay was also highly sensitive with a detection limit as low as 100 copies/μL for both wild-type and resistant strains. The performance of the assay was evaluated in terms of amplification efficiency (100%). The results obtained by real-time RT-PCR were in complete agreement with direct nucleotide sequencing. However, real-time RT-PCR provided more detail on the relative quantities of ratios between resistant and sensitive strains in each individual. The results revealed that four of 1288 (0.31%) patients were infected with the oseltamivir resistant strain. The number of patients infected by resistant strains was higher during the third (0.61%) and second (0.24%) waves than during the first (0%) outbreak. In conclusion, the real-time RT-PCR assay for H275Y detection is advantageous because it is specific, sensitive, and provides quantitative data. And it would be useful for large-scale testing and monitoring of oseltamivir resistant strains of the pandemic influenza A virus (H1N1).

PMID: 21820468 PMCID: [PMC3391650](#) DOI: [10.1016/j.jviromet.2011.07.008](#)

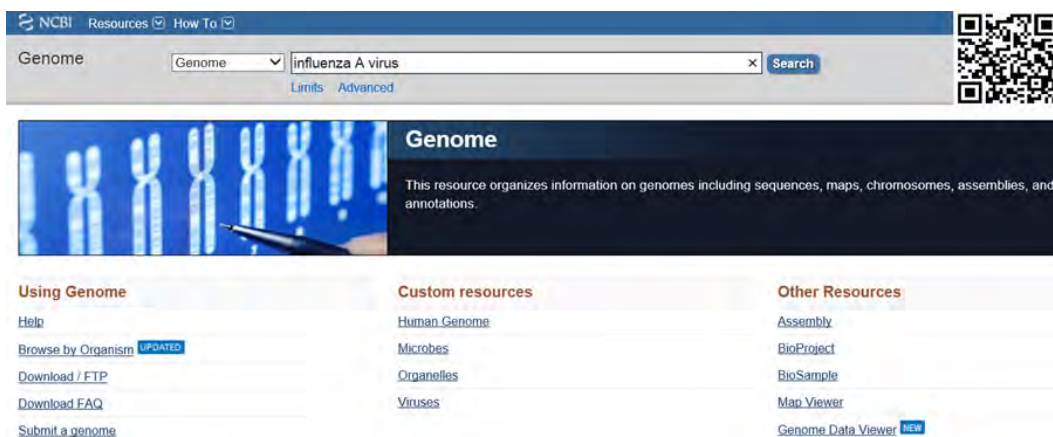
รูปที่ 1.6 รูปแบบ summary (A) และ abstract (B) ที่ได้จากการค้นหาบทความใน PubMed

ในหน้าผลการแสดงผลบทความใน PubMed ทางด้านขวา (รูปที่ 1.6) จะมี link ที่เชื่อมโยงไปสู่ข้อมูลต่างๆ ที่เกี่ยวข้องกับบทความที่กำลังสืบค้นอยู่ ซึ่งได้แก่

- Similar articles แสดงบทความอื่นๆ ที่มีเนื้อหาคล้ายๆ กับบทความที่กำลังสืบค้นอยู่
- Cited by PubMed Central แสดงบทความอื่นๆ ที่มีการอ้างอิงถึงบทความที่กำลังสืบค้นอยู่
- Related information แสดง link เชื่อมโยงไปสู่ข้อมูลอื่นๆ ที่เกี่ยวข้องกับบทความที่กำลังสืบค้นอยู่ เช่น Genome, Nucleotide, Protein, Structure, Taxonomy และ OMIM เป็นต้น

### ฐานข้อมูล Genome [https://www.ncbi.nlm.nih.gov/genome]

Genome เป็นฐานข้อมูลที่ประกอบด้วยรหัสพันธุกรรมทั้งจีโนมของสิ่งมีชีวิตมากกว่า 1,000 ชนิด ซึ่งประกอบไปด้วยข้อมูลด้านต่างๆ ได้แก่ รหัสพันธุกรรม โครโมโซม แผนที่ยีน และคุณลักษณะของยีน เป็นต้น การเข้าสู่ฐานข้อมูล Genome สามารถเข้าจากหน้าแรกของ NCBI website (รูปที่ 1.1) โดยคลิกที่คำว่า Genome (บริเวณด้านขวาภายใต้หัวข้อ Popular Resources) ก็จะเชื่อมโยงไปยังฐานข้อมูล Genome ได้ (รูปที่ 1.7)



รูปที่ 1.7 ฐานข้อมูล Genome [https://www.ncbi.nlm.nih.gov/genome]

Organism Overview ; Genome Assembly and Annotation report [7]

ID: 10290

## Influenza A virus

Influenza A virus overview

Lineage: Viruses[10431]; ssRNA viruses[2098]; ssRNA negative-strand viruses[546]; Orthomyxoviridae[10]; Alphainfluenzavirus[1]; Influenza A virus[1]

### NLM/NCBI 2009 H1N1 Flu Virus Resources:

- Newest [2009 H1N1 Flu \(Swine Flu\) sequences](#)
- Citations [recently added](#) to PubMed
- [MedlinePlus \(consumer health information\)](#)
- [Enviro-Health Links](#)

The Centers for Disease Control and Prevention (CDC) and other health organizations are actively investigating the recent outbreak of [2009 H1N1 Flu Virus \(Swine Flu\)](#) in humans. Cases in the U.S. were first reported in late March and early April 2009. CDC has determined that this swine influenza A (H1N1) virus is contagious and is spreading from human to human. Swine Influenza is a respiratory disease of pigs (swine) caused by type A influenza virus that regularly causes outbreaks of flu in pigs. Like all influenza viruses, swine flu viruses change constantly. Pigs can be infected by avian influenza and human influenza viruses as well as swine influenza viruses. When influenza viruses from different species infect pigs, the viruses can reassort (i.e., swap genes) and new viruses that are a mix of swine, human and/or avian influenza viruses can emerge. There are four main influenza type A virus subtypes that have been isolated in pigs: H1N1, H1N2, H3N2, and H3N1, but most of the recently isolated influenza viruses from pigs have been H1N1 viruses. While swine flu viruses do not normally infect humans, sporadic human infections with swine flu have occurred. Most commonly, these cases occur in persons with direct exposure to pigs; human-to-human transmission of swine flu can also occur, as is the case with the 2009 outbreak.

### Summary

**Sequence data:** genome assemblies: 7; sequence reads: 9  
**Statistics:** median total length (Mb): 0.013498  
 median protein count: 12  
 median GC%: 43.6638

### Publications

1. New genetic variants of influenza A(H1N1)pdm09 detected in Cuba during 2011-2013. Arencibia A, et al. Infect Genet Evol 2015 Jun
2. Host Adaptation and the Alteration of Viral Properties of the First Influenza A/H1N1pdm09 Virus Isolated in Japan. Aina A, et al. PLoS One 2015
3. Molecular genetic analysis of the Influenza A(H1N1)pdm09 virus from lethal and recovered cases in Russia from 2009 to 2014: Deletions in the nucleoprotein. Yatsyshina S, et al. Infect Genet Evol 2015 Aug

[More...](#)

### Representative (genome information for reference and representative genomes)

#### Reference genomes:

- [Influenza A virus \(A/New York/392/2004\(H3N2\)\)](#)

Submitter: The NIAID Influenza Genome Sequencing Consortium

Human Pathogen

Type	Name	RefSeq	INSDC	Size (Kb)	GC%	Protein	Gene
gment	segment 1	NC_007373.1	CY002071.1	2.34	42.5	1	1
gment	segment 2	NC_007372.1	CY002070.1	2.34	42.2	2	2
gment	segment 3	NC_007371.1	CY002069.1	2.23	41.7	2	2
gment	segment 4	NC_007366.1	CY002064.1	1.76	42.0	1	1
gment	segment 5	NC_007369.1	CY002067.1	1.57	45.8	1	1
gment	segment 6	NC_007368.1	CY002066.1	1.47	42.7	1	1
gment	segment 7	NC_007367.1	CY002065.1	1.03	46.7	2	2
gment	segment 8	NC_007370.1	CY002068.1	0.89	41.7	2	2

รูปที่ 1.8 ลักษณะผลที่ได้จากการสืบค้นจีโนมจากฐานข้อมูล Genome

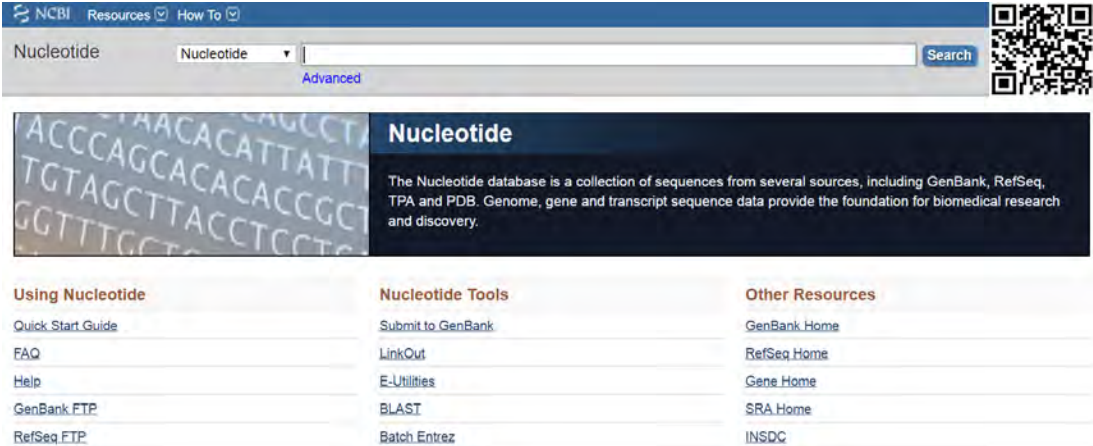
การค้นหาข้อมูลจีโนม สามารถพิมพ์คำที่ต้องการค้นหาใส่ลงไปในส่วนช่องว่างด้านบน หรือเลือกที่คำว่า Advanced เพื่อใช้ระบบ Advanced search ในการค้นหาข้อมูลจีโนมอย่างจำเพาะได้ นอกจากนี้สามารถเข้าฐานข้อมูล Genome โดยเชื่อมโยงมาจาก Related information ของบทความที่ค้นหาได้จากฐานข้อมูล PubMed ก็ได้ ตัวอย่างเช่นพิมพ์คำว่า influenza A virus แล้วคลิกที่คำว่า Search ก็จะได้ผลการค้นหาจีโนมดังรูปที่ 1.8 ผลที่ได้จากการค้นหา Genome ที่สำคัญ ได้แก่

- Lineage เป็นการระบุเชื้อสาย และอนุกรมวิธานของสิ่งมีชีวิต
- Summary เป็นข้อมูลสรุปเกี่ยวกับจีโนม เช่น ขนาดจีโนม จำนวนโปรตีน และ GC%
- Publications เป็นการอ้างอิงบทความที่เกี่ยวข้องกับจีโนมที่กำลังสืบค้น
- Representative เป็นข้อมูล reference genome (จีโนมต้นแบบของสิ่งมีชีวิตแต่ละชนิด) สำหรับใช้ในการเปรียบเทียบอ้างอิงต่อไป ซึ่งประกอบไปด้วยข้อมูล reference sequence (RefSeq) และ accession number (เลขรหัสของนิวคลีโอไทด์) ซึ่งสามารถเชื่อมโยงไปสู่ฐานข้อมูล Nucleotide ได้

### ฐานข้อมูล Nucleotide [<https://www.ncbi.nlm.nih.gov/nucleotide>]

Nucleotide เป็นฐานข้อมูลที่รวบรวมลำดับนิวคลีโอไทด์ของสิ่งมีชีวิตจากหลายแหล่งข้อมูล ได้แก่ GenBank, RefSeq, TPA และ PDB การเข้าสู่ฐานข้อมูล Nucleotide สามารถเข้าจากหน้าแรกของ NCBI website (รูปที่ 1.1) โดยคลิกที่คำว่า Nucleotide (บริเวณด้านขวาภายใต้หัวข้อ Popular Resources) ก็เชื่อมโยงไปยังฐานข้อมูล Nucleotide ได้ (รูปที่ 1.9) โดยสามารถพิมพ์คำสำคัญที่ต้องการค้นหาใส่ลงไปในส่วนช่องว่างด้านบน หรือเลือกที่คำว่า Advanced เพื่อใช้ระบบ Advanced search ในการค้นหาข้อมูลนิวคลีโอไทด์อย่างจำเพาะได้ นอกจากนี้สามารถเข้าฐานข้อมูล Nucleotide โดยเชื่อมโยงมาจาก Related information ของบทความที่ค้นหาได้จากฐานข้อมูล PubMed หรือเชื่อมโยงจาก RefSeq จากฐานข้อมูล Genome ก็ได้ การแสดงผลข้อมูลนิวคลีโอไทด์มีหลายรูปแบบ (รูปที่ 1.10 และ 11) ได้แก่

- GenBank เป็นการแสดงผลที่มีข้อมูลรายละเอียด ดังนี้ ชื่อสิ่งมีชีวิต ชื่อยีน Accession no. อนุกรมวิธาน บทความอ้างอิง แหล่งที่มาของตัวอย่าง ความยาวของยีน ส่วนของยีนที่ใช้สร้างโปรตีน ลำดับกรดอะมิโน และลำดับนิวคลีโอไทด์
- Summary เป็นการสรุปชื่อสิ่งมีชีวิต ชื่อยีน ความยาวของยีน และ Accession no.
- FASTA เป็นการสรุปชื่อสิ่งมีชีวิต ชื่อยีน Accession no. และลำดับนิวคลีโอไทด์



The screenshot shows the NCBI Nucleotide database search interface. At the top, there's a search bar with 'Nucleotide' selected and a 'Search' button. Below the search bar is a banner for the Nucleotide database with a background image of DNA sequences. Underneath the banner are three columns of links: 'Using Nucleotide' (Quick Start Guide, FAQ, Help, GenBank FTP, RefSeq FTP), 'Nucleotide Tools' (Submit to GenBank, LinkOut, E-Utilities, BLAST, Batch Entrez), and 'Other Resources' (GenBank Home, RefSeq Home, Gene Home, SRA Home, INSDC).

รูปที่ 1.9 ฐานข้อมูล Nucleotide [<https://www.ncbi.nlm.nih.gov/nucleotide>]

นอกจากนี้ยังสามารถเลือกคำสั่ง Send to (รูปที่ 1.11 ด้านขวาบน) เพื่อถ่ายโอนข้อมูลและบันทึกเข้าสู่คอมพิวเตอร์ของเราได้ ซึ่งเป็นวิธีที่เหมาะสมสำหรับการบันทึกข้อมูลหลายๆ ยีนพร้อมๆ กัน โดยมีรูปแบบให้เลือก ดังนี้

- Complete record บันทึกข้อมูลโดยละเอียดตามที่แสดงผลใน GenBank format
- Coding sequences บันทึกข้อมูลเป็นแบบ FASTA format ของลำดับนิวคลีโอไทด์หรือลำดับกรดอะมิโนที่สร้างมาจากยีนนั้นๆ
- Gene features บันทึกข้อมูลเป็นแบบ FASTA format

ในหน้าตาการแสดงผลนิวคลีโอไทด์แบบ summary (รูปที่ 1.11A) จะมี link ที่เชื่อมโยงไปสู่ข้อมูลต่างๆ ที่เกี่ยวข้อง ได้แก่ BioProject, Protein, PubMed และ Taxonomy เป็นต้น



ฐานข้อมูลทางด้านชีวภาพ

```

LOCUS      HM446344                1429 bp    cRNA    linear    VRL 12-OCT-2011
DEFINITION Influenza A virus (A/Thailand/CU-H88/2009(H1N1)) segment 6
            neuraminidase (NA) gene, complete cds.
ACCESSION  HM446344
VERSION    HM446344.1
DBLINK     BioProject: PEJNA37913
KEYWORDS   .
SOURCE     Influenza A virus (A/Thailand/CU-H88/2009(H1N1))
ORGANISM   Influenza A virus (A/Thailand/CU-H88/2009(H1N1))
            Viruses; ssRNA viruses; ssRNA negative-strand viruses;
            Orthomyxoviridae; Influenzavirus A.
REFERENCE  1 (bases 1 to 1429)
AUTHORS    Prachayangprocha,S., Makkoch,J., Payungporn,S., Chieochansin,T.,
            Vuthitanachot,C., Vuthitanachot,V., Theamboonlers,A. and
            Poovorawan,Y.
TITLE      Serological analysis of human pandemic influenza (H1N1) in Thailand
JOURNAL    J Health Popul Nutr 28 (6), 537-544 (2010)
PUBMED     21261190

FEATURES             Location/Qualifiers
     source           1..1429
                     /organism="Influenza A virus
                     (A/Thailand/CU-H88/2009(H1N1))"
                     /mol_type="viral cRNA"
                     /strain="A/Thailand/CU-H88/2009"
                     /serotype="H1N1"
                     /isolate="CU_H88_NA-1"
                     /isolation_source="nasopharyngeal swab"
                     /host="Homo sapiens"
                     /db_xref="taxon:794876"
                     /segment="6"
                     /country="Thailand"
                     /collection_date="Aug-2009"
                     /note="lineage: sw1;
                     passaged once in embryonated chicken eggs"
     gene             1..1410
                     /gene="NA"
     CDS              1..1410
                     /gene="NA"
                     /codon_start=1
                     /product="neuraminidase"
                     /protein_id="A0149821.1"
                     /translation="MNPNQKIITIGSVCMTIGMANLILQIGNIISIWISHSIQLGNQN
                     QIETCNQSVITYENNTWVNQTYVNI SNTNFAAGQSVVSVKLAGNSSLCPVSGWAIYSK
                     DNSIRIGSKGDVVFVIREPFISCSPLECRTFFLTQGALLNDKHSNGTIKDRSPYRTLMS
                     CPIGEVFPSPYNSRFESVAMSASACHDGINWLTIGISGPDNGAVAVLKYNGIITDTIKS
                     WRNNILRTQSEACVNGSCFTVMTDGPSPDQASYKIFRIEKGKIVKSVEMNAPNYHY
                     EECSCYPDSSEITCVCRDNWHGSRFPWVSFNQNLLEYQLGYLCSGLFGDNPRFPNDKTS
                     CGPVSSNGANGVKGFSFKYGNQVNIQRKTSISSRNGFEMINWDPNGWTGTDNNFSIKQD
                     TVGINWSGYSGSFVQHPELTGLDCIRPCFWVELIRGRPKENTWTSGSSISFCGVNS
                     DTVGWSWPDGAELFFTIK"

ORIGIN
1 atgatccaa accaaaagat aataaccatt ggttcggtct gtatgacaat tggaatggct
61 aacttaatat tacaaattgg aaacataatc tcaatatgga ttagccactc aattcaactt
121 ggggaatcaa atcagattga aacatgcaat caaagcgtca ttacttatga aaacaacact
181 tgggtaaate agacatatgt taacatcagc saacaccaact ttgctgctgg acagtcagtg
241 gtttcogtga aattagggg caattcctct otctgccctg ttagtggatg ggotatatac
301 agtaagaca acagtataag aatcggttcc aaggggatg tgtttgtcat aagggaaeca

```

รูปที่ 1.10 รูปแบบ GenBank format ในการแสดงผลนิวคลีโอไทด์จากฐานข้อมูล Nucleotide

A) Summary +

[Influenza A virus \(A/Thailand/CU-H88/2009\(H1N1\)\) segment 6 neuraminidase \(NA\) gene, complete cds](#)  
 1,429 bp linear cRNA  
 Accession: HM446344.1 GI: 297650472  
[BioProject](#) [Protein](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

Send to: -

- Complete Record
- Coding Sequences
- Gene Features

---

Choose Destination

- File
- Clipboard
- Collections
- Analysis Tool

B) Influenza A virus (A/Thailand/CU-H88/2009(H1N1)) segment 6 neuraminidase (NA) gene, complete cds

GenBank: HM446344.1

[GenBank](#) [Graphics](#)

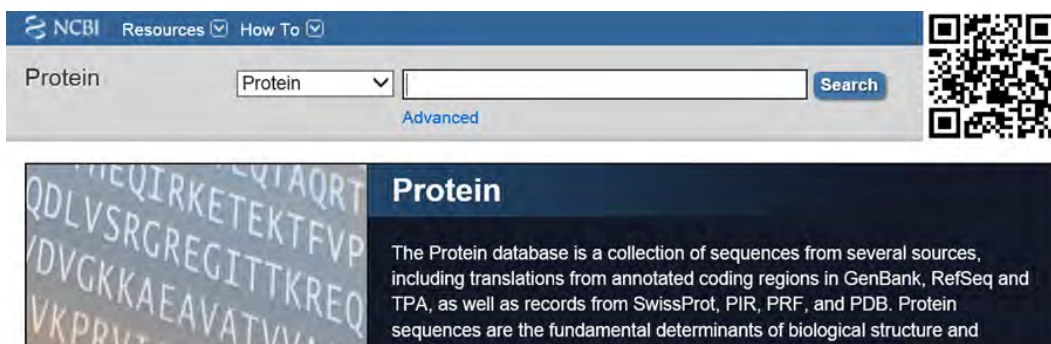
```
>HM446344.1 Influenza A virus (A/Thailand/CU-H88/2009(H1N1)) segment 6
neuraminidase (NA) gene, complete cds
ATGAATCCAAACCAAAGATAATAACCATTTGGTTCGGTCTGTATGACAATTGGAATGGCTAACTTAATAT
TACAAATTGGAAACATAATCTCAATATGGATTAGCCACTCAATTCAACTGGGAATCAAATCAGATTGA
AACATGCAATCAAAGCGTCATTACTTTATGAAAACAACACTTGGGTAAATCAGACATATGTTAACATCAGC
AACACCAACTTTGCTGCTGGACAGTCAGTGGTTTCCGTGAAATTAGCGGGCAATTCCTCTCTGCCCCTG
TTAGTGGATGGGCTATATACAGTAAAGACAACAGTATAAGAATCGGTTCCAAGGGGGATGTGTTTGTTCAT
AAGGGAACCATTCATATCATGCTCCCCCTTGGAAATGCAGAACCTTCTTCTTGACTCAAGGGGCCCTTGCTA
AATGACAAACATTTCCAATGGAACCATTAAGACAGGAGCCCATATCGAACCCATAATGAGCTGTCTTATG
GTGAAGTTCCCTCTCCATACAACCTCAAGATTTGAGTCAGTCGCTTGGTCAGCAAGTGTTCATGATGG
CATCAATTTGGCTAACAAATTTGGAAATTTCTGGCCAGACAATGGGGCAGTGGCTGTGTTAAAGTACAACGGC
ATAATAACAGACACTATCAAGAGTTGGAGAAACAATATATTTGAGAACACAAGAGTCTGAATGTGCATGTG
TAAATGGTTCTTGCTTTACTGTAATGACCGATGGACCAAGTGTGATGGACAGGCCCTCATACAAGATCTTCAG
AATAGAAAAGGGGAAAGATAGTCAAATCAGTCGAAATGAATGCCCTAATTATCACTATGAGGAATGCTCC
TGTTATCCTGATTCTAGTGAATCACATGTGTGTGCAGGGATAACTGGCATGGCTCGAATCGACCGTGGG
TGCTTTTCAACCAGAATCTGGAATATCAGATAGGATACATATGCAGTGGGATTTTCGGAGACAATCCACG
CCCTAATGATAAGACAGGCAGTTGTGGTCCAGTATCGTCTAATGGAGCAAATGGAGTAAAGGATTTTCA
TTCAAATACGGCAATGGTGTGGATAGGGAGAACTAAAAGCATTAGTTCAGAAACCGGTTTGGAGATGA
TTTGGGATCCGAACGGATGGACTGGGACAGACAATAACTTCTCAATAAAGCAAGATATCGTAGGAATAAA
TGAGTGGTCAGGATATAGCGGGAGTTTGTTCAGCATCCAGAACTAACAGGGCTGGATTGTATAAGACCT
TGCTTCTGGGTTGAACATAATCAGAGGGCGACCCAAAGAGAACACAATCTGGACTAGCGGGAGCAGCATAT
CCTTTTGTGGTGTAAACAGTGACACTGTGGGTTGGTCTTGGCCAGACGGTGTGAGTTGCCATTTACCAT
TGACAAGTAATTTGTTCAAAAACCTCCTG
```

รูปที่ 1.11 รูปแบบ summary (A) และ FASTA (B) ในการแสดงผลจากฐานข้อมูล Nucleotide

## ฐานข้อมูล Protein [https://www.ncbi.nlm.nih.gov/protein]

Protein เป็นฐานข้อมูลที่รวบรวมโปรตีนจากหลายแหล่งข้อมูล ได้แก่ GenPept, RefSeq, Swiss-Prot, PIR, PRF และ PDB การเข้าสู่ฐานข้อมูล Protein สามารถเข้าจากหน้าแรกของ NCBI website (รูปที่ 1.1) เมื่อคลิกที่คำว่า Protein (บริเวณด้านขวาภายใต้หัวข้อ Popular Resources) ก็จะเชื่อมโยงไปยังฐานข้อมูล Protein ได้ (รูปที่ 1.12) โดยสามารถพิมพ์คำสำคัญที่ต้องการค้นหาใส่ลงไปในส่วนช่องว่างด้านบน หรือเลือกที่คำว่า Advanced เพื่อใช้ระบบ Advanced search ในการค้นหาข้อมูลโปรตีนได้อย่างจำเพาะ นอกจากนี้สามารถเข้าฐานข้อมูล Protein โดยเชื่อมโยงมาจาก Related information จากฐานข้อมูล PubMed หรือ Nucleotide ก็ได้ การแสดงผลข้อมูลโปรตีนมีหลายรูปแบบ (รูปที่ 1.13 และ 14) ได้แก่

- GenPept เป็นการแสดงผลที่มีข้อมูลรายละเอียด ดังนี้ ชื่อสิ่งมีชีวิต ชื่อโปรตีน Accession no. อนุกรมวิธาน บทความอ้างอิง แหล่งที่มาของตัวอย่าง ส่วนต่างๆของโปรตีน ความยาวของกรดอะมิโน และลำดับกรดอะมิโน
- Summary เป็นการสรุปชื่อสิ่งมีชีวิต ชื่อโปรตีน ความยาวของกรดอะมิโน และ Accession no.
- FASTA เป็นการสรุปชื่อสิ่งมีชีวิต ชื่อโปรตีน Accession no. และลำดับกรดอะมิโน



รูปที่ 1.12 หน้าแรกของฐานข้อมูล Protein [https://www.ncbi.nlm.nih.gov/protein]



```

LOCUS      ADI49821                469 aa                linear   VRL 12-OCT-2011
DEFINITION  neuraminidase [Influenza A virus (A/Thailand/CU-H88/2009(H1N1))].
ACCESSION  ADI49821
VERSION    ADI49821.1
DBLINK     BioProject: PRJNA37813
DBSOURCE   accession HM446344.1
KEYWORDS   .
SOURCE     Influenza A virus (A/Thailand/CU-H88/2009(H1N1))
ORGANISM   Influenza A virus \(A/Thailand/CU-H88/2009\(H1N1\)\)
           Viruses; ssRNA viruses; ssRNA negative-strand viruses;
           Orthomyxoviridae; Influenzavirus A.
REFERENCE  1 (residues 1 to 469)
AUTHORS    Prachayangprecha,S., Makkoch,J., Payungporn,S., Chieochansin,T.,
           Vuthitanachot,C., Vuthitanachot,V., Theamboonlers,A. and
           Poovorawan,Y.
TITLE      Serological analysis of human pandemic influenza (H1N1) in Thailand
JOURNAL    J Health Popul Nutr 28 (6), 537-544 (2010)
PUBMED     21261198
FEATURES   Location/Qualifiers
           source                1..469
                                   /organism="Influenza A virus
                                   (A/Thailand/CU-H88/2009(H1N1))"
                                   /strain="A/Thailand/CU-H88/2009"
                                   /serotype="H1N1"
                                   /isolate="CU_H88_NA-1"
                                   /isolation_source="nasopharyngeal swab"
                                   /host="Homo sapiens"
                                   /db_xref="taxon:794876"
                                   /segment="6"
                                   /country="Thailand"
                                   /collection_date="Aug-2009"
                                   /note="lineage: sw1;
                                   passaged once in embryonated chicken eggs"
           Protein
           .....                1..469
                                   /product="neuraminidase"
           Region
           .....                85..465
                                   /region_name="Influenza_NA"
                                   /note="Sialidase or neuraminidase (EC 3.2.1.18) of
                                   Influenza viruses A and B; cd15483"
                                   /db_xref="CDD:271235"
           Site
           .....                order(98..99,102,104,107..108,110..113,126,136..137,139,
                                   141..144,153..155,163..166,168..170,172,174,179,196..198,
                                   201,203,205,207,211..213,215,217,408,419,448,450,452..462,
                                   465)
                                   /site_type="other"
                                   /note="tetramer interface [polypeptide binding]"
                                   /db_xref="CDD:271235"
           CDS
           .....                1..469
                                   /gene="NA"
                                   /coded_by="HM446344.1:1..1410"
ORIGIN
1 mnpnqkiiti gsvcmctigma nlilqignii siwishsiql gnqnqietcn qsvityennt
61 wvnqtyvnis ntnfaagqsv vsvklagnss lcpvsgwaiy skdnsirigs kgdvvfirep
121 fiscsplecr tffltqgall ndkhsngtik drspyrtlms cpigevpspy nsrfesvaws
181 asachdginw ltigisgpdn gavavlkyng iitdtikswr nnilrtqese cacvngscft
241 vmtdgpsdgg asykifriek gkivksvemn apnyhyeecs cypdsseitc vcrdnwhgsn
301 rpwvsfnqnl eyqigyicsg ifgdnprpnd ktgscgpvss ngangvkgfs fkyngvwig

```

รูปที่ 1.13 รูปแบบ GenPept format ในการแสดงผลโปรตีนจากฐานข้อมูล Protein

A) Summary Send to: ▼

[neuraminidase \[Influenza A virus \(A/Thailand/CU-H88/2009\(H1N1\)\)\]](#)  
 469 aa protein  
 Accession: ADI49821.1 GI: 297660473  
[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#) [Related Sequences](#)  
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

**Choose Destination**

File  Clipboard

Collections  Analysis Tool

B) FASTA Send to: ▼

**neuraminidase [Influenza A virus (A/Thailand/CU-H88/2009(H1N1))]**  
 GenBank: ADI49821.1  
[GenPept](#) [Identical Proteins](#) [Graphics](#)

```
>ADI49821.1 neuraminidase [Influenza A virus (A/Thailand/CU-H88/2009(H1N1))]
MNPNQKIITIGSVCMTIGMANLILQIGNIISIWISHSIQLGNQNIETCNQSVITYENNTWVWVQTYVNIS
NTNFAAGQSVVSVKLAGNSSLCPVSGWAIYSKDNSIRIGSKGDVFEVIREPFISCSPLECRTFFFLTQGALL
NDKHSNGTIKDRSPYRTLMSCPIGEVPSPYNSRFESVAWSASACHDGINWLTIGISGPDNGAVAVLKYNG
IITDTIKSWRNILRTQESECACVNGSCFTVMTDGPSSDGGQASYKIFRIERKIVKSVEMNAPNYHYEES
CYPDSSEITCVCRDNWHGSNRPWVSFNQNLQYQIGYICSGIFGDNPRPNDKTGSCGVPVSSNGANGVKGFS
FKYNGVWIGRTKSISSRNGFEMIWDPNGTGTDNDFSQKQDIVGINESWGSYSGSFVQHPELTGLDCIRP
CFWVELIRGRPKENTIWTSGSSISFCGVNSDTVGSWPDGAELPFTIDK
```

รูปที่ 1.14 รูปแบบการแสดงผล summary (A) และ FASTA (B) จากฐานข้อมูล Protein

นอกจากนี้ยังสามารถเลือกคำสั่ง Send to (รูปที่ 1.14 ด้านขวาบน) เพื่อถ่ายโอนข้อมูล โปรตีนหรือลำดับกรดอะมิโนจากฐานข้อมูล และบันทึกเข้าสู่คอมพิวเตอร์ของเราได้ ซึ่งเป็นวิธีที่ เหมาะสมสำหรับการบันทึกข้อมูลหลายๆ โปรตีนพร้อมๆ กันได้อย่างรวดเร็ว สำหรับการแสดงผลโปรตีนแบบ summary (รูปที่ 1.14A) จะมี link ที่เชื่อมโยงไปสู่ข้อมูลต่างๆ ที่เกี่ยวข้อง ได้แก่ Structure, Nucleotide, PubMed, Taxonomy และ Related sequences เป็นต้น

NCBI Resources How To

Structure Structure influenza AND neuraminidase AND H1N1 Search

Advanced

**Structure**

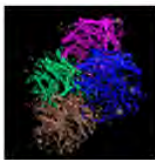
Three dimensional structures provide a wealth of information on the biological function and the evolutionary history of macromolecules. They can be used to examine sequence-structure-function relationships, interactions, active sites, and more.

Using Structure Structure Tools More Resources

รูปที่ 1.15 หน้าแรกของฐานข้อมูล Structure [https://www.ncbi.nlm.nih.gov/structure]

## ฐานข้อมูล Structure [https://www.ncbi.nlm.nih.gov/structure]

Structure เป็นฐานข้อมูลโครงสร้างสามมิติของสารชีวโมเลกุล ที่รวบรวมจาก Protein Data Bank (PDB) และประกอบด้วยเครื่องมือสำหรับการจำลองภาพสามมิติและการวิเคราะห์เปรียบเทียบโครงสร้าง การเข้าสู่ฐานข้อมูล Structure (รูปที่ 1.15) โดยสามารถพิมพ์คำสำคัญที่ต้องการค้นหาใส่ลงไปในส่วนช่องว่างด้านบน หรือเลือกที่คำว่า Advanced เพื่อใช้ระบบ Advanced search ในการค้นหาข้อมูลโครงสร้างสามมิติได้อย่างจำเพาะ นอกจากนี้สามารถเข้าสู่ฐานข้อมูล Structure โดยเชื่อมโยงมาจาก Related information จากฐานข้อมูล PubMed หรือ Protein ก็ได้ การแสดงผลข้อมูลโครงสร้างสามมิติ (รูปที่ 1.16 และ 17) มีลักษณะดังนี้



### [H1n1 2009 Pandemic Influenza Virus: Resistance Of The I223r Neuraminidase Mutant Explained By Kinetic And Structural Analysis\[Hydrolase\]](#)

Taxonomy: **Influenza A virus (A/California/07/2009(H1N1))**

Proteins: 4 Chemicals: 30 modified: 2013-07-18

MMDB ID: 103520 PDB ID: 4B7R

[View in iCn3D](#) [Similar Structures](#) [PubMed](#) [Proteins](#) [Conserved Domains](#) [PubChem Compound](#)

## รูปที่ 1.16 รูปแบบ summary ในการแสดงผลโครงสร้างสามมิติจากฐานข้อมูล Structure

การแสดงผลการค้นหาโครงสร้างสามมิติแบบ summary จะแสดงข้อมูลต่างๆ ได้แก่

- ชื่อโครงสร้างสามมิติ ซึ่งสามารถคลิกเพื่อเชื่อมโยงไปดูรายละเอียดของโครงสร้างสามมิติได้
- Taxonomy เป็นการแสดงอนุกรมวิธานของสิ่งมีชีวิต
- Proteins เป็นการระบุจำนวนโปรตีนที่เป็นองค์ประกอบอยู่ในโครงสร้าง 3 มิติดังกล่าว
- Chemicals เป็นการระบุจำนวนสารเคมีที่สามารถจับกับโครงสร้างสามมิติดังกล่าว
- Modified เป็นการระบุวัน/เดือน/ปี ที่มีการปรับปรุงข้อมูลของโครงสร้างสามมิติดังกล่าว
- MMDB ID เป็นการระบุรหัสประจำตัวในฐานข้อมูล Molecular Modeling database (MMDB)
- PDB ID เป็นการระบุรหัสประจำตัวในฐานข้อมูล Protein Data Bank (PDB)
- Link เชื่อมโยงไปสู่ข้อมูลอื่นๆ ได้แก่
  - View in iCn3D ดูรายละเอียดโครงสร้างสามมิติ โดยใช้โปรแกรม Cn3D

- Similar Structures ค้นหาสารชีวโมเลกุลที่มีโครงสร้างคล้ายคลึงกัน
- PubMed เชื่อมโยงไปสู่การอ้างอิงบทความวิจัย ในฐานข้อมูล PubMed
- Proteins เชื่อมโยงไปสู่โปรตีน ในฐานข้อมูล Protein
- Conserved Domains ค้นหาส่วน domains ของโปรตีนที่มีลักษณะอนุรักษ์
- PubChem Compound ค้นหาสารเคมีที่มีปฏิสัมพันธ์กับโครงสร้างดังกล่าว

รายละเอียดโครงสร้างสามมิติที่ได้จากฐานข้อมูล Structure (รูปที่ 1.17) ที่สำคัญ ได้แก่

- Citation เป็นข้อมูลอ้างอิงถึงบทความวิจัยที่รายงานโครงสร้างสามมิติดังกล่าว
- Biological Unit เป็นข้อมูลทางชีวภาพของโครงสร้าง ประกอบด้วย
  - Molecular graphic เป็นการแสดงโครงสร้างสามมิติ และระบุจำนวน subunits ของโปรตีน
  - Interactions ระบุชนิดของสารชีวโมเลกุลที่มีปฏิสัมพันธ์กับโครงสร้างสามมิติ
  - Download Structure Data สามารถดาวน์โหลดข้อมูลโครงสร้างสามมิติได้
- Molecular Components เป็นข้อมูลทางอณูชีววิทยาของโครงสร้าง ประกอบด้วย
  - Proteins ระบุชนิดและจำนวนของโปรตีนที่เป็นส่วนประกอบในโครงสร้าง
  - Chemicals ระบุชนิดและจำนวนของสารเคมีที่มีปฏิสัมพันธ์กับโครงสร้างสามมิติดังกล่าว

ตัวอย่างเช่น จากรูปที่ 1.17 เป็นการค้นหาข้อมูลโครงสร้างสามมิติของเอนไซม์ neuraminidase ของเชื้อไวรัสไข้หวัดใหญ่สายพันธุ์ H1N1 (H1N1 influenza A virus) พบว่ามีข้อมูลที่สนใจ คือ เอนไซม์ดังกล่าวมีลักษณะเป็น tetrameric ประกอบด้วยโปรตีน 4 หน่วยย่อย โดยมีปฏิสัมพันธ์กับสารเคมี อื่นๆ ได้แก่

- calcium ion 8 โมเลกุล
- 5-N-Acetyl-3-(1-Ethylpropyl)-1-Cyclohexene-1-Carboxylic Acid 4 โมเลกุล
- 4-(2-Hydroxyethyl)-1-Piperazine Ethanesulfonic Acid 4 โมเลกุล
- N-Acetyl-D-Glucosamine 14 โมเลกุล

Biological Unit for 4B7R: tetrameric; determined by software (PISA) [?](#)

Molecular Graphic [?](#)

Interactions [?](#)

Mouse over symbols to identify molecules  
and click to highlight in 3D viewer

Protein  
  Nucleotide  
  Chemical

Download Structure Data [?](#)

Download

Format: ASN.1 (Cn3D) ▾

Data Set: Single 3D structure ▾

Molecular Components in 4B7R [?](#)

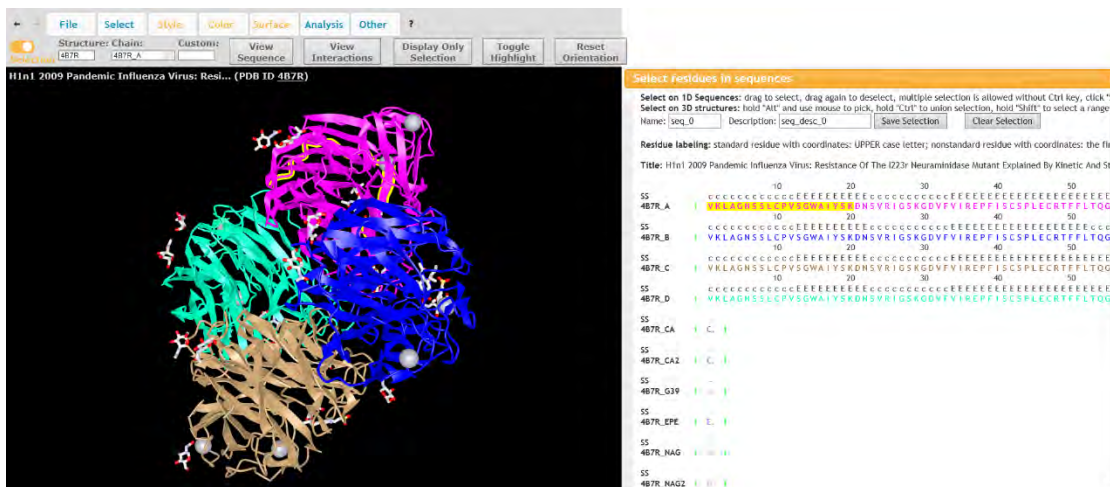
Label	Count	Molecule
<b>Proteins (4 molecules)</b>		
	4	<div style="display: flex; align-items: center;"> <div style="margin-right: 10px;"> <p style="font-size: x-small; margin: 0;">4 Proteins</p> <p style="font-size: x-small; margin: 0;">30 Domains</p> <p style="font-size: x-small; margin: 0;">Domain Families</p> <p style="font-size: x-small; margin: 0;">Specific Hits</p> <p style="font-size: x-small; margin: 0;">Super Families</p> </div> <div style="flex-grow: 1;"> <p style="text-align: center; font-weight: bold; font-size: small;">Neuraminidase</p> <p style="text-align: center; font-size: x-small; margin-top: 5px;">Influenza_NA</p> <p style="text-align: center; font-size: x-small; margin-top: 5px;">Sialidase superfamily</p> </div> <div style="margin-left: 10px; text-align: right;"> <p style="font-size: x-small; margin: 0;">See in 3D</p> <p style="font-size: x-small; margin: 0;">See in 3D</p> </div> </div>
<b>Chemicals (30 molecules)</b>		
	8	Calcium Ion
	4	5-N-Acetyl-3-(1-Ethylpropyl)-1-Cyclohexene-1-Carboxylic Acid
	4	4-(2-Hydroxyethyl)-1-Piperazine Ethanesulfonic Acid
	14	N-Acetyl-D-Glucosamine

รูปที่ 1.17 รายละเอียดโครงสร้างสามมิติที่ได้จากฐานข้อมูล Structure

## การวิเคราะห์โครงสร้างสามมิติด้วยโปรแกรม Cn3D

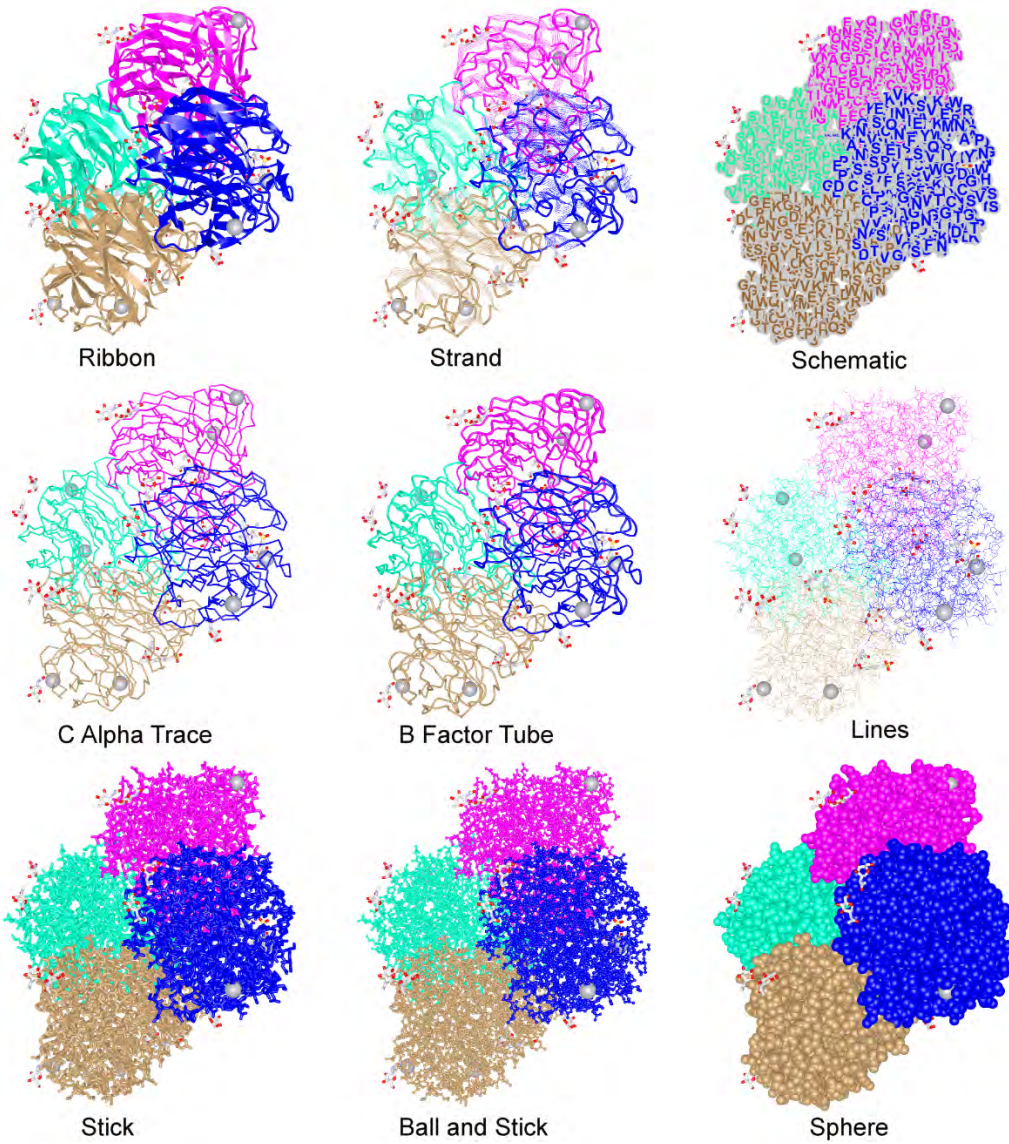
จากผลโครงสร้างสามมิติจากฐานข้อมูล Structure (รูปที่ 1.16) หากคลิกที่ View in iCn3D จะเชื่อมโยงไปพิจารณาโครงสร้างสามมิติด้วยโปรแกรม Cn3D ซึ่งสามารถปรับเปลี่ยนการดูรายละเอียดของโครงสร้าง ได้แก่ (รูปที่ 1.18 และ 19)

- File สามารถเปิดดูข้อมูล หรือบันทึกข้อมูลโครงสร้างสามมิติ
- Select เลือกแสดง หรือเน้นเฉพาะส่วนที่สนใจโครงสร้างสามมิติ
- Style ปรับรูปแบบของ protein, side chains, nucleotide, ligand, ions และ water
- Color เปลี่ยนสีของโมเลกุล หรืออะตอมต่างๆ ในโครงสร้างสามมิติ
- Surface เปลี่ยนรูปแบบพื้นผิวของโมเลกุลในโครงสร้างสามมิติ
- Analysis วิเคราะห์โครงสร้าง เช่น ระยะห่าง ปฏิสัมพันธ์ และพันธะทางเคมีต่างๆ
- Other ปรับมุมมอง การย่อขนาด การขยายขนาด และการหมุนโครงสร้าง เป็นต้น



รูปที่ 1.18 การแสดงผลโครงสร้างสามมิติด้วยโปรแกรม Cn3D

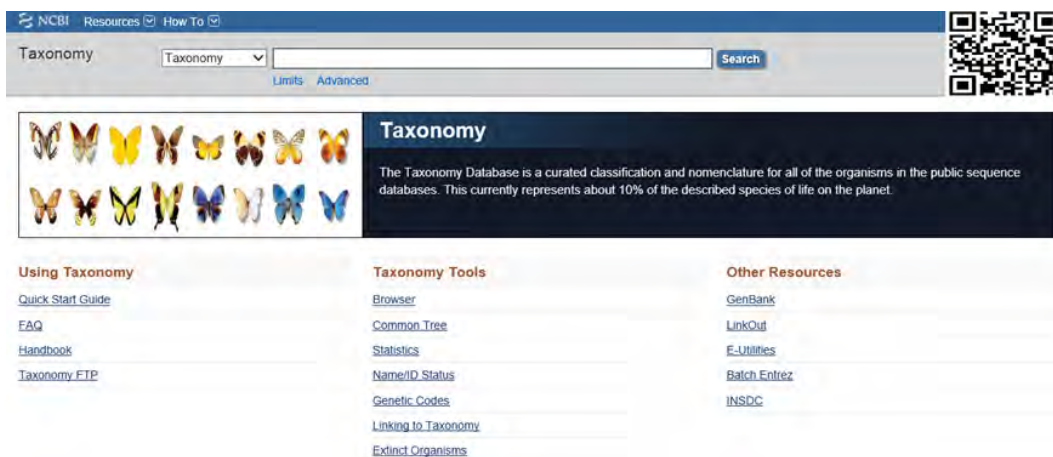




รูปที่ 1.19 การปรับเปลี่ยนรูปแบบต่างๆ (style) ของโครงสร้างสามมิติ ด้วยโปรแกรม Cn3D

## ฐานข้อมูล Taxonomy [https://www.ncbi.nlm.nih.gov/taxonomy]

Taxonomy เป็นฐานข้อมูลที่แสดงชื่อและอนุกรมวิธานทางพันธุกรรมของสิ่งมีชีวิต การเข้าสู่ฐานข้อมูล Taxonomy สามารถเข้าจาก <https://www.ncbi.nlm.nih.gov/taxonomy> (รูปที่ 1.20) โดยสามารถพิมพ์คำสำคัญที่ต้องการค้นหาใส่ลงไปในส่วนช่องว่างด้านบน หรือเลือกที่คำว่า Advanced เพื่อใช้ระบบ Advanced search ในการค้นหาข้อมูลอนุกรมวิธานของสิ่งมีชีวิตได้อย่างจำเพาะ นอกจากนี้สามารถเข้าฐานข้อมูล Taxonomy โดยเชื่อมโยงมาจาก Related information จากฐานข้อมูล PubMed หรือ Nucleotide, Protein หรือ Structure ก็ได้ การแสดงผลข้อมูลอนุกรมวิธาน (รูปที่ 1.21)



รูปที่ 1.20 หน้าแรกของฐานข้อมูล Taxonomy

[https://www.ncbi.nlm.nih.gov/taxonomy]

ผลการค้นหาอนุกรมวิธานของสิ่งมีชีวิต จะแสดงข้อมูลต่างๆ ได้แก่

- Taxonomy ID เป็นการระบุรหัสประจำตัวในฐานข้อมูล Taxonomy
- Scientific name ชื่อทางวิทยาศาสตร์ของสิ่งมีชีวิต
- Lineage เป็นการระบุอนุกรมวิธานของสิ่งมีชีวิต โดยแบ่งเป็น Kingdom, Family, Genus, Species และ Subtypes เป็นต้น



**Influenza A virus (A/Thailand/CU-H88/2009(H1N1))**

*Taxonomy ID:* 794876 (for references in articles please use NCBI:txid794876)

*Scientific name:* **Influenza A virus (A/Thailand/CU-H88/2009(H1N1))**

*Inherited blast name:* **viruses**

*Rank:* no rank

*Genetic code:* [Translation table 1 \(Standard\)](#)

*Host:* vertebrates| human

[Lineage\( full \)](#)  
 Viruses; ssRNA viruses; ssRNA negative-strand viruses; [Orthomyxoviridae](#);  
 Alphainfluenzavirus; [Influenza A virus](#); [H1N1 subtype](#)

[ICTV homepage](#)

**External Information Resources (NCBI LinkOut)**

LinkOut	Subject	LinkOut Provider
<a href="#">Influenza Virus Resource</a>	taxonomy/phylogenetic	<a href="#">NCBI taxonomy bookmarks</a>

**Entrez records**

Database name	Direct links
Nucleotide	<a href="#">8</a>
Protein	<a href="#">10</a>
Taxonomy	<a href="#">1</a>

รูปที่ 1.21 ผลการค้นหอนุกรมวิธานของสิ่งมีชีวิตจากฐานข้อมูล Taxonomy

ฐานข้อมูล OMIM [<https://www.ncbi.nlm.nih.gov/omim>]

OMIM (Online Mendelian Inheritance in Man) เป็นฐานข้อมูลของยีนและความผิดปกติทางพันธุกรรมของมนุษย์ที่เกี่ยวข้องในการเกิดโรค การเข้าสู่ฐานข้อมูล OMIM สามารถเข้าจาก <https://www.ncbi.nlm.nih.gov/omim> (รูปที่ 1.22) โดยสามารถพิมพ์คำสำคัญที่ต้องการค้นหาใส่ลงไปในส่วนช่องว่างด้านบน หรือเลือกที่คำว่า Advanced เพื่อใช้ระบบ Advanced search ในการค้นหาข้อมูลได้อย่างจำเพาะ การแสดงผลข้อมูลที่ได้จาก OMIM มีลักษณะ (รูปที่ 1.23) ดังนี้

- Title ชื่อยีนทั่วไป
- Alternative titles or symbols ชื่อยีน หรือสัญลักษณ์ชื่อย่อของยีนอื่นๆ ที่ใช้
- HGNC Approved Gene Symbol เป็นชื่อยีนมาตรฐานที่ใช้
- Cytogenetic location เป็นการระบุตำแหน่งของยีนบนโครโมโซม
- Genomic coordinates เป็นการระบุตำแหน่งของยีนใน genome sequences

- Cloning and Expression เป็นการอ้างอิงการทดลองศึกษาการแสดงออกของยีน
- Gene Function เป็นการระบุหน้าที่ของยีนโดยอ้างอิงจากการศึกษาวิจัย
- References เป็นการอ้างอิงถึงบทความวิจัยที่เกี่ยวข้อง



รูปที่ 1.22 หน้าแรกของฐานข้อมูล OMIM [<https://www.ncbi.nlm.nih.gov/omim>]

\* 609209

## INFLUENZA VIRUS NS1A PROTEIN-BINDING PROTEIN; IVNS1ABP

*Alternative titles; symbols*

NS1-BINDING PROTEIN; NS1BP  
NCX DOWNSTREAM GENE 1; ND1  
KIAA0850

*HGNC Approved Gene Symbol: **IVNS1ABP***

*Cytogenetic location: **1q25.3** Genomic coordinates (GRCh38): **1:185,292,366-185,317,328** (from NCBI)*

### TEXT

#### ► Cloning and Expression

#### ► Gene Function

#### ▼ Mapping

By radiation hybrid analysis, [Nagase et al. \(1998\)](#) mapped the IVNS1ABP gene to chromosome 1. [+](#)

#### ► REFERENCES

รูปที่ 1.23 ผลการค้นหายีนและความผิดปกติทางพันธุกรรมจากฐานข้อมูล OMIM

## เอกสารอ้างอิง และแนะนำอ่านเพิ่มเติม

1. How to: Find published information on a gene or sequence  
[<https://www.ncbi.nlm.nih.gov/guide/howto/find-published-info-gene-sequence>]
2. How to: Find the function of a gene or gene product  
[<https://www.ncbi.nlm.nih.gov/guide/howto/find-func-gene>]
3. How to: Find transcript sequences for a gene  
[<https://www.ncbi.nlm.nih.gov/guide/howto/find-transcript-gene>]
4. McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2002.  
[[https://www.ncbi.nlm.nih.gov/books/NBK21101/pdf/Bookshelf\\_NBK21101.pdf](https://www.ncbi.nlm.nih.gov/books/NBK21101/pdf/Bookshelf_NBK21101.pdf)]



# 2

ฐานข้อมูลทางด้านไวรัสวิทยา

Virological Databases

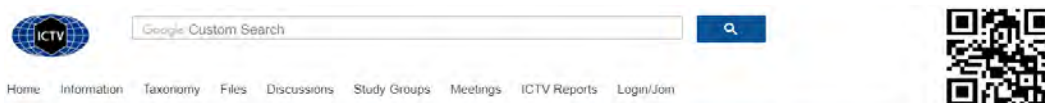
## ฐานข้อมูลทางด้านไวรัสวิทยา

### Virological Databases

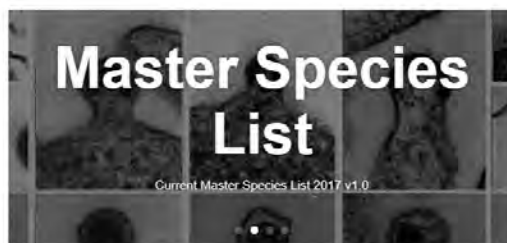
นอกจากฐานข้อมูลทางด้านชีวภาพแล้ว ปัจจุบันยังมีฐานข้อมูลเฉพาะทางด้านไวรัสวิทยาเพื่อใช้ในการเก็บรวบรวมข้อมูลเชิงลึกต่างๆ ที่เกี่ยวกับเชื้อไวรัส และโรคที่เกิดจากการติดเชื้อไวรัส ในที่นี้จึงขอกล่าวถึงฐานข้อมูลทางด้านไวรัสวิทยาที่สำคัญและนิยมใช้อ้างอิงในงานวิจัยทางด้านไวรัสวิทยา ดังนี้

#### ฐานข้อมูล ICTV [<https://talk.ictvonline.org>]

เป็นฐานข้อมูลของ International Committee on Taxonomy of Viruses (ICTV) ที่นิยมใช้เป็นแหล่งอ้างอิงมาตรฐานสากลในการจัดจำแนกอนุกรมวิธาน (taxonomy) ของเชื้อไวรัส (รูปที่ 2.1) โดยฐานข้อมูลจะมีการ update ข้อมูลให้ทันสมัยอยู่เสมอ นอกจากนี้ผู้ใช้สามารถดาวน์โหลด ICTV Master Species Lists ซึ่งเป็นไฟล์ที่รวบรวมข้อมูลอนุกรมวิธานของเชื้อไวรัสทั้งหมดในฐานข้อมูลมาใช้ได้ ปัจจุบัน ICTV (ปี 2017) มีการเก็บข้อมูลอนุกรมวิธานของเชื้อไวรัส โดยแบ่งออกเป็น 9 orders, 131 families, 46 subfamilies, 803 genera และ 4853 species



### International Committee on Taxonomy of Viruses ICTV



รูปที่ 2.1 ฐานข้อมูล ICTV [<https://talk.ictvonline.org>]

## ฐานข้อมูล ViralZone [https://viralzone.expasy.org]

เป็นฐานข้อมูลที่รวบรวมเนื้อหาเกี่ยวกับอนุชีววิทยาของเชื้อไวรัส ได้แก่ อนุกรมวิธาน (taxonomy), อนุภาคไวรัส (virions), การเพิ่มจำนวน (replication cycles), ตัวรับบนผิวเซลล์ (cell receptors), ปฏิสัมพันธ์ระหว่างไวรัสและเซลล์เจ้าบ้าน (host-virus interactions) และ วิวัฒนาการของจีโนม (virus genome evolution) โดยแสดงผลจำแนกตามลักษณะจีโนมของเชื้อไวรัส (Baltimore classification system) ดังนี้ (รูปที่ 2.2)

- Group I: double strand DNA viruses
- Group II: single strand DNA viruses
- Group III: double strand RNA viruses
- Group IV: positive single strand RNA viruses
- Group V: negative single strand RNA viruses
- Group VI: RNA reverse transcribing viruses (retroviruses)
- Group VII: DNA reverse transcribing viruses

The screenshot shows the ViralZone website interface. At the top, there is a blue header with the ViralZone logo and navigation links for 'Contact Us' and 'Home'. Below the header, there are three main categories: 'DNA viruses', 'Retro-transcribing viruses', and 'RNA viruses'. Each category has sub-menus for different genome types: 'ds DNA', 'ss DNA', 'ds DNA (RT)', 'ss RNA (RT)', 'dsRNA', 'ss RNA (+)', 'ss RNA (-)', and 'Circular ss RNA'. On the left side, there are navigation options: 'Browse by Baltimore', 'Browse by host', 'Browse by virion', 'Resources' (including HIV, HBV, Herpesvirus), 'Human viruses' (including Human Virus Size), 'Reference sequences', and 'Viral molecular biology' (including Virion, Virus entry, Transcription, replication, translation, and Virus exit). The main content area features 'VIRALZONE NEWS' with two articles: 'Virus Acronyms February 2018' and 'Herpesvirus resource September 2017'. On the right side, there is a 'ViralZone current statistics' section for May 2018, showing 622 Virus description pages, 121 Families, 492 Genera, and 9 individual Species. It also mentions 216 Viral molecular biology pages and provides a link to UniProt release 2018\_04 with 6,279 reference proteomes and 16,792 manually reviewed proteins.

รูปที่ 2.2 ฐานข้อมูล ViralZone [https://viralzone.expasy.org]

### ฐานข้อมูล ViPR [<http://www.viprbrc.org>]

เป็นฐานข้อมูลสำหรับการค้นหาข้อมูล และการวิเคราะห์ข้อมูลทางอณูชีววิทยาของเชื้อไวรัส ซึ่งประกอบด้วย 3 ส่วนหลัก (รูปที่ 2.3) ได้แก่

- ViPR Search Tools ใช้สำหรับการค้นหาข้อมูลต่างๆ ที่เกี่ยวข้องกับเชื้อไวรัส เช่น genome, protein, 3D structure, strain, immune epitopes, antiviral drugs และ host factors เป็นต้น
- Analysis Tools ใช้สำหรับวิเคราะห์ข้อมูลทางอณูชีววิทยาของเชื้อไวรัส เช่น identify similar sequences (BLAST), identify short peptides in proteins, metadata-driven comparative analysis, sequence format conversion, PCR primer design, genome annotator, align sequences, analyze sequence variations, visualized aligned sequences และ generate phylogenetic tree เป็นต้น
- Workbench เป็นส่วนที่ใช้ในการเก็บบันทึกข้อมูลที่ได้จากการค้นหา หรือการวิเคราะห์ดังกล่าวข้างต้น และสามารถ uploads เพื่อเผยแพร่ หรือส่งต่อให้เพื่อนร่วมงานได้

### ฐานข้อมูล VIPERdb [<http://viperdbscripps.edu/index.php>]

Virus Particle Explorer database (VIPERdb) เป็นฐานข้อมูลที่รวบรวมโครงสร้าง icosahedral capsid ของเชื้อไวรัส ซึ่งมีข้อมูลหลายรูปแบบ (รูปที่ 2.4) ได้แก่ X-ray structures, Cryo EM, helical structure, non-icosahedral structure, modeled structures, crystal information, nucleic acid organization และ hetero atoms นอกจากนี้ยังมี utilities เพื่อใช้ในการวิเคราะห์ ได้แก่ family association energies, oligomer generator, icosahedral server, map a residue, virosphere, gallery maker, contact finder, lattice matrices, PDB to VIPER, amino acid information, secondary structure information และ CapsidMaps เป็นต้น



**ViPR**  
Virus Pathogen Resource

About Us | Community | Announcements | Links | Resources | Support

**Search**    **Analyze**    **Save to Workbench**

**Search our comprehensive database for:**

- Sequences & strains
- Immune epitopes
- 3D protein structures
- Host Factor Data
- Antiviral Drugs

**Analyze data online:**

- Sequence Alignment
- Phylogenetic Tree
- Sequence Variation (SNP)
- Metadata-driven Comparative Analysis
- BLAST

**Sign up for a workbench to:**

- Store and share data
- Combine working sets
- Integrate your data with ViPR data
- Store and share analyses
- Custom search alert

Browse All Search Types    Browse All Tools    Sign In

**Supported Programs**

Click on a supported program of interest to go to program description page.

NIH NIAID Functional Genomics    NIH NIAID Systems Biology

**Virus Families**

Click on icon of family or species of interest. Click [here](#) to view all families and species in list format. Don't know family of species?

รูปที่ 2.3 ฐานข้อมูล ViPR [http://www.viprbrc.org]

**VIPERdb**    Main    Data    Utilities    Links    Help    Contact Us    Find a Virus

**Virus Particle Explorer<sup>2</sup>**    Family list    Genus list    Enter PDBID

**WELCOME**

VIPERdb is a database for [icosahedral virus capsid structures](#). The emphasis of the resource is on providing data from structural and computational analyses on these systems, as well as high quality renderings for visual exploration. In addition, all virus capsids are placed in a [single icosahedral orientation convention](#), facilitating comparison between different structures. The web site includes powerful [search utilities](#), links to other relevant databases, background information on virus capsid structure, and useful [database interface tools](#).

You have different options on how to start using VIPERdb. You can jump to any section in the site using the [Top Menu](#). If you are looking for a specific entry, you can look for it on the [X-Ray](#) or [Cryo-EM](#) drop-down lists on the top, or type its [PDB-ID](#) in the corresponding field above. If you are not sure, you can access the [Search Page here](#) and look for it by Name. All virus entries are grouped into [Families](#), you can see a list [here](#) with links to all their members.

All database information (from [668 viruses](#)) is accessed through the [Info Page](#), which shows all pertinent data using different database interfaces ([Biodata](#), [3D Structure](#), [2D  \$\Phi\$ - \$\Psi\$  Maps](#)) for each individually selected entry.

If you are new to VIPERdb, you can get familiar with the site by accessing the [Info Page](#) for some of the most popular spherical viruses. Here is just a small sample (click images to open corresponding [Info Page](#)):

รูปที่ 2.4 ฐานข้อมูล VIPERdb [http://viprdb.scripps.edu/index.php]

## ฐานข้อมูล IEDB [http://www.iedb.org]

Immune Epitope DataBase (IEDB) เป็นฐานข้อมูลที่รวบรวมเนื้อหาเกี่ยวกับ immune epitope ที่กระตุ้นระบบภูมิคุ้มกัน โดยแบ่งออกเป็น 2 ส่วน (รูปที่ 2.5) ได้แก่

- Search สำหรับใช้ในการค้นหา immune epitope โดยสามารถเลือกตั้งค่าในการค้นหา ได้แก่ epitope, antigen, host, assay, MHC restriction และ disease เป็นต้น
- Epitope Analysis Resource ใช้ในการวิเคราะห์และทำนายต่างๆ เช่น
  - T cell epitope prediction
  - B cell epitope prediction
  - Epitope analysis tools

**Summary Metrics**

Peptidic Epitopes	448,585
Non-Peptidic Epitopes	2,636
T Cell Assays	337,318
B Cell Assays	413,385
MHC Ligand Assays	930,129
Epitope Source Organisms	3,661
Restricting MHC Alleles	764
References	19,374

รูปที่ 2.5 ฐานข้อมูล IEDB [http://www.iedb.org]

## ฐานข้อมูล VirHostnet [http://virhostnet.prabi.fr]

เป็นฐานข้อมูลที่รวบรวมเนื้อหาเกี่ยวกับปฏิสัมพันธ์ระหว่างโปรตีนของเชื้อไวรัส และโปรตีนของเซลล์เจ้าบ้าน โดยแบ่งออกเป็น 3 ส่วน (รูปที่ 2.6) ได้แก่

- Search ใช้สำหรับการค้นหาข้อมูลต่างๆ ได้แก่ protein, domain, pathway taxonomy และ publication เป็นต้น
- Browse ใช้สำหรับให้แสดงข้อมูลตามรูปแบบต่างๆ เช่น Baltimore group, Family rank และ Species rank เป็นต้น
- Interology web service สำหรับใส่ข้อมูลลำดับกรดอะมิโนที่สนใจลงไป จากนั้นโปรแกรมจะเปรียบเทียบกับฐานข้อมูลเพื่อระบุปฏิสัมพันธ์ของโปรตีน



Baltimore	VH network			VV network	
	ppi	v	h	ppi	v
Deltavirus	154	2	154	0	0
Retro-transcribing viruses	2848	151	1512	125	93
Satellites	0	0	0	3	3
dsDNA viruses, no RNA stage	12831	751	3781	4348	1515
dsRNA viruses	68	13	62	19	17
ssDNA viruses	293	6	286	31	16
ssRNA viruses	15063	414	3945	1019	395
unassigned viruses	0	0	0	15	5
unclassified viruses	0	0	0	2	1

รูปที่ 2.6 ฐานข้อมูล VirHostnet [http://virhostnet.prabi.fr]

## ฐานข้อมูล NCBI-VVR [<https://www.ncbi.nlm.nih.gov/genome/viruses/variation>]

VVR (Virus Variation Resource) เป็นฐานข้อมูลหนึ่งของ NCBI ที่รวบรวมลำดับนิวคลีโอไทด์ และกรดอะมิโนของเชื้อไวรัสที่มีความหลากหลายทางพันธุกรรมสูง ซึ่งมักจะเป็นเชื้อไวรัสชนิดที่สามารถก่อโรคติดเชื้ออุบัติใหม่ (รูปที่ 2.7) ได้แก่

- Influenza virus (รูปที่ 2.8)
- Dengue virus (รูปที่ 2.9)
- Zika virus (รูปที่ 2.10)
- West Nile virus (รูปที่ 2.11)
- Rotavirus (รูปที่ 2.12)
- MERS coronavirus (รูปที่ 2.13)
- Ebolavirus (รูปที่ 2.14)

โดยสามารถตั้งค่าในการค้นหาข้อมูลได้มากมาย เช่น Type/Species, Host, Region/Country, Protein/Gene, Genome region, Subtype, Sequence length, Collection date และ Release date เป็นต้น



รูปที่ 2.7 NCBI-VVR [<https://www.ncbi.nlm.nih.gov/genome/viruses/variation>]



**Influenza Virus Resource**  
Information, Search and Analysis

Influenza Virus Database

Related links New features How to cite Contact us Help

Overview Database Genome Set BLAST Annotation Submission FTP Virus resources

Protein or nucleotide sequences can be retrieved from the database using GenBank accession numbers or search terms. Multiple queries can be built by clicking the "Add Query" button every time a new query is made, and queries in any combination from the Query Builder can be selected to get sequences in the database. Sequences can be downloaded, and it is possible to analyze them using the multiple sequence alignment or tree building tool integrated to the database.

**Get sequences by accession**  
Enter a comma or space separated list of sequence accessions or upload text file with this list.  
Upload เลือกไฟล์ ไม่เลือกไฟล์ Accessions

Add query Show results

**Select sequence type:**  
 Protein  Protein coding region  Nucleotide

**Search for keyword:**  
Keyword  Search in: strain name

**Define search set:**

Type	Host	Country/Region	Protein	Subtype	H any	N any	Sequence length	Collection date	Release date
any	any	any	any	any	1	1	Min: <input type="text"/> From: <input type="text"/>	<input type="text"/>	<input type="text"/>
A	Avian	regions	PB2		2	2	Max: <input type="text"/> To: <input type="text"/>	<input type="text"/>	<input type="text"/>
B	Bat	Northern temperate	PB1					<input type="text"/>	<input type="text"/>
C	Blow fly	Southern temperate	PB1-F2		3	3		<input type="text"/>	<input type="text"/>

Additional filters:  Collapse identical sequences

Add query Show results Clear form

รูปที่ 2.8 ฐานข้อมูล Influenza Virus Resource ที่เชื่อมโยงมาจาก NCBI-WVR

**Virus Variation** Dengue virus database

Virus Variation home Virus resources

**Select sequence type**  
 Protein  Nucleotide  Full-length sequences only

**Define search set**

Structural Non-structural

	C	M	E	NS1	NS2A	NS2B	NS3
Type	any	any	any	any	any	any	any
Disease	1 known	2 DF	3 DHF	4 DSS			
Host	Human	Mammal	Mosquito	Primate			
Region/Country	regions	Africa	Asia	Europe			
Genome region	C	M	E	NS1			
Isolation source	abdominal cavity	blood	brain	cerebrospinal fluid			

Collection date:    to

Release date:    to

Additional filters:  Collapse identical sequences

Add query Show results

รูปที่ 2.9 ฐานข้อมูล Dengue virus database ที่เชื่อมโยงมาจาก NCBI-WVR

**Virus Variation** Zika virus

Virus Variation home Virus resources

**Select sequence type**

Protein  Nucleotide  Full-length sequences only

**Define search set**

Structural Non-structural

C	prM	E	NS1	NS2A	NS2B	NS3
---	-----	---	-----	------	------	-----

**Host** any Human Mammal Mosquito Primate

**Region/Country** any regions Africa Asia Europe

**Genome region** any C prM E NS1

**Isolation source** any abdominal cavity amniotic fluid blood brain

Collection date: [ ] [ ] [ ] to [ ] [ ] [ ]

Release date: [ ] [ ] [ ] to [ ] [ ] [ ]

Year Month Day Year Month Day

**Additional filters** [ ] [ ]

Add query Show results  Collapse identical sequences

รูปที่ 2.10 ฐานข้อมูล Zika virus database ที่เชื่อมโยงมาจาก NCBI-VR

**Virus Variation** West Nile virus database

Virus Variation home Virus resources

**Select sequence type**

Protein  Nucleotide  Full-length sequences only

**Define search set**

Structural Non-structural

C	prM	E	NS1	NS2A	NS2B	NS3
---	-----	---	-----	------	------	-----

**Host** any Amphibian Bird Human Mammal

**Region/Country** any regions Africa Asia Europe

**Genome region** any C prM E NS1

**Isolation source** any abdominal cavity blood brain cerebrospinal fluid

Collection date: [ ] [ ] [ ] to [ ] [ ] [ ]

Release date: [ ] [ ] [ ] to [ ] [ ] [ ]

Year Month Day Year Month Day

**Additional filters** [ ] [ ]

Add query Show results  Collapse identical sequences

รูปที่ 2.11 ฐานข้อมูล West Nile virus database ที่เชื่อมโยงมาจาก NCBI-VR

**Virus Variation** Rotavirus database

Virus Variation home Virus resources

**Select sequence type**

Protein  Nucleotide  Full-length sequences only

**Define search set**

Species	Host	Region/Country	Protein	Isolation source
any Rotavirus A	any Bird Human Mammal Primate	any <b>regions</b> Africa Asia Europe	any VP1 VP2 VP3 VP4	any abdominal cavity blood cecum cerebrospinal fluid

Collection date: [ ] [ ] [ ] to [ ] [ ] [ ]  
 Release date: [ ] [ ] [ ] to [ ] [ ] [ ]  
Year Month Day Year Month Day

Additional filters

Add query Show results  Collapse identical sequences

รูปที่ 2.12 ฐานข้อมูล Rotavirus database ที่เชื่อมโยงมาจาก NCBI-WR

**Virus Variation** MERS coronavirus database

Virus Variation home Virus resources

**Select sequence type**

Protein  Nucleotide  Full-length sequences only

**Define search set**

Host	Region/Country	Genome region	Isolation source
any Unknown Camel Gorilla Human	any <b>regions</b> Africa Asia Europe	any ORF 1a protein ORF1ab protein Spike glycoprotein NS3 protein	any blood lung oronasopharynx stool

Collection date: [ ] [ ] [ ] to [ ] [ ] [ ]  
 Release date: [ ] [ ] [ ] to [ ] [ ] [ ]  
Year Month Day Year Month Day

Additional filters

Add query Show results  Collapse identical sequences

รูปที่ 2.13 ฐานข้อมูล MERS coronavirus database ที่เชื่อมโยงมาจาก NCBI-WR

รูปที่ 2.14 ฐานข้อมูล Ebolavirus database ที่เชื่อมโยงมาจาก NCBI-VVR

นอกจากนี้ยังมีฐานข้อมูลเฉพาะของเชื้อไวรัสแต่ละชนิดที่สามารถก่อโรครุนแรงในมนุษย์ ได้แก่ Hepatitis B virus, Hepatitis C virus, Human immunodeficiency virus (HIV), Influenza virus และ human papillomavirus (HPV) เป็นต้น ในที่นี้จึงขอแนะนำฐานข้อมูลที่สำคัญสำหรับเชื้อไวรัสแต่ละชนิด ดังนี้

#### ฐานข้อมูลสำหรับเชื้อไวรัสตับอักเสบบี (Databases for hepatitis B virus)

- SeqHepB [<http://www.seqhepb.com>]

เป็นฐานข้อมูลเกี่ยวกับจีโนม และการวิเคราะห์การกลายพันธุ์ที่เกี่ยวข้องกับการดื้อยาต้านไวรัสของเชื้อไวรัสตับอักเสบบี (รูปที่ 2.15)


- HBVdb [<https://hbvdb.ibcp.fr/HBVdb/>]

เป็นฐานข้อมูลเกี่ยวกับเชื้อไวรัสตับอักเสบบีโดยเฉพาะ ผู้ใช้สามารถค้นหาข้อมูลจีโนม และความหลากหลายทางพันธุกรรม พร้อมทั้งสามารถวิเคราะห์ลำดับนิวคลีโอไทด์ จำแนกสายพันธุ์ และวิเคราะห์การกลายพันธุ์ที่เกี่ยวข้องกับการดื้อยาต้านไวรัสของเชื้อไวรัสตับอักเสบบี (รูปที่ 2.16)



# SeqHepB

Hepatitis B Viral Genome Analysis



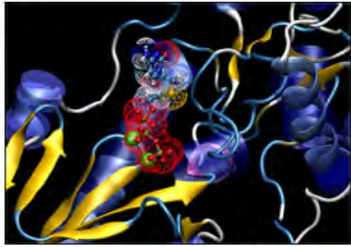
Home
Log-In
About Us
Services
Drug Resistance
Contact Us

### Drug Resistance

As treatment options for the management of Hepatitis B continue to grow, the clinician's challenge to manage drug selection in the face of single and multi-drug resistance is increasing daily.

Based on a dynamic global mutational database, SeqHepB has been designed to enable clinicians to manage their patients' treatment regime on an informed and accurate basis.

Genomic sequence analysis software within SeqHepB rapidly and accurately identifies those mutations and sub-mutations that lead to drug resistance and multi-drug resistance.



Lamivudine bound in HBV polymerase

รูปที่ 2.15 ฐานข้อมูล SeqHepB [<http://www.seqhepb.com>]

## HBVdb

The Hepatitis B Virus database



HOME
HBV
QUERY
ANALYSIS
HBVDB
LINKS

### Welcome to the HBVdb home page !

The specialized Hepatitis B Virus (HBV) database, HBVdb, allows the researchers to investigate the genetic variability of the virus and the viral resistance to treatment. HBV is a major health problem worldwide with more than 350 million individuals being chronically infected. The main drugs used to treat infected patients are nucleos(t)ides analogs (reverse transcriptase inhibitors). Unfortunately, HBV mutants resistant to these drugs may be selected and be responsible for treatment failure. HBVdb contains a collection of computer-annotated sequences based on manually annotated reference genomes. The database web interface allows static and dynamic queries and provides integrated analysis tools including sequence annotation, genotyping and drug resistance profiling.

The current HBVdb release is 43.0 with 88219 entries and was last updated on 2018-05-10 ([see news](#)).

The HBVdb service has been accessed 285465 times by 17940 unique users since 2012 ([see statistics](#)).

If you use HBVdb in your work, please cite *Hayer J. et al., Nucleic Acids Research, 2013* ([see about](#)).

Please use the menu bar above or the submit buttons below in order to access the Hepatitis B Virus Database (HBVdb) resources. A brief description of each resource will appear when the pointing device pauses over a menu or a button. You can find more explanations about how to use these resources in the help page from the HBVdb menu ([see help](#)).

Genome

HBV
Nomenclature

Proteins
Core
HBx
Surface
Polymerase

Query
Dataset
Nucleotide
Protein

Generic N
Blast N
ClustalW N
FASTA N

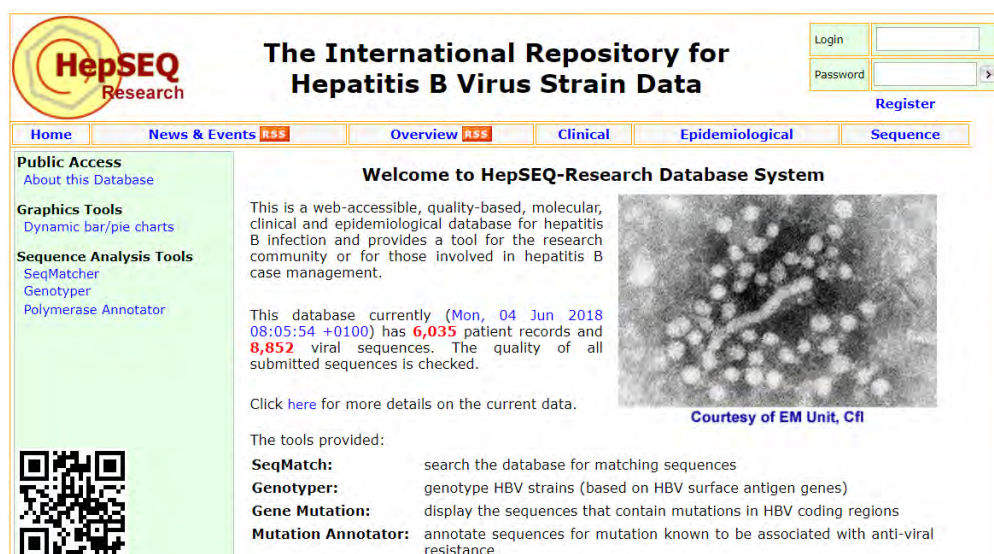
Analysis
Generic P
Blast P
ClustalW P
FASTA P

Specialized
Annotate
Genotype
Resistance

รูปที่ 2.16 ฐานข้อมูล HBVdb [<https://hbvdb.ibcp.fr/HBVdb/>]

- HepSEQ [<http://www.hepseq.org>]

เป็นฐานข้อมูลที่รวบรวมข้อมูลทางคลินิกของผู้ป่วยที่ติดเชื้อ และระบาดวิทยาของเชื้อไวรัสตับอักเสบบี ในปัจจุบันฐานข้อมูลนี้ได้รวบรวมข้อมูลของผู้ป่วยไว้มากกว่า 6 พันราย และมีลำดับนิวคลีโอไทด์ของเชื้อไวรัสตับอักเสบบีมากกว่า 8 พันตัวอย่าง นอกจากนี้ยังมีเครื่องมือในการวิเคราะห์เปรียบเทียบลำดับนิวคลีโอไทด์ การจำแนกสายพันธุ์ และวิเคราะห์การกลายพันธุ์ที่เกี่ยวข้องกับการดื้อยาต้านไวรัสของเชื้อไวรัสตับอักเสบบี (รูปที่ 2.17)



รูปที่ 2.17 ฐานข้อมูล HepSEQ [<http://www.hepseq.org>]

- HBVRegDB [<http://lancelot.otago.ac.nz>]

เป็นฐานข้อมูลที่เน้นการวิเคราะห์หา regulatory sequences บนลำดับนิวคลีโอไทด์หรือจีโนมของเชื้อไวรัสตับอักเสบบี (รูปที่ 2.18) พร้อมทั้งมีเครื่องมือในการวิเคราะห์เปรียบเทียบลำดับนิวคลีโอไทด์ การค้นหาบริเวณอนุรักษ์ (conserved regions) และโครงสร้างทติยภูมิของอาร์เอ็นเอ (RNA secondary structures)

**Hepatitis B Virus Regulatory Sequence Database (HBVRegDB)**

Introduction | Hepadnavirus genomes | Single genomic sequences | Annotated sequence groups (gbrowse) | Genomic alignments | Tutorials | Links

**Resources available here:**

**Introduction to HBVRegDB**

This site provides regulatory sequence resources for hepadnaviridae genomes with annotation and links to related resources.

The target audience is scientists studying HBV molecular biology and regulation of gene expression from the HBV genome. Users please cite this publication: Panjavorayan, Roessner, Firth and Brown, *Virology Journal* 2007, 4: 136 (free access).

Tutorials on how you might use the key resources here are available:

1. Comparison of your sequence to a well annotated HBV genome
  - BLAST 2 sequences to map the numbering from a HBVRegDB reference sequence (e.g. AM282986)
2. Testing for conservation of a sequence, a potential regulatory element, across genomes
3. Testing for conservation of an RNA secondary structures and elements across genomes
4. Repeating similarity searches against HBVRegDB virus sequences, RefSeq viral genomes and proteins
  - BLASTN against local nucleotide databases (RefSeq Virus, HBVRegDB)
  - BLASTP, TBLASTX, BLASTX against local protein databases (RefSeq Virus, Uniprot50)

รูปที่ 2.18 ฐานข้อมูล HBVRegDB [<http://lancelot.otago.ac.nz>]

## ฐานข้อมูลสำหรับเชื้อไวรัสตับอักเสบซี (Databases for hepatitis C virus)

- LANL HCV Database [<http://hcv.lanl.gov>]

เป็นฐานข้อมูลที่รวบรวมลำดับนิวคลีโอไทด์ของเชื้อไวรัสตับอักเสบซี (HCV sequence database) และข้อมูลระบบภูมิคุ้มกันที่ต่อต้านเชื้อไวรัสตับอักเสบซี (HCV immunology database) เช่น CTL epitopes, T helper epitopes และ Antibodies พร้อมทั้งมีเครื่องมือสำหรับการวิเคราะห์ต่างๆ เช่น search, alignment, sequence analysis, phylogenetics และ format converter เป็นต้น (รูปที่ 2.19)

- euHCVdb [<http://euhcvdb.ibcp.fr>]

เป็นฐานข้อมูลที่รวบรวมข้อมูลเกี่ยวกับลำดับนิวคลีโอไทด์ โปรตีน และโครงสร้างสามมิติที่เกี่ยวข้องเชื้อไวรัสตับอักเสบซี พร้อมทั้งมีเครื่องมือสำหรับการวิเคราะห์ลำดับนิวคลีโอไทด์ โปรตีน และโครงสร้างสามมิติ (รูปที่ 2.20)

**HCV DATABASES**

The Hepatitis C Virus (HCV) Database Project was initially funded by the Division of Microbiology and Infectious Diseases of the National Institute of Allergies and Infectious Diseases (NIAID). The HCV database group strives to present HCV-associated, hand-annotated genetic data in a userfriendly way, by providing access to the central database via web-accessible search interfaces and supplying a number of analysis tools.

This project is deeply indebted to the HIV Database Group, led by Dr. Bette Korber, for much of its tools, infrastructure and philosophy; the two databases continue to collaborate closely and to share software and resources.

To our users Please note that the HCV database site is no longer funded by NIH. We try to keep the database updated and the tools running, but we cannot guarantee we can quickly resolve problems.

**HCV SEQUENCE DATABASE**      **HCV IMMUNOLOGY DATABASE**

รูปที่ 2.19 ฐานข้อมูล LANL HCV Database [<http://hcv.lanl.gov>]

**HepCVax**  
**VIRGIL**

Today: 2018-6-4  
Server: UP  
Release: 122 . 0  
Date: 2011-01-26  
Entries: 96645

About  
News  
Help  
Nomenclature  
Static:  
◦ Genome  
◦ Polyprotein  
Query:  
◦ Sequences  
◦ 3D-Models  
HCV Tools:  
◦ Extract  
◦ Number  
Other Tools:  
◦ Geno3D  
◦ Modeome3D  
◦ NPS@  
◦ PIG  
◦ SuMo

**euHCVdb**

**HCV polyprotein**

ER Lumen  
ER Membrane  
Cytosol

To view protein data, click on it

3011 residues (AF009606 - isolate HCV-H77)  
10 proteins

รูปที่ 2.20 ฐานข้อมูล euHCVdb [<http://euhcvdb.ibcp.fr>]



## ฐานข้อมูลสำหรับเชื้อไวรัสเอชไอวี (Databases for human immunodeficiency virus)

- Stanford HIV Drug Resistance DB [<http://hivdb.stanford.edu>]

เป็นฐานข้อมูลที่รวบรวมข้อมูลการรักษาผู้ป่วยด้วยยาต้านไวรัส ผลทางคลินิก คุณสมบัติของเชื้อไวรัส ความหลากหลายทางพันธุกรรม และจำแนกเป็นสายพันธุ์ต่างๆ ของเชื้อไวรัสเอชไอวี แล้วสรุปเป็นข้อมูลความสัมพันธ์ด้านต่างๆ (รูปที่ 2.21) ได้แก่

- สายพันธุ์ของเชื้อไวรัสและการรักษาผู้ป่วย (genotype-treatment)
- สายพันธุ์ของเชื้อไวรัส และคุณสมบัติของเชื้อไวรัส (genotype-phenotype)
- สายพันธุ์ของเชื้อไวรัส และผลทางคลินิก (genotype-clinical outcomes)

Stanford University  
**HIV DRUG RESISTANCE DATABASE**  
A curated public database to represent, store and analyze HIV drug resistance data.

HOME GENOTYPE-RX GENOTYPE-PHENO GENOTYPE-CLINICAL HIVdb PROGRAM ABOUT HIVdb

**Version 8.5 of HIVDB**  
**INSTI: a new drug and comments**

**Query Pages**

- Genotype-treatment**  
Retrieve sequences (and/or mutations) from persons receiving selected HIV drugs  
Retrieve sequences and treatments from viruses with specific mutations
- Genotype-phenotype**  
Retrieve drug susceptibility data for isolates with selected mutations  
Download genotype-phenotype research datasets
- Genotype-clinical**  
Summaries of genotype-clinical outcome studies  
Genotype-clinical outcome datasets (download)
- References**  
Published drug resistance studies in HIVDB  
Published studies by Stanford database group

**New Submissions** [View All](#)

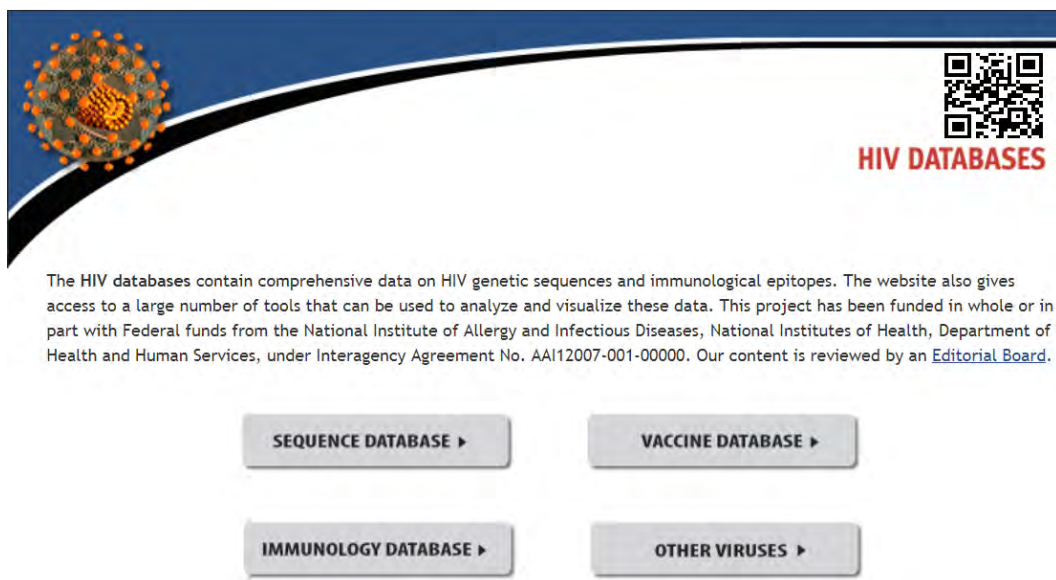
HIVdb Program  
Drug Resistance Summaries (Download PDF)  
PIs NRTIs NNRTIs INSTIs  
HIVseq Program  
HIVaig Program  
HIV-1 Genetic Variability for Drug Resistance

Calibrated Population Resistance  
INTERACTIVE MAP  
Surveillance Mutations  
Point-of-Care / Essential Mutations

รูปที่ 2.21 ฐานข้อมูล Stanford HIV Drug Resistance DB [<http://hivdb.stanford.edu>]

- **LANL HIV database** [<http://hiv.lanl.gov>]

เป็นฐานข้อมูลที่รวบรวมลำดับนิวคลีโอไทด์ของเชื้อไวรัสเอชไอวี (HIV sequence) ผลการทดสอบวัคซีน (vaccine trials) ใน nonhuman primates และข้อมูลทางด้านระบบภูมิคุ้มกัน (immunology) ที่ต่อต้านเชื้อไวรัสเอชไอวี พร้อมทั้งมีเครื่องมือสำหรับการวิเคราะห์รูปแบบต่างๆ เช่น search, alignments และ tools ต่างๆ เป็นต้น (รูปที่ 2.22)



รูปที่ 2.22 ฐานข้อมูล LANL HIV database [<http://hiv.lanl.gov>]

- **bNAber** [<http://bnaber.org>]

เป็นฐานข้อมูลที่รวบรวมแอนติบอดีชนิดต่างๆ ที่จำเพาะต่อเชื้อไวรัสเอชไอวี (broadly neutralizing HIV antibodies) และสามารถวิเคราะห์เปรียบเทียบการจับกันระหว่างแอนติบอดีชนิดต่างๆ และเชื้อไวรัสเอชไอวีสายพันธุ์ต่างๆ ซึ่งช่วยในกระบวนการพัฒนาวัคซีนต้านเชื้อไวรัสเอชไอวีได้ (รูปที่ 2.23)

รูปที่ 2.23 ฐานข้อมูล bNAber [<http://bnaber.org>]

### ฐานข้อมูลสำหรับเชื้อไวรัสปาปิลโลมา (Databases for papillomavirus)

- PAVE [<https://pave.niaid.nih.gov>]

Papillomavirus Episteme (PAVE) เป็นฐานข้อมูลที่รวบรวมข้อมูลจีโนม ยีน และโปรตีนของ papillomavirus ที่พบการติดเชื้อทั้งในคน และสัตว์ชนิดต่างๆ ซึ่งประกอบด้วย 3 ส่วนหลัก (รูปที่ 2.24) ได้แก่

- Search เป็นการค้นหาข้อมูล จีโนม ยีน โปรตีน โครงสร้างสามมิติ และรูปภาพที่เกี่ยวข้องกับ papillomavirus สายพันธุ์ต่างๆ
- Analyze เป็นการวิเคราะห์ข้อมูล ได้แก่ multiple sequence alignment, protein structure viewer และ L1 taxonomy tool เป็นต้น
- Explore เป็นการสำรวจฐานข้อมูล โดยแบ่งออกเป็น reference genomes, variants, transcript maps, proteins, epigenetics, taxonomy, review chapters, image viewer, movies และ phylogenetic tree เป็นต้น



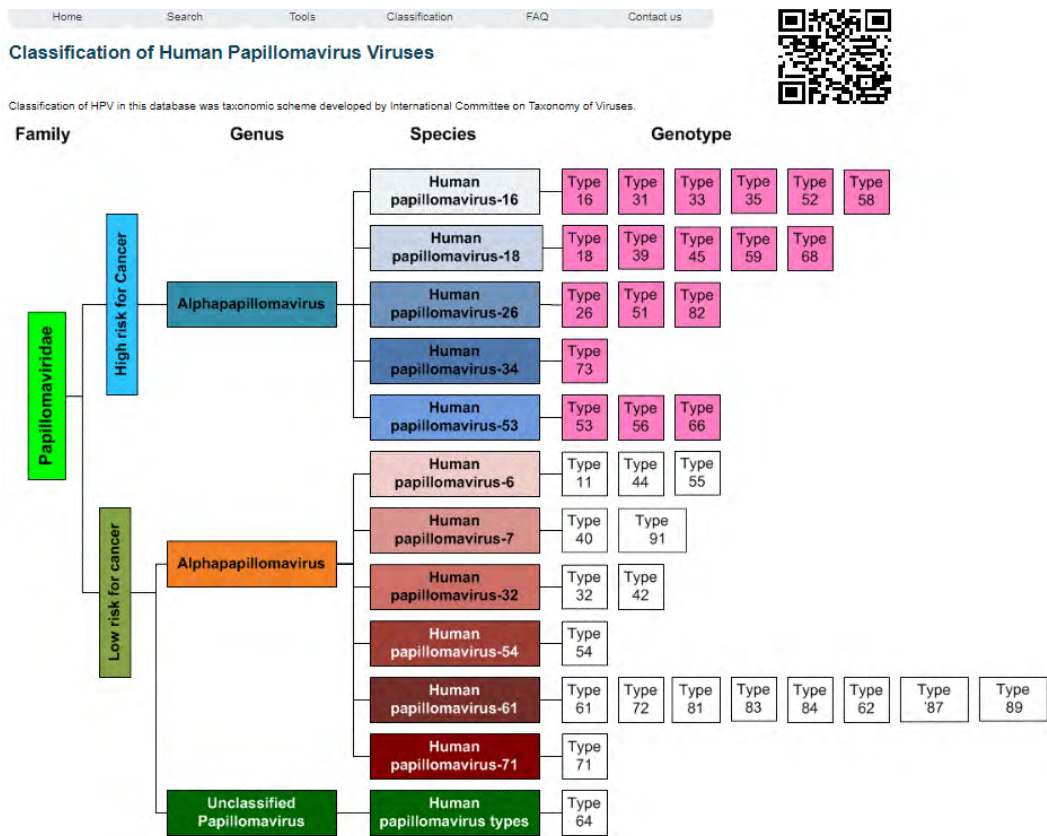
**Search**

**Analyze**

**Explore**

รูปที่ 2.24 ฐานข้อมูล PAVE [<https://pave.niaid.nih.gov>]

- HPV T cell Antigen Database [<http://cvc.dfci.harvard.edu/hpv/index.php>]  
เป็นฐานข้อมูลที่รวบรวมข้อมูล และการวิเคราะห์เกี่ยวกับระบบภูมิคุ้มกันที่ตอบสนองต่อเชื้อ human papillomavirus (HPV) ซึ่งประกอบด้วย 3 ส่วนหลัก ได้แก่
  - Search การค้นหาข้อมูล ได้แก่ HPV viral sequences, HLA ligands และ T-cell epitopes ที่ตอบสนองต่อ HPV เป็นต้น
  - Tools เครื่องมือในการวิเคราะห์ ได้แก่ BLAST, alignment, variability analysis และ view epitope เป็นต้น
  - Classification เป็นระบบการจำแนกสายพันธุ์ของ HPV โดยแบ่งออกเป็น family, genus, species และ genotypes (รูปที่ 2.25)



รูปที่ 2.25 HPV T cell Antigen Database [<http://cvc.dfc.harvard.edu/hpv/index.php>]

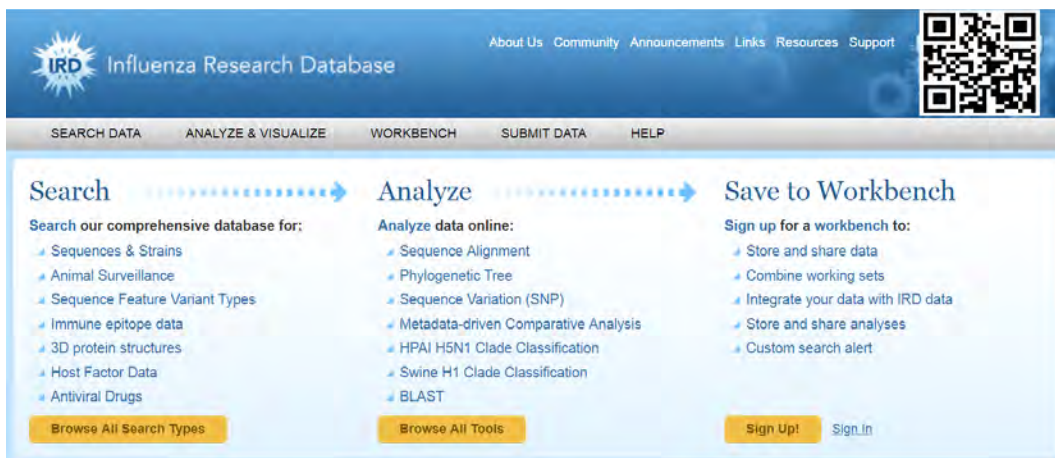
### ฐานข้อมูลสำหรับเชื้อไวรัสไข้หวัดใหญ่ (Databases for influenza virus)

นอกจากฐานข้อมูล Influenza Virus Resource (IVR) ที่อยู่ใน NCBI-VVR ที่กล่าวไปแล้วข้างต้น ยังมีฐานข้อมูลสำคัญอื่นๆ ที่เกี่ยวข้องกับเชื้อไวรัสไข้หวัดใหญ่ ดังนี้

- IRD (Influenza Research Database) [<http://www.fludb.org>]

เป็นฐานข้อมูลที่มีรูปแบบสอดคล้องคล้ายคลึงกับ VIPR แต่ IRD จะเน้นสำหรับการค้นคว้าข้อมูล และการวิเคราะห์ข้อมูลทางอนุชีววิทยาของเชื้อไวรัสไข้หวัดใหญ่โดยเฉพาะ ซึ่งประกอบด้วย 3 ส่วนหลัก (รูปที่ 2.26) ได้แก่

- Search ใช้สำหรับการค้นหาข้อมูลต่างๆ ที่เกี่ยวข้องกับเชื้อไวรัสไข้หวัดใหญ่ เช่น sequences & strains, animal surveillance, sequence feature variant types, 3D structure, immune epitopes, antiviral drugs และ host factors เป็นต้น
- Analysis Tools ใช้สำหรับวิเคราะห์ข้อมูลทางอณูชีววิทยาของเชื้อไวรัส เช่น sequence alignment, phylogenetic tree, sequence variation (SNP), Metadata-driven comparative analysis, HPAI H5N1 clade classification, swine H1 clade classification และ BLAST เป็นต้น
- Workbench เป็นส่วนที่ใช้ในการเก็บบันทึกข้อมูลที่ได้จากการค้นหา หรือการวิเคราะห์ดังกล่าวข้างต้น และสามารถ uploads เพื่อเผยแพร่ หรือส่งต่อให้เพื่อนร่วมงานได้



รูปที่ 2.26 ฐานข้อมูล IRD (Influenza Research Database) [<http://www.fludb.org>]

- **IVDB** [<http://influenza.big.ac.cn>]

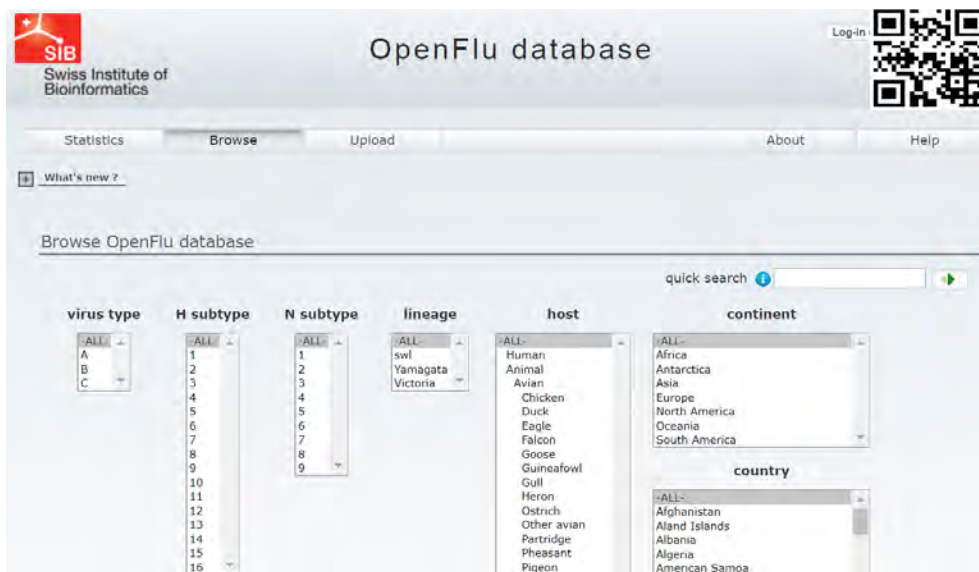
เป็นฐานข้อมูลที่รวบรวมลำดับนิวคลีโอไทด์ และลำดับกรดอะมิโนของเชื้อไวรัสไข้หวัดใหญ่ โดยสามารถตั้งค่าในการค้นหาข้อมูลได้มากมาย เช่น Type, Host, Country/Region, Segment, Year และ Sequence length เป็นต้น (รูปที่ 2.27) นอกจากนี้ยังมีเครื่องมือสำหรับใช้ในการวิเคราะห์ข้อมูล เช่น sequence distribution tool, sequence quality filter system, sequence BLAST, multi-alignment, phylogenetic tree builder, protein structure view และ sequence polymorphism เป็นต้น

รูปที่ 2.27 ฐานข้อมูล IVDB [<http://influenza.big.ac.cn>]

- **OpenFluDB** [<http://platform.gisaid.org>]

เป็นฐานข้อมูลที่รวบรวมลำดับนิวคลีโอไทด์ และลำดับกรดอะมิโนของเชื้อไวรัสไข้หวัดใหญ่ โดยสามารถตั้งค่าในการค้นหาข้อมูลได้มากมาย เช่น virus type, HA subtype, NA subtype, lineage, host, continent และ country เป็นต้น (รูปที่ 2.28) นอกจากนี้ยังสามารถแสดงข้อมูลทางสถิติด้านระบาดวิทยาของเชื้อไวรัสไข้หวัดใหญ่ โดยแบ่งออกเป็น

- Graphic plot กราฟระหว่างปีที่ระบาด และจำนวนผู้ติดเชื้อทั่วโลก
- Geolocation แสดงพื้นที่ที่มีการระบาดบนแผนที่โลก
- Similarity map กราฟแสดงความเหมือนระหว่างสารพันธุกรรมของเชื้อไวรัสไข้หวัดใหญ่สายพันธุ์ต่างๆ



รูปที่ 2.28 ฐานข้อมูล OpenFluDB [<http://platform.gisaid.org>]

### เอกสารอ้างอิง และแนะนำอ่านเพิ่มเติม

1. Carrillo-Tripp M, Shepherd CM, Borelli IA, Venkataraman S, Lander G, Natarajan P, Johnson JE, Brooks CL 3rd, Reddy VS. VIPERdb2: an enhanced and web API enabled relational database for structural virology. *Nucleic Acids Res.* 2009 Jan;37(Database issue):D436-42.
2. Chang S, Zhang J, Liao X, Zhu X, Wang D, Zhu J, Feng T, Zhu B, Gao GF, Wang J, Yang H, Yu J, Wang J. Influenza Virus Database (IVDB): an integrated information resource and analysis platform for influenza virus research. *Nucleic Acids Res.* 2007 Jan;35(Database issue):D376-80.
3. Combet C, Garnier N, Charavay C, Grando D, Crisan D, Lopez J, Dehne-Garcia A, Geourjon C, Bettler E, Hulo C, Le Mercier P, Bartenschlager R, Diepolder H, Moradpour D, Pawlotsky JM, Rice CM, Trépo C, Penin F, Deléage G. euHCVdb: the European hepatitis C virus database. *Nucleic Acids Res.* 2007 Jan;35(Database issue):D363-6.

4. Eroshkin AM, LeBlanc A, Weekes D, Post K, Li Z, Rajput A, Butera ST, Burton DR, Godzik A. bNAber: database of broadly neutralizing HIV antibodies. *Nucleic Acids Res.* 2014 Jan;42(Database issue):D1133-9.
5. Gnaneshan S, Ijaz S, Moran J, Ramsay M, Green J. HepSEQ: International Public Health Repository for Hepatitis B. *Nucleic Acids Res.* 2007 Jan;35(Database issue):D367-70.
6. Hayer J, Jadeau F, Delage G, Kay A, Zoulim F, Combet C. HBVdb: a knowledge database for Hepatitis B Virus. *Nucleic Acids Res.* 2013 Jan;41(Database issue):D566-70.
7. Hulo C, de Castro E, Masson P, Bougueleret L, Bairoch A, Xenarios I, Le Mercier P. ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.* 2011 Jan;39(Database issue):D576-82.
8. Kuiken C, Korber B, Shafer RW. HIV sequence databases. *AIDS Rev.* 2003 Jan-Mar;5(1):52-61. Review.
9. Kuiken C, Yusim K, Boykin L, Richardson R. The Los Alamos hepatitis C sequence database. *Bioinformatics.* 2005 Feb 1;21(3):379-84.
10. Liechti R, Gleizes A, Kuznetsov D, Bougueleret L, Le Mercier P, Bairoch A, Xenarios I. OpenFluDB, a database for human and animal influenza virus. *Database (Oxford).* 2010 Jul 6;2010:baq004.
11. Navratil V, de Chassey B, Meyniel L, Delmotte S, Gautier C, Andr  P, Lotteau V, Rabourdin-Combe C. VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. *Nucleic Acids Res.* 2009 Jan;37(Database issue):D661-8.



12. Panjaworayan N, Roessner SK, Firth AE, Brown CM. HBVRegDB: annotation, comparison, detection and visualization of regulatory elements in hepatitis B virus sequences. *Virol J.* 2007 Dec 17;4:136.
13. Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, Liu M, Kumar S, Zaremba S, Gu Z, Zhou L, Larson CN, Dietrich J, Klem EB, Scheuermann RH. ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res.* 2012 Jan;40(Database issue):D593-8.
14. Shafer RW. Rationale and uses of a public HIV drug-resistance database. *J Infect Dis.* 2006 Sep 15;194 Suppl 1:S51-8. Review.
15. Squires RB, Noronha J, Hunt V, Garcia-Sastre A, Macken C, Baumgarth N, Suarez D, Pickett BE, Zhang Y, Larsen CN, Ramsey A, Zhou L, Zaremba S, Kumar S, Dietrich J, Klem E, Scheuermann RH. Influenza research database: an integrated bioinformatics resource for influenza research and surveillance. *Influenza Other Respir Viruses.* 2012 Nov;6(6):404-16.
16. Van Doorslaer K, Tan Q, Xirasagar S, Bandaru S, Gopalan V, Mohamoud Y, Huyen Y, McBride AA. The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. *Nucleic Acids Res.* 2013 Jan;41(Database issue):D571-8.
17. Yuen LK, Ayres A, Littlejohn M, Colledge D, Edgely A, Maskill WJ, Locarnini SA, Bartholomeusz A. SeqHepB: a sequence analysis program and relational database system for chronic hepatitis B. *Antiviral Res.* 2007 Jul;75(1):64-74.
18. Yusim K, Richardson R, Tao N, Dalwani A, Agrawal A, Szinger J, Funkhouser R, Korber B, Kuiken C. Los alamos hepatitis C immunology database. *Appl Bioinformatics.* 2005;4(4):217-25.

# 3

การวิเคราะห์ข้อมูลลำดับนิวคลีโอไทด์

Nucleotide sequence analysis

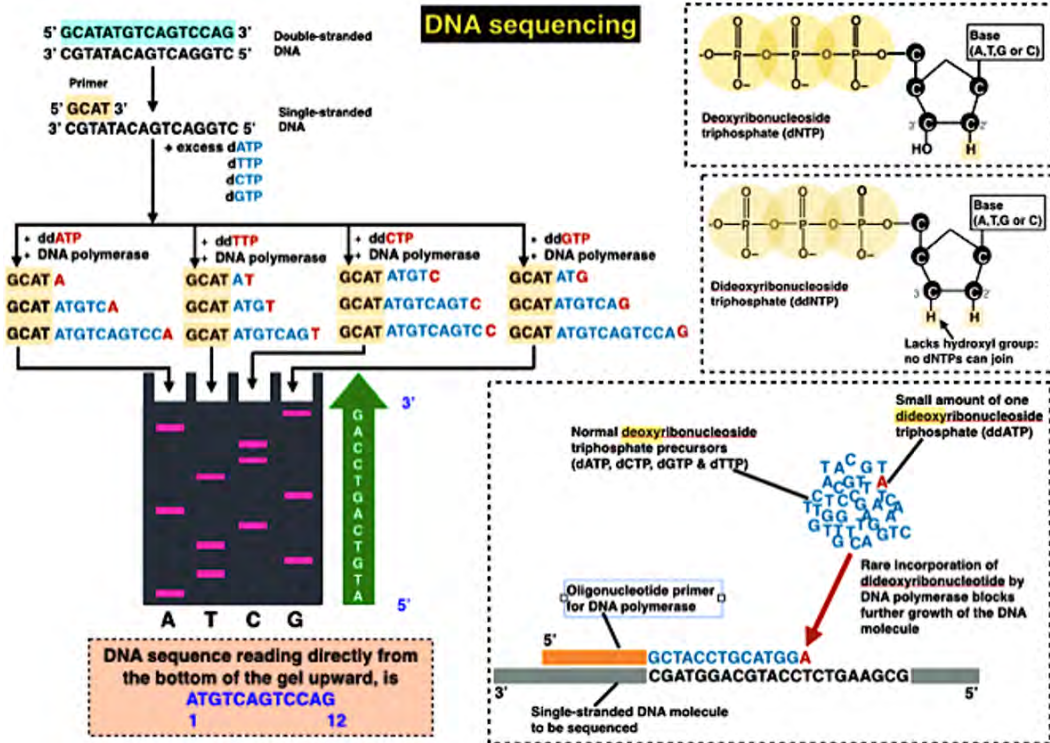
## การวิเคราะห์ข้อมูลลำดับนิวคลีโอไทด์

### Nucleotide sequence analysis

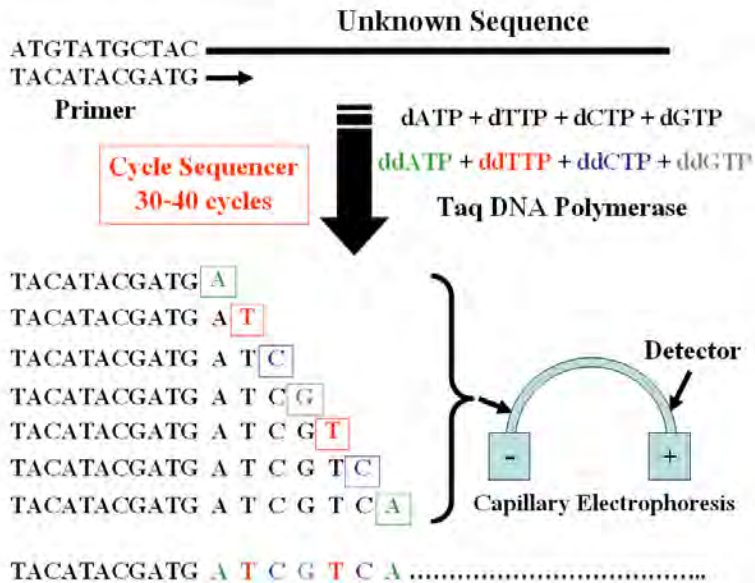
#### เทคนิคการหาลำดับสารพันธุกรรม (DNA sequencing)

การหาลำดับนิวคลีโอไทด์ด้วยวิธี Dideoxy chain termination method เป็นที่นิยม เนื่องจากเป็นวิธีที่ง่ายและมีประสิทธิภาพ โดยในปฏิกิริยาจะประกอบด้วย DNA polymerase, deoxyribonucleotide triphosphate (dNTPs), single primer และ DNA template แต่จะมีการใส่ dideoxyribonucleotide triphosphate (ddNTPs) ได้แก่ ddATP, ddTTP, ddGTP และ ddCTP ซึ่งติดฉลากด้วยสารกัมมันตรังสีหรือสารเรืองแสงที่มีสีแตกต่างกัน โดยที่สาร ddNTPs นี้จะมีโครงสร้างคล้ายคลึงกับ dNTPs ปกติ แต่จะมีความแตกต่างเฉพาะที่คาร์บอนตำแหน่งที่ 3 จะมี H แทนที่ OH (รูปที่ 3.1) ซึ่งคุณสมบัตินี้จะทำให้ไม่สามารถเกิดพันธะ phosphodiester bond ต่อสาย DNA ได้ จึงเกิดการหยุดต่อสาย (Chain termination) ซึ่งในปฏิกิริยาการต่อสาย DNA polymerase จะมีการนำเอาทั้ง dNTP และ ddNTP มาใช้ในการสร้างสาย DNA ใหม่ ดังนั้นปฏิกิริยาที่เกิดขึ้นจึงมีทั้งการต่อสาย (เมื่อใช้ dNTP) และการหยุดต่อสาย (เมื่อใช้ ddNTP) ผสมกัน เมื่อนำผลิตภัณฑ์ที่ได้จากปฏิกิริยามา run ใน polyacrylamide gel electrophoresis แล้วทำ autoradiography โดยใช้ฟิล์ม X-rays ก็จะได้แถบที่ใช้ในการอ่านลำดับ DNA ได้ (รูปที่ 3.1)

ในปัจจุบันมีการพัฒนาหลักการของวิธี Chain termination method มาใช้กับเครื่องอ่านลำดับ DNA อัตโนมัติ (Automated DNA sequencer) โดยใช้ ddNTP ที่ติดฉลากด้วยสารเรืองแสงที่มีสีแตกต่างกันใน ddNTP แต่ละชนิด จึงทำให้สามารถใส่รวมกันให้เกิดปฏิกิริยาในหลอดเดียวกันได้ หลังจากปฏิกิริยาเสร็จสิ้น สามารถนำมา Run ใน Capillary electrophoresis ซึ่งมีความสามารถในการแยกได้ดี นอกจากนี้มีตัวตรวจวัด (Detector) ซึ่งสามารถแยกสัญญาณการเรืองแสงที่มีความยาวคลื่นแตกต่างกันของสารเรืองแสงแต่ละชนิด ทำให้สามารถแยกและวิเคราะห์ลำดับ DNA ได้อย่างมีประสิทธิภาพ (รูปที่ 3.2)



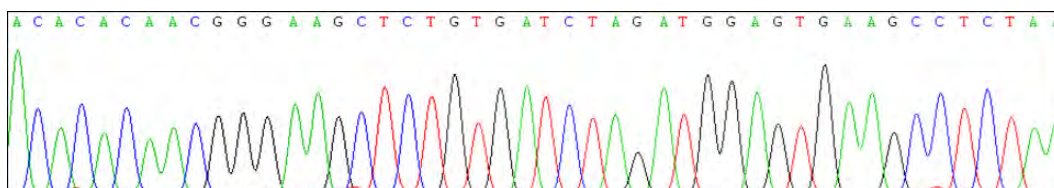
รูปที่ 3.1 หลักการหาลำดับสารพันธุกรรม ด้วยวิธี Chain termination method



รูปที่ 3.2 หลักการทำงานของเครื่อง automated DNA sequencer

### ➤ กราฟลำดับนิวคลีโอไทด์ (Nucleotide chromatogram)

ข้อมูลลำดับนิวคลีโอไทด์ที่ได้จากเครื่อง automated DNA sequencer จะเป็นไฟล์ข้อมูล (.ab1 file) ที่เรียกว่ากราฟลำดับนิวคลีโอไทด์ (nucleotide chromatogram) เมื่อเปิดไฟล์ในโปรแกรมที่ใช้ในการวิเคราะห์ จะแสดงผลเป็นกราฟที่แสดงระดับ (intensity level) ของสัญญาณเรืองแสง 4 สี (A สีเขียว, T สีแดง, C สีน้ำเงิน และ G สีดำ) และลำดับเบสที่ได้จากการแปลผลอัตโนมัติ (รูปที่ 3.3)

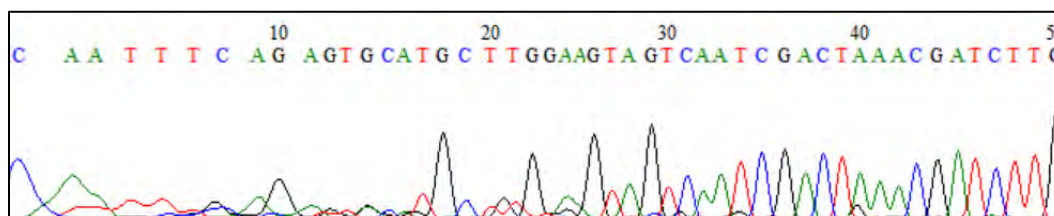


รูปที่ 3.3 ตัวอย่างข้อมูลที่ได้จากเครื่อง automated DNA sequencer

### ➤ การวิเคราะห์กราฟลำดับนิวคลีโอไทด์

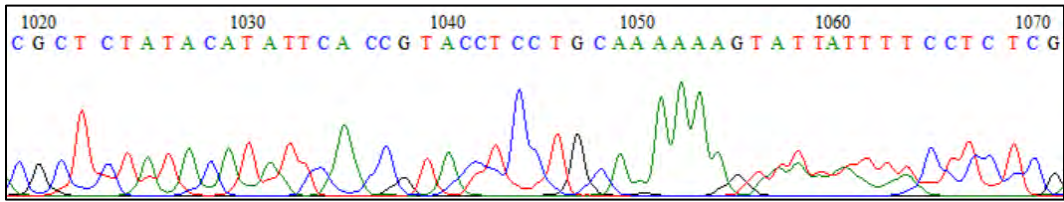
โดยทั่วไปก่อนที่จะนำลำดับนิวคลีโอไทด์ที่ได้จากเครื่อง automated DNA sequencer มาวิเคราะห์ ควรตรวจสอบเบื้องต้นก่อนว่าข้อมูลที่ได้นั้นมีความถูกต้องน่าเชื่อถือหรือไม่ และจำเป็นต้องตัดลำดับนิวคลีโอไทด์ (nucleotide trimming) ในส่วนที่เป็น noise signal ซึ่งเป็นข้อมูลที่ไม่ถูกต้องหรือมีความน่าเชื่อถือน้อยออกไป ดังนี้

#### ➤ กราฟลำดับนิวคลีโอไทด์ช่วงต้น มักอยู่ในช่วงตำแหน่ง 50-100 bp (รูปที่ 3.4)



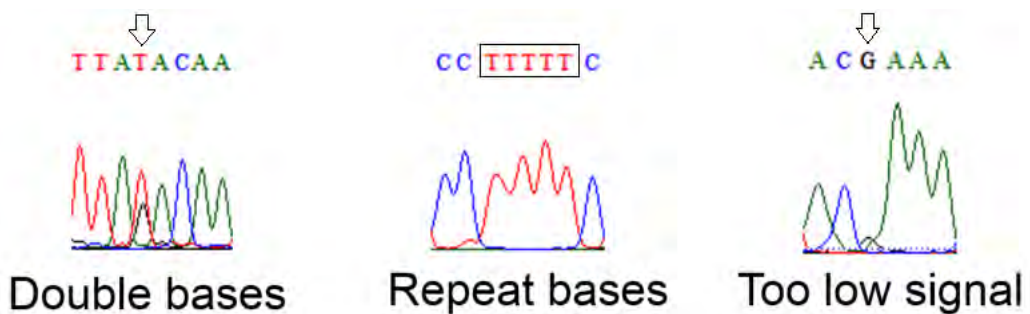
รูปที่ 3.4 ตัวอย่าง Noise signal ในช่วงต้นของกราฟลำดับนิวคลีโอไทด์

#### ➤ กราฟลำดับนิวคลีโอไทด์ช่วงท้าย มักอยู่ในช่วงตำแหน่งที่มากกว่า 900 bp (รูปที่ 3.5)



รูปที่ 3.5 ตัวอย่าง Noise signal ในช่วงท้ายของกราฟลำดับนิวคลีโอไทด์

- Unclear signals กราฟลำดับนิวคลีโอไทด์ไม่ชัดเจน (รูปที่ 3.6) ได้แก่
  - Double peaks จากรูปที่ 3.6 จะเห็นกราฟลำดับนิวคลีโอไทด์ซ้อนทับกันระหว่างกราฟสีแดง และกราฟสีดํา แต่กราฟสีแดงมี peak ที่สูงกว่า จึงทำให้เครื่องอ่านผลเป็นเบส T เท่านั้น ซึ่งกรณีที่พบกราฟซ้อนกัน อาจแปลผลได้หลายแบบ เช่น sequencing error, heterozygous alleles หรือ minor mutation ดังนั้นจึงควรทำ DNA sequencing ซ้ำอีกครั้งหรือทำ bidirectional sequencing เพื่อเพิ่มความถูกต้องแม่นยำในการแปลผล
  - Repeat bases ในกรณีที่ลำดับเบสซ้ำๆ กัน อาจทำให้เครื่องอ่านสัญญาณเรืองแสงผิดพลาด จากรูปที่ 3.6 จะเห็นกราฟสีแดงจำนวน 4 peaks ติดต่อกัน แต่เครื่องอ่านผลเป็นเบส T จำนวน 5 ตัว ดังนั้นจึงควรสังเกตลักษณะ peaks ดังกล่าวเพื่อแก้ไขข้อมูลลำดับเบสให้ถูกต้อง
  - Too low signal ในบางกรณีอาจพบสัญญาณเรืองแสงที่มี peak ต่ำมาก จึงควรพิจารณาว่าเป็น noise signal หรือ เป็น peak ของลำดับเบสจริงๆ (รูปที่ 3.6)

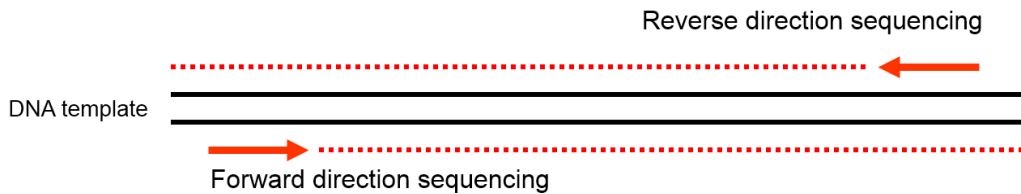


รูปที่ 3.6 ตัวอย่าง unclear signals ชนิดต่างๆ



➤ การหาลำดับนิวคลีโอไทด์สองทิศทาง (Bidirectional sequencing)

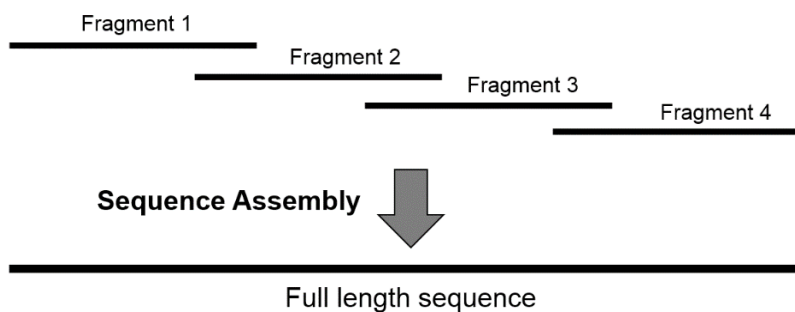
เป็นการหาลำดับนิวคลีโอไทด์ 2 ครั้ง โดยใช้ forward primer และ reverse primer ซึ่งมีทิศสวนทางกัน จะทำให้ได้ผลลำดับนิวคลีโอไทด์บริเวณยืนเดียวกัน จึงเป็นการยืนยันผลซึ่งกันและกัน (รูปที่ 3.7) แต่ในบางกรณีอาจพบว่าผลที่ได้จาก bidirectional sequencing มีความขัดแย้งกัน จึงควรทำ DNA sequencing ซ้ำแบบ duplicates หรือ triplicates เพื่อให้ได้ผลลำดับนิวคลีโอไทด์ที่มีความถูกต้องมากขึ้น



รูปที่ 3.7 แผนภาพแสดงกระบวนการ bidirectional sequencing

การนำข้อมูลลำดับนิวคลีโอไทด์มาเชื่อมต่อกัน (Nucleotide sequence assembly)

เนื่องจากลำดับนิวคลีโอไทด์ที่ได้จากเครื่อง automated DNA sequencer มีความยาวจำกัดอยู่ที่ประมาณ 800-1,000 bp แต่ในการศึกษาหาลำดับของยีนต่างๆ หรือจีโนมจะมีความยาวของลำดับนิวคลีโอไทด์มากกว่านั้น จึงมีความจำเป็นในการเพิ่มจำนวน DNA เป็นท่อนๆ ที่มีส่วนซ้อนทับกันอย่างน้อย 20-30 bp (overlapping fragments) จากนั้นหาลำดับนิวคลีโอไทด์ของแต่ละท่อน แล้วนำข้อมูลลำดับนิวคลีโอไทด์มาเชื่อมต่อกัน (nucleotide sequence assembly) เพื่อให้ได้เป็นยีนที่สมบูรณ์ (full length gene) หรือจีโนม (genome) ดังรูปที่ 3.8



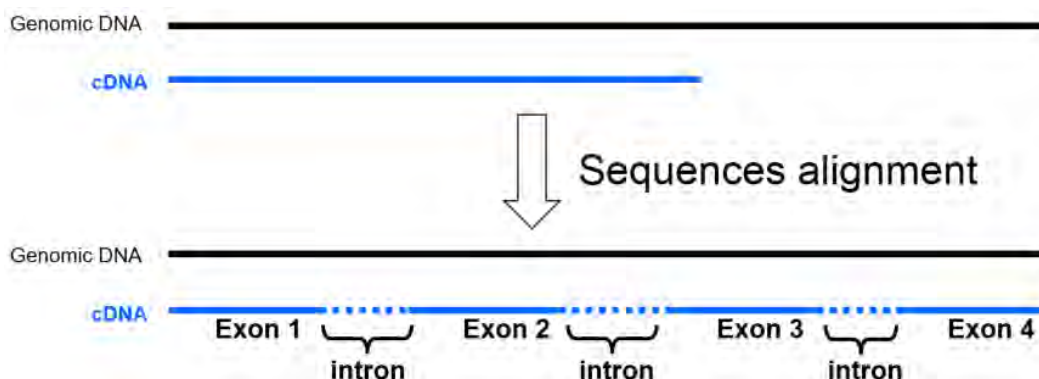
รูปที่ 3.8 แผนภาพแสดง nucleotide sequence assembly

## การจัดเรียงลำดับนิวคลีโอไทด์ (Nucleotide sequences alignment)

การจัดเรียงลำดับนิวคลีโอไทด์ ใช้ในการวิเคราะห์เปรียบเทียบว่าลำดับนิวคลีโอไทด์ของยีนที่สนใจมีลำดับเบสเหมือนหรือแตกต่างกันอย่างไร โดยสามารถเปรียบเทียบลำดับนิวคลีโอไทด์เป็นคู่ จำนวน 2 สาย (pair sequences alignment) หรือเปรียบเทียบลำดับนิวคลีโอไทด์จำนวนหลายสายพร้อมๆ กัน (multiple sequences alignment) ซึ่งจัดได้ว่าเป็นการวิเคราะห์ที่มีความสำคัญมาก สามารถประยุกต์ใช้ในการศึกษาทางอณูชีววิทยาด้านต่างๆ ดังนี้

### ➤ การศึกษารูปแบบ RNA splicing

การแสดงออกของยีนเกี่ยวข้องกับกระบวนการสำคัญที่เรียกว่า Central Dogma ซึ่งประกอบด้วยกระบวนการถอดรหัส (transcription) จาก DNA ไปสู่ RNA และกระบวนการแปลรหัส (translation) จาก RNA ไปเป็นโปรตีน (polypeptide) ในสิ่งมีชีวิตที่เป็น eukaryotes จะมีเยื่อหุ้มนิวเคลียสแบ่ง compartment อย่างชัดเจนทำให้กระบวนการถอดรหัสเกิดขึ้นในนิวเคลียส จากนั้น RNA ที่ได้จะต้องผ่านกระบวนการ RNA processing ซึ่งประกอบด้วย การเติม 5'-capping, RNA splicing และ poly A-tails ทำให้ได้เป็น mature RNA ที่มีความเสถียร แล้วเคลื่อนที่ออกไปสู่ไซโตพลาสซึมของเซลล์เพื่อเกิดกระบวนการแปลรหัสให้ได้เป็นโปรตีน ซึ่งในขั้นตอน RNA splicing จะเป็นการนำเอาส่วนที่เป็น introns ออกไป แล้วนำส่วนที่เป็น exons มาเชื่อมต่อกัน โดยรูปแบบการเกิด RNA splicing มีได้หลายรูปแบบ (alternative splicing) หากต้องการศึกษารูปแบบดังกล่าว สามารถจัดเรียงเปรียบเทียบลำดับนิวคลีโอไทด์ของยีนที่สนใจระหว่าง genomic DNA (มีทั้ง introns และ exons) และ complementary DNA (มีแต่ exons) โดยลำดับนิวคลีโอไทด์ส่วนที่เหมือนกันระหว่าง genomic DNA และ complementary DNA จะเป็นส่วน exons ในขณะที่ส่วนที่ขาดหายไปเป็น complementary DNA ก็คือส่วนที่เป็น introns ดังนั้นด้วยการเปรียบเทียบแบบ pair sequences alignment จะทำให้ศึกษารูปแบบการเกิด RNA splicing ของยีนที่สนใจได้ (รูปที่ 3.9)



รูปที่ 3.9 การศึกษา RNA splicing ด้วยวิธี pair sequences alignment

➤ การศึกษาความเหมือน (identity) และความแตกต่าง (distance)

การศึกษาเปรียบเทียบลำดับนิวคลีโอไทด์ หรือลำดับกรดอะมิโน แบบ multiple sequences alignment สามารถใช้ในการคำนวณหาค่าความเหมือน (identity score) และค่าความแตกต่าง (distance score) ระหว่างสิ่งมีชีวิตแต่ละ species หรือ ไวรัสแต่ละสายพันธุ์ได้ ซึ่งสามารถประยุกต์ใช้ประโยชน์ได้ดังนี้

- ค่าความเหมือน (identity score) นิยมใช้ในการบ่งชี้ว่าลำดับนิวคลีโอไทด์ หรือลำดับกรดอะมิโนที่เราสนใจ (query sequence) นั้นเป็นยีน หรือโปรตีนชนิดใด และเป็นสิ่งมีชีวิตชนิดใด ตัวอย่างเช่น หากมีลำดับนิวคลีโอไทด์ของเชื้อไวรัสตับอักเสบบี (HBV) ที่แยกได้จากผู้ป่วย แล้วต้องการจำแนกว่าเชื้อ HBV ดังกล่าวเป็นสายพันธุ์ (genotype) ใด ก็สามารถใช้ multiple sequences alignment เปรียบเทียบกับลำดับนิวคลีโอไทด์ของเชื้อไวรัสตับอักเสบบี (HBV) genotypes ต่างๆ แล้วคำนวณหาค่าความเหมือน (identity) จากตารางที่ 3.1 บ่งชี้ว่า query sequence น่าจะจำแนกเป็น HBV genotype C เนื่องจากมีความเหมือนมากที่สุด (97.70%) นอกจากนี้หลักการดังกล่าวยังประยุกต์ใช้เป็น เครื่องมือในการค้นหาเปรียบเทียบความคล้ายคลึงกันทางชีวภาพ (Basic Local Alignment Search Tool) [BLAST] ซึ่งจะกล่าวรายละเอียดในบทต่อไป

### ตารางที่ 3.1 ค่าความเหมือน (identity) ระหว่างเชื้อไวรัสตับอักเสบบี (HBV) genotypes ต่างๆ

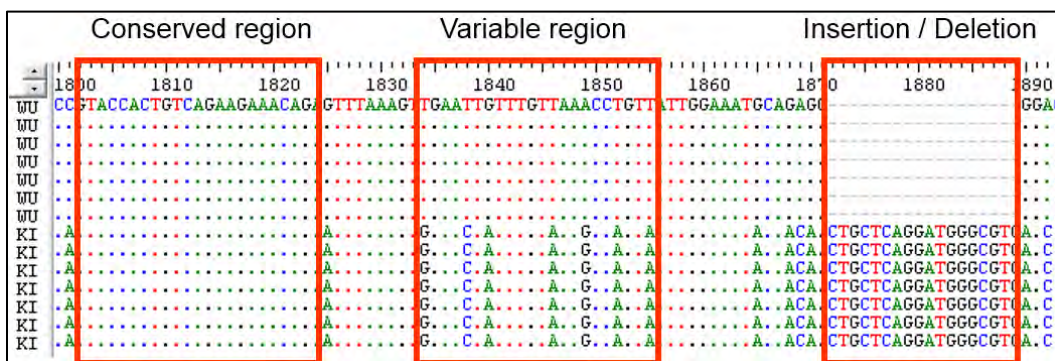
	Genotype_A	Genotype_B	Genotype_C	Genotype_D	Genotype_E	Genotype_F	Genotype_G	Genotype_H	Query
Genotype_A	ID								
Genotype_B	90.40%	ID							
Genotype_C	90.70%	91.10%	ID						
Genotype_D	89.50%	88.90%	89.00%	ID					
Genotype_E	89.70%	89.30%	89.90%	91.90%	ID				
Genotype_F	85.70%	86.10%	85.90%	85.90%	86.50%	ID			
Genotype_G	87.50%	86.60%	86.70%	87.20%	88.60%	84.60%	ID		
Genotype_H	85.50%	86.00%	86.20%	85.70%	86.00%	91.30%	84.40%	ID	
Query	90.30%	91.10%	97.70%	89.20%	89.60%	86.00%	86.90%	86.30%	ID

- ค่าความแตกต่าง (distance score) เป็นค่าที่ตรงกันข้ามกับค่าความเหมือน นิยมใช้ในการวิเคราะห์ความสัมพันธ์เชิงวิวัฒนาการ (Phylogenetic analysis) ชนิด distance method ซึ่งจะกล่าวรายละเอียดในบทต่อไป

#### ➤ การศึกษาบริเวณ conserved และ variable regions

การศึกษาเปรียบเทียบลำดับนิวคลีโอไทด์แบบ multiple sequences alignment สามารถใช้ในการจำแนกบริเวณที่เป็นลักษณะสำคัญของยีน (รูปที่ 3.10) ได้ดังนี้

- บริเวณอนุรักษ์ (conserved region) เป็นบริเวณที่ไม่พบการกลายพันธุ์ ซึ่งบ่งชี้ว่าลำดับนิวคลีโอไทด์บริเวณนี้ ถูกใช้ในการสร้างกรดอะมิโนที่เป็นองค์ประกอบสำคัญในด้านโครงสร้างและหน้าที่ของโปรตีน (เช่น เป็น binding site หรือ active site) ถ้าหากเกิดการกลายพันธุ์บริเวณนี้จะทำให้โปรตีนทำงานไม่ได้ และสิ่งมีชีวิตจะไม่สามารถดำรงชีพได้ จึงจำเป็นต้องอนุรักษ์บริเวณดังกล่าวให้คงอยู่แบบเดิม นอกจากนี้ส่วน conserved regions บนลำดับนิวคลีโอไทด์ ได้รับความนิยมในการเลือกใช้เพื่อออกแบบ primers สำหรับใช้ในเทคนิค polymerase chain reaction (PCR) ซึ่งจะกล่าวรายละเอียดในบทต่อไป
- บริเวณแปรเปลี่ยน (variable region) เป็นบริเวณที่พบการกลายพันธุ์ค่อนข้างมาก ซึ่งบ่งชี้ว่าสิ่งมีชีวิตมีการตอบสนองต่อสิ่งแวดล้อม ทำให้มีการปรับเปลี่ยนลำดับนิวคลีโอไทด์บริเวณนี้ เพื่อให้สามารถดำรงชีพอยู่ได้
- บริเวณที่เพิ่มขึ้น หรือขาดหายไป (insertion/deletion) เป็นบริเวณที่พบการเปลี่ยนแปลงที่อาจส่งผลต่อโครงสร้างและหน้าที่ของยีนและโปรตีนได้



รูปที่ 3.10 ตัวอย่างการเปรียบเทียบลำดับนิวคลีโอไทด์ระหว่าง WU และ KI polyomavirus เพื่อใช้จำแนก conserved/variable regions และ insertion/deletion regions

➤ การวิเคราะห์การกลายพันธุ์ (Mutation analysis)

การกลายพันธุ์ระดับยีน แบ่งออกได้เป็นหลายรูปแบบ ดังนี้

- Substitution การเปลี่ยนแปลงลำดับเบส ซึ่งแบ่งออกเป็น
  - Purine transition เปลี่ยนระหว่าง purine bases (A ↔ G)
  - Pyrimidine transition เปลี่ยนระหว่าง pyrimidine bases (C ↔ T)
  - Transversion เปลี่ยนระหว่าง pyrimidine (C/T) ↔ purine (A/G)
- Point mutation ที่เกิดขึ้นบริเวณที่เป็น coding sequence ของยีน แบ่งออกเป็น
  - Synonymous (silent) mutation มีการเปลี่ยนแปลงลำดับเบสของยีน แต่ลำดับของกรดอะมิโนคงเดิม
  - Nonsynonymous mutation มีการเปลี่ยนแปลงลำดับเบสของยีน แล้วทำให้การแปลรหัสเปลี่ยนแปลงไป (ตารางที่ 3.2) ได้แก่
    - Missense mutation มีการเปลี่ยนแปลงลำดับเบสของยีน แล้วทำให้ลำดับของกรดอะมิโนเปลี่ยนแปลงไป
    - Nonsense mutation มีการเปลี่ยนแปลงลำดับเบสของยีน แล้วกลายเป็น stop codon ซึ่งจะทำให้หยุดกระบวนการแปลรหัส

- Insertion mutation มีลำดับนิวคลีโอไทด์เพิ่มขึ้น
- Deletion mutation มีลำดับนิวคลีโอไทด์ขาดหายไป
- Frame shift mutation เป็นผลมาจากมีลำดับนิวคลีโอไทด์เพิ่มขึ้น (insertion) หรือ ลำดับนิวคลีโอไทด์ขาดหายไป (deletion) ทำให้เกิดการเปลี่ยนแปลง coding frame ของรหัส codon ที่ประกอบด้วยลำดับเบสจำนวน 3 ตัว จึงมีผลทำให้ได้ ลำดับกรดอะมิโนเปลี่ยนแปลงไปจากเดิมอย่างมาก

ตารางที่ 3.2 รหัส codon บน mRNA สำหรับการแปลรหัสเป็นกรดอะมิโน

First position (5' end)	Second position				Third position (3' end)
	A	G	C	U	
A	Lys	Arg	Thr	Ile	A
A	Lys	Arg	Thr	<b>Met</b>	G
A	Asn	Ser	Thr	Ile	C
A	Asn	Ser	Thr	Ile	U
G	Glu	Gly	Ala	Val	A
G	Glu	Gly	Ala	Val	G
G	Asp	Gly	Ala	Val	C
G	Asp	Gly	Ala	Val	U
C	Gln	Arg	Pro	Leu	A
C	Gln	Arg	Pro	Leu	G
C	His	Arg	Pro	Leu	C
C	His	Arg	Pro	Leu	U
U	<b>STOP</b>	<b>STOP</b>	Ser	Leu	A
U	<b>STOP</b>	Trp	Ser	Leu	G
U	Tyr	Cys	Ser	Phe	C
U	Tyr	Cys	Ser	Phe	U



## โปรแกรมในการวิเคราะห์ลำดับนิวคลีโอไทด์ และลำดับกรดอะมิโน

ในปัจจุบันมีโปรแกรมสำหรับการวิเคราะห์ลำดับนิวคลีโอไทด์ และลำดับกรดอะมิโน จำนวนหลายโปรแกรม ในที่นี้จะขอกล่าวถึงเฉพาะโปรแกรม BioEdit (Biological sequence alignment editor for Window) เนื่องจากเป็นโปรแกรมที่สามารถใช้งานได้ฟรีโดยไม่เสียค่าใช้จ่าย (freeware) มีรูปแบบการใช้งานง่ายไม่ซับซ้อน มีคำสั่งให้เลือกใช้งานวิเคราะห์ได้หลากหลายประเภท เป็นโปรแกรมที่มีมาตรฐานเป็นที่ยอมรับสำหรับงานวิจัยในระดับนานาชาติ และผู้เขียนมีประสบการณ์ในการใช้งานโปรแกรมนี้มาหลายปี จึงขอสรุปขั้นตอนการใช้งานของโปรแกรม BioEdit ดังนี้

1. Download โปรแกรม ได้จาก <http://www.mbio.ncsu.edu/BioEdit/bioedit.html> (รูปที่ 3.11) โดยคลิกที่คำว่า “BioEdit.zip (full installation 12.6 Mb)” เพื่อดำเนินการ download โปรแกรม เมื่อ download เสร็จแล้ว ก็ดำเนินการติดตั้งโปรแกรมโดยเลือก file ที่ชื่อว่า “setup.exe”



Copyright © 1997-2005  
Tom Hall  
Ibis Therapeutics  
Carlsbad, CA 92008

**BioEdit**  
Biological sequence alignment editor for Win95/98/NT/2K/XP

<b>BioEdit.zip</b> (full installation 12.6 Mb)
<b>Bug fixes / changes</b>
<b>BioEdit</b> General information
<b>BioDoc.pdf</b> (pdf format help doc)
<b>View Screenshots</b>

**BioEdit** is a biological sequence alignment editor written for Windows 95/98/NT/2000/XP. An intuitive multiple document interface with convenient features makes alignment and manipulation of sequences relatively easy on your desktop computer. Several sequence manipulation and analysis options and links to external analysis programs facilitate a working environment which allows you to view and manipulate sequences with simple point-and-click operations.

**BioEdit's features include:**

- Several modes of hand alignment
- Automated ClustalW alignment
- Automated Blast searches (local and WWW)
- Plasmid drawing and annotation
- Accessory application configuration
- Restriction mapping
- RNA comparative analysis tools
- Graphical matrix data viewing tools
- Shaded alignment figures
- Translation-based nucleic acid alignment
- ABI trace viewing, editing and printing
- Customizable ...other features

**New version is WinXP compatible**

รูปที่ 3.11 การ download โปรแกรม BioEdit

[<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>]

2. เมื่อติดตั้งโปรแกรมเสร็จแล้ว ก็สามารถเปิดโปรแกรมขึ้นมาใช้งานงานได้ทันที จากนั้นสามารถเลือก file ที่ต้องการเปิดใช้งานในการวิเคราะห์ ซึ่งโปรแกรม BioEdit สามารถรองรับ files ชนิดต่างๆ ได้แก่

- GenBank Files (\*.gbk, \*.gen, \*.gb, \*.gnk)
- FASTA Files (\*.fas, \*.fst, \*.fsa, \*.fasta)
- NBRF/PIR Files (\*.pir, \*.nbr)
- Phylip Files (\*.phy)
- Text Files (\*.txt, \*.rtf)
- Plasmid Files (\*.pmd)
- BioEdit Project (\*.bio)
- Clustal Files (\*.aln)
- ABI Files (\*.abi, \*.ab1)
- GCG Files (\*.gcg)
- XML Files (\*.xml)

3. การใช้โปรแกรม BioEdit ในการวิเคราะห์ลำดับนิวคลีโอไทด์ ประเภทต่างๆ

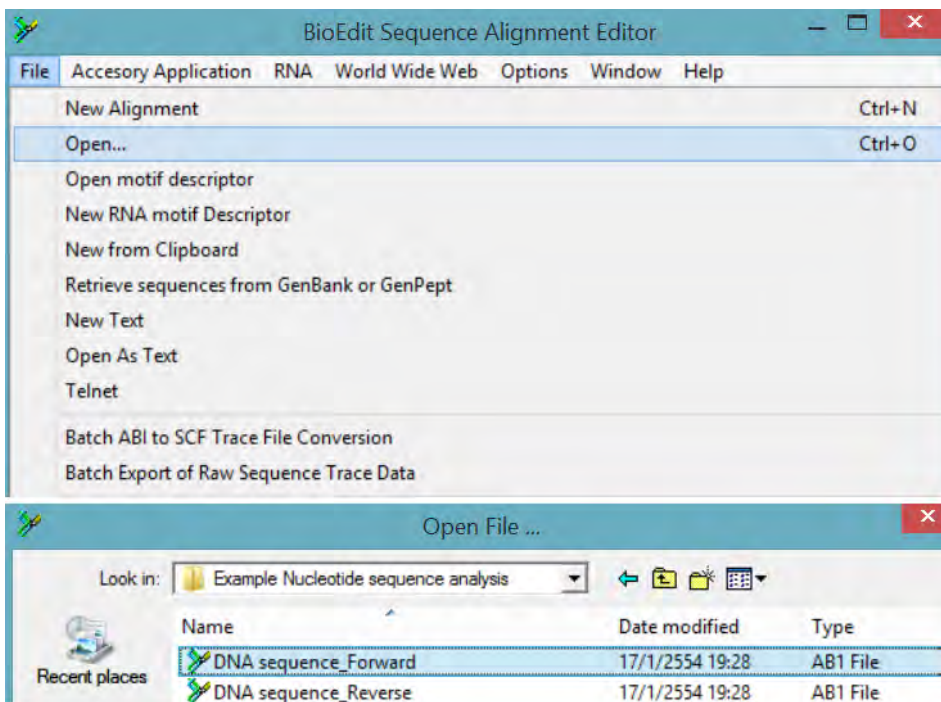
#### ➤ การวิเคราะห์กราฟลำดับนิวคลีโอไทด์โดยใช้โปรแกรม BioEdit

1. เริ่มต้นจากการเปิด ABI File (.ab1) ที่ได้จากเครื่อง automated DNA sequencer โดยคลิกที่ File > Open แล้วเลือก file ที่ต้องการเปิด (รูปที่ 3.12) โปรแกรมจะแสดงผลเป็น 2 ส่วน คือหน้าแสดงกราฟลำดับนิวคลีโอไทด์ และหน้าแสดงเฉพาะลำดับนิวคลีโอไทด์ (รูปที่ 3.13)

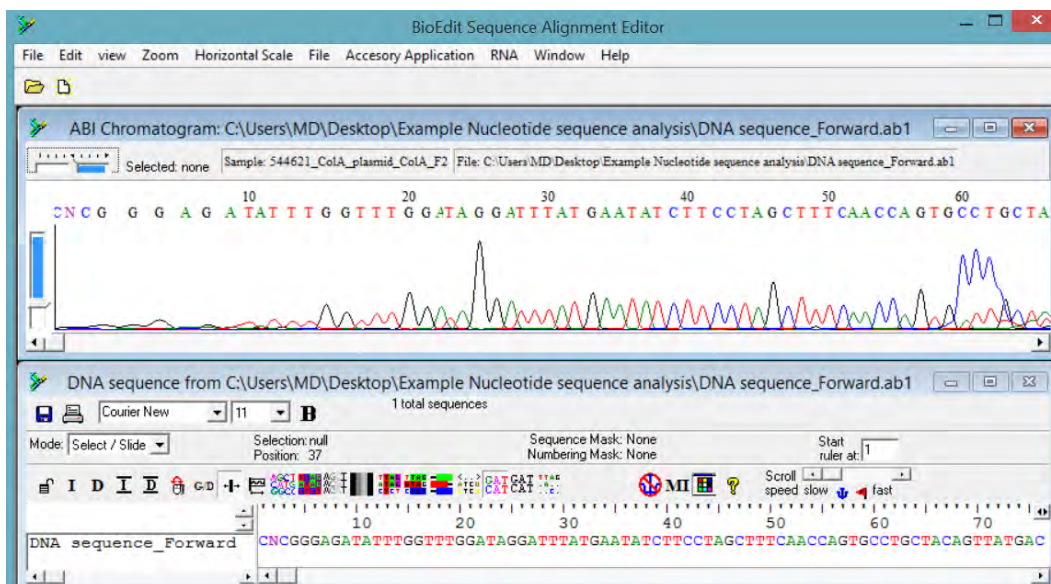
2. ในหน้าที่แสดงกราฟลำดับนิวคลีโอไทด์ (รูปที่ 3.13 ส่วนบน) ผู้ใช้สามารถปรับขนาดของกราฟลำดับนิวคลีโอไทด์ ได้ดังนี้

- Vertical scale เพื่อปรับความสูงของกราฟ
- Horizontal scale เพื่อปรับความกว้างของกราฟ

การวิเคราะห์ข้อมูลลำดับนิวคลีโอไทด์



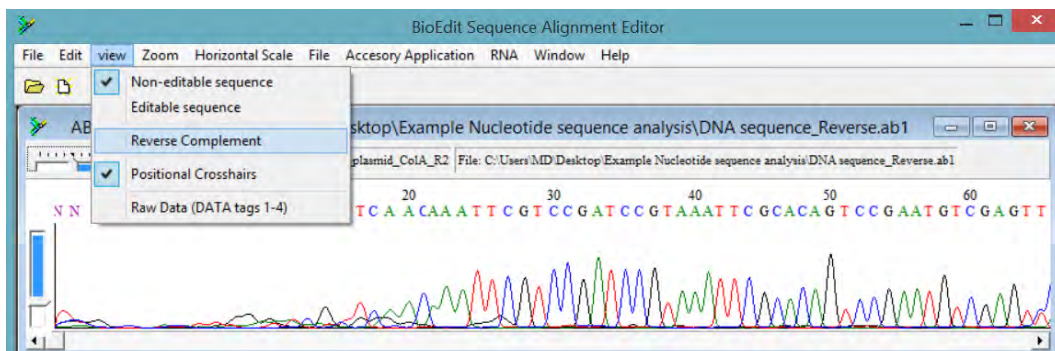
รูปที่ 3.12 การเปิด ABI file โดยใช้โปรแกรม BioEdit



รูปที่ 3.13 การแสดงผลกราฟลำดับนิวคลีโอไทด์ โดยใช้โปรแกรม BioEdit

3. ในกรณีที่ทำ bidirectional sequencing คือการหาลำดับนิวคลีโอไทด์ 2 ครั้ง โดยใช้ forward primer และ reverse primer ซึ่งมีทิศสวนทางกัน ในการวิเคราะห์ผลของกราฟลำดับนิวคลีโอไทด์ที่ได้จาก reverse primer จำเป็นจะต้องทำการกลับทิศทาง และเปลี่ยนเป็นเบสคู่สม (Reverse and Complement) เพื่อให้ได้ผลลำดับนิวคลีโอไทด์ เป็น strand ที่ตรงกับผลที่ได้จาก forward primer เนื่องจากโปรแกรมการวิเคราะห์ลำดับนิวคลีโอไทด์นิยมแสดงผลและวิเคราะห์ลำดับนิวคลีโอไทด์ของสายที่เป็น sense strand (+) เท่านั้น ซึ่งโปรแกรม BioEdit สามารถทำ Reverse and Complement ได้ ดังนี้

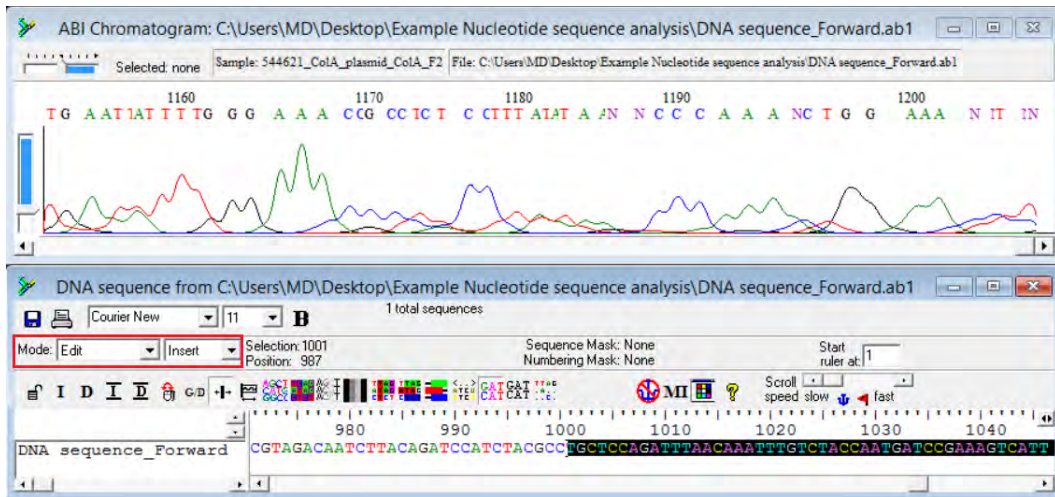
- ในหน้าแสดงกราฟลำดับนิวคลีโอไทด์ ให้เลือกที่คำสั่ง View > Reverse Complement (รูปที่ 3.14)
- ในหน้าที่แสดงเฉพาะลำดับนิวคลีโอไทด์ ให้คลิกที่ชื่อของลำดับนิวคลีโอไทด์ (เช่น DNA sequence\_Reverse) แล้วเลือกที่คำสั่ง Sequence > Nucleic acid > Reverse Complement



รูปที่ 3.14 การ Reverse Complement สำหรับลำดับนิวคลีโอไทด์ที่ได้จาก reverse primer

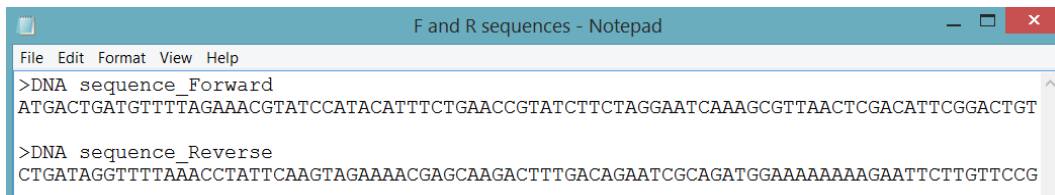
4. ผู้ใช้สามารถเลื่อนดูกราฟลำดับนิวคลีโอไทด์ที่ตำแหน่งต่างๆ เพื่อตรวจสอบเบื้องต้นว่าลำดับนิวคลีโอไทด์ที่กำลังวิเคราะห์นั้น มีบริเวณใดที่ให้ผลถูกต้องน่าเชื่อถือ และมีบริเวณใดบ้างที่เป็น noise signals (บริเวณช่วงต้น และช่วงปลายของลำดับนิวคลีโอไทด์) หรือ unclear signals (double peaks, repeat bases, too low signal) ซึ่งเป็นข้อมูลที่ไม่ถูกต้องหรือมีความน่าเชื่อถือน้อย

5. เมื่อพบบริเวณที่เป็น noise signals และต้องการตัดลำดับนิวคลีโอไทด์ (nucleotide trimming) ในส่วนที่เป็น noise signals ออกไป โดยให้เลือก Mode ในหน้าต่างที่แสดงเฉพาะลำดับนิวคลีโอไทด์เป็น Edit จากนั้นเลือกลำดับนิวคลีโอไทด์เฉพาะส่วนที่เป็น noise signal โดยทำเป็นแถบสีดำ (รูปที่ 3.15 ส่วนล่าง) แล้วกดปุ่ม Del (บนแป้นพิมพ์) เพื่อทำการตัดลำดับนิวคลีโอไทด์บริเวณดังกล่าวออกไป เหลือแต่ผลลำดับนิวคลีโอไทด์ที่มีความถูกต้องน่าเชื่อถือ



รูปที่ 3.15 การเลือกตัดลำดับนิวคลีโอไทด์ ส่วนที่เป็น noise signal โดยใช้โปรแกรม BioEdit

6. จากนั้นสามารถบันทึกผลลำดับนิวคลีโอไทด์โดยคลิกที่คำสั่ง File > Save as จากนั้นสามารถตั้งชื่อ file และเลือกนามสกุลของ file ได้ ซึ่งนิยมเลือกเป็น Fasta file อีกวิธีที่นิยมใช้คือ copy โดยคลิกชื่อของลำดับนิวคลีโอไทด์ (เช่น DNA sequence\_Forward) จากนั้นคลิกที่คำสั่ง Edit > Copy Sequences to clipboard (Fasta Format) จากนั้นเปิดโปรแกรม Notepad แล้วเลือก Edit > Paste แล้วสามารถ save เป็น text file (ที่เป็น Fasta format) เก็บไว้ (รูปที่ 3.16)



รูปที่ 3.16 ตัวอย่างลำดับนิวคลีโอไทด์ที่เป็น Fasta format ในโปรแกรม Notepad



➤ การเชื่อมต่อลำดับนิวคลีโอไทด์ โดยใช้โปรแกรม BioEdit

1. เริ่มจากการเตรียมลำดับนิวคลีโอไทด์ที่ผ่านการตัด (nucleotide trimming) ในส่วนที่เป็น noise signals ออกไปแล้ว โดยรวมลำดับนิวคลีโอไทด์ที่มีทิศทางเดียวกันหลายๆ สายที่ต้องการเชื่อมต่อกันไว้ใน text file (ที่เป็น Fasta format) ของโปรแกรม Notepad (รูปที่ 3.17)



```

>Fragment1
ATGATAACGAATATAAAGTATAAGCTAGCAATAACGTTAAGTGTTCCTTGTAACCTTTTCG
TTAGGATGTAATAGCCAAAATAATGGCTCTAAAAATGATCTTAGTTTTGTTTGCAACAATG
TTTGCTCAGGCTACACAACAACTACAATGAGATCGACAACAGTAAGCGTTAGCTCGTTA
AACGAACTGATAGGTTTTAAACCTATTCAAGTAGAAAATGAGCAAGACTTTGACAGAATC
GCAGATGGAAAAAAGAATTCTTGTTCGGTTCGGTTATGACTTTCCGACCAAGTAATTCT
TATTCGAATGGAAGTGCACAAAATATAACTTCAAATAAATTGTAATGGAAGTTACGTTGCA
CGTTTTAAACCCGAAAGAATTCGTCAGATTCTTGAAACACATGATATAAATTGTGTGGAT
AAATTTGTTTGAATTACGATCAGTATTCGGACTATATTTATTCACAAGAGAACATGCTT
GAAGTAGTCAATCGACTAAACGATCTTGCTCCTTTTTACAACGGTAACAACGATTTGAAC
TTTATACAATTGTTTCAGAATGTTTTGGGCCGGATATTATGTAAAACATTCTCACCCCTCC

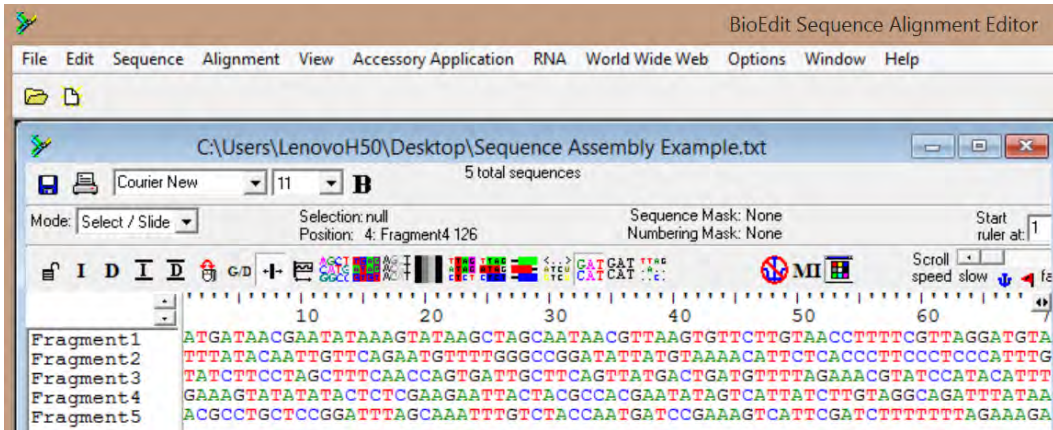
>Fragment2
TTTATACAATTGTTTCAGAATGTTTTGGGCCGGATATTATGTAAAACATTCTCACCCCTCC
CTCCCATTTGATACAAACCAAATCTCTCAGGCTTTAGTTACTCCAATGCAGATTTTTGCA
AGTAGTGCAGACTTCTTAGATGGAACCAACGATGCCGAAAAGTTTTAGAGTTCTTCTTT
ACAGTAGCGGATAGTACAAAATTTGGACATACGATTTATCCACGTATCCTTTCTTTTCTT
GAAGCTACTATAAACGATCCACAAAAGATTAAGAAACAATCTTCTCAGGCCATCGCTTTG
AATGCTGTTTTTAGATTGTTTCAACGTCATATCCATTCAAACAGCAACGAATTCCTTACA
ATGATCGATTATAGATTGATTAGTAAACTGAGAAGACTTGCTCTTGATACAAGTCTCAAC
ACAGATTCTCAAGTCTGGATCATCAACAATGCGATATTTGGTTTGGATAGGATTTATGAA
TATCTTCCTAGCTTTCAACCAGTGATTGCTTCAGTTATGACTGATGTTTTAGAAAACGTAT

>Fragment3
TATCTTCCTAGCTTTCAACCAGTGATTGCTTCAGTTATGACTGATGTTTTAGAAAACGTAT
CCATACATTTCTGAACCGTATCTTCTAGGAATCAAAGCGTTAACTCGACATTCGGACTGT
GCGAATTTACGGATCGGACGAATTTGTTTGAGCGATATCAAAGAACTGTGAAAAAGGCG
GTTTTGTCAAATACATATATTTTACGATAAAAACACAAATCGTACACACAGCTCTATCA
ATCGATGAAATTCAACCTCTATATCACGCTCTTAAACAGGTAGAATCTCAATTTTTCCGT
TTGATAGGAATTCATGCTCCCGTGTCTGGAGATACGACGGATTCAATTACCATGTATGTG
TATAAATCTAGAAAAGGATTACGAAACTTTTCATCCTTTTCTATTCGATCTTTCTACAGAT
AATGGTGGGATCTACATAGAACGAGACAAAACGCTCTATACATATCAACGTAATCCTGCA
GAAAGTATATATACTCTCGAAGAATTACTACGCCACGAATATAGTCATTATCTTGTAGGC

```

รูปที่ 3.17 ตัวอย่างการเตรียมลำดับนิวคลีโอไทด์ใน text file เพื่อใช้ในการ assembly

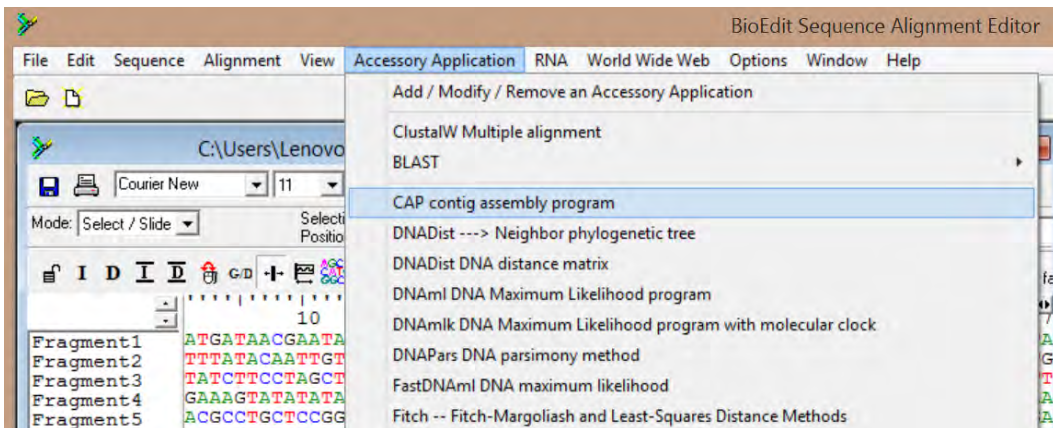
2. เปิดใช้งานโปรแกรม BioEdit แล้วเปิด text file ที่เตรียมไว้ (รูปที่ 3.18)



รูปที่ 3.18 การแสดงผลลำดับนิวคลีโอไทด์จาก text file (Fasta Format) ในโปรแกรม BioEdit

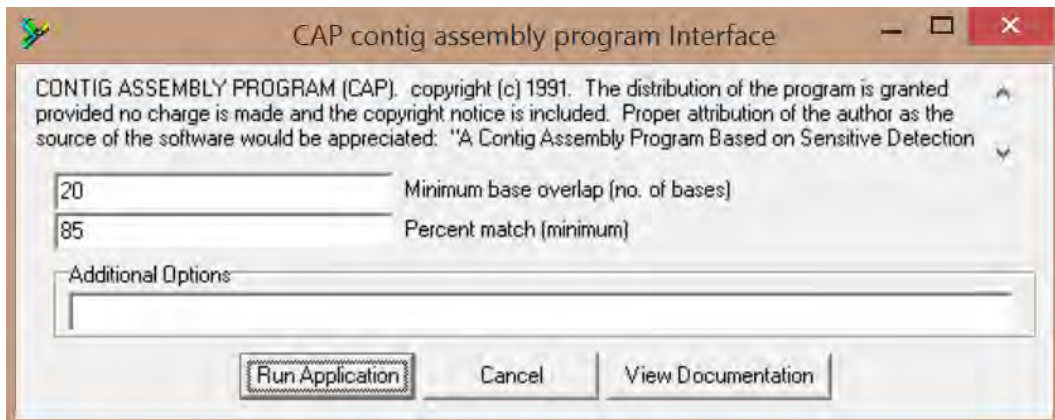
3. คลิกที่คำสั่ง Accessory Application > CAP contig assembly program (รูปที่ 3.19) จะมีหน้าต่างปรากฏขึ้นเพื่อตั้งค่าที่จะใช้ในการเชื่อมต่อลำดับนิวคลีโอไทด์ (รูปที่ 3.20) ดังนี้

- Minimum base overlap กำหนดจำนวนนิวคลีโอไทด์ที่ซ้อนทับกันน้อยที่สุดระหว่างสายของนิวคลีโอไทด์ที่จะเชื่อมต่อกัน (ค่า default = 20)
- Percent match (minimum) กำหนดค่าเปอร์เซ็นต์ความเหมือนที่น้อยที่สุดระหว่างลำดับนิวคลีโอไทด์ที่จะเชื่อมต่อกัน (ค่า default = 85)



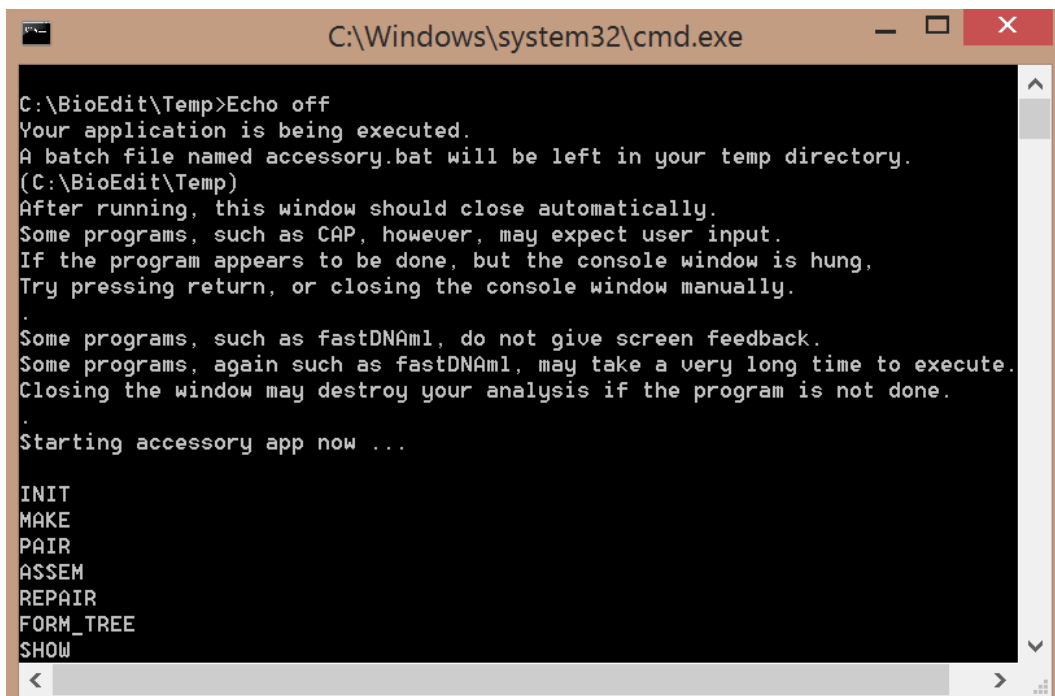
รูปที่ 3.19 การใช้คำสั่ง CAP contig assembly program



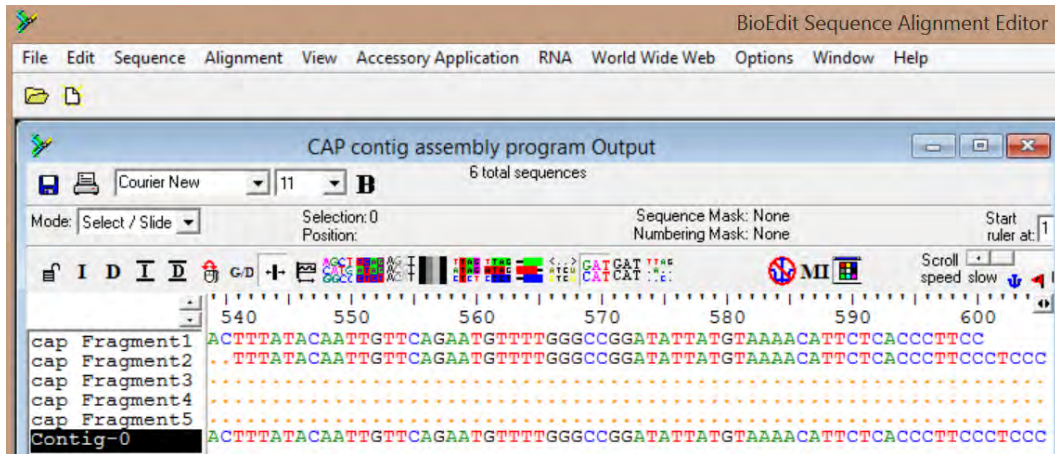


รูปที่ 3.20 การตั้งค่าสำหรับการเชื่อมต่อลำดับนิวคลีโอไทด์ด้วย CAP contig assembly

4. เมื่อตั้งค่าเสร็จแล้ว (นิยมใช้ค่า default) ให้คลิกที่คำว่า Run Application จะปรากฏหน้าต่างเป็น command อัตโนมัติ (รูปที่ 3.21) เมื่อพบคำว่า SHOW ที่บรรทัดล่างสุด แสดงว่าการวิเคราะห์เสร็จแล้ว ให้กดปุ่ม Enter (บนแป้นพิมพ์) โปรแกรมก็จะแสดงผล (รูปที่ 3.22)

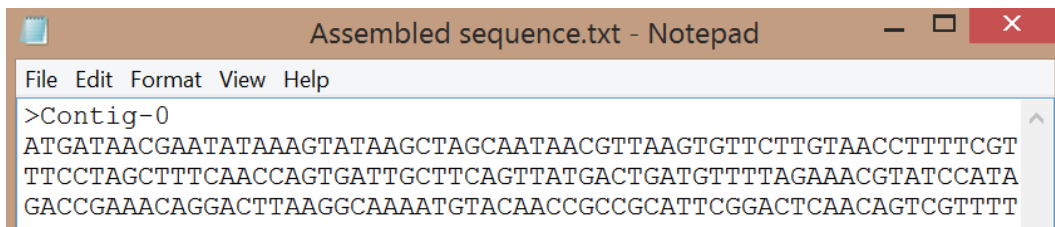


รูปที่ 3.21 การทำงานของ CAP contig assembly program เพื่อเชื่อมต่อลำดับนิวคลีโอไทด์



รูปที่ 3.22 ผลการเชื่อมต่อลำดับนิวคลีโอไทด์ด้วย CAP contig assembly program

5. จากรูปที่ 3.22 โปรแกรมจะแสดงลำดับนิวคลีโอไทด์ที่ซ้อนทับกัน (overlapping sequences) สำหรับใช้เชื่อมต่อกันระหว่างสาย หากต้องการข้อมูลลำดับนิวคลีโอไทด์ที่ผ่านการเชื่อมต่อกันเสร็จแล้ว (full length gene) ให้คลิกที่ชื่อของลำดับนิวคลีโอไทด์ (Contig-0) จากนั้นคลิกที่คำสั่ง Edit > Copy Sequences to clipboard (Fasta Format) จากนั้นเปิดโปรแกรม Notepad แล้วเลือก Edit > Paste แล้วสามารถ save เป็น text file (ที่เป็น Fasta Format) เก็บไว้ (รูปที่ 3.23) โดยอาจเปลี่ยนชื่อจาก Contig-0 เป็นชื่อของยีนตามความเหมาะสม

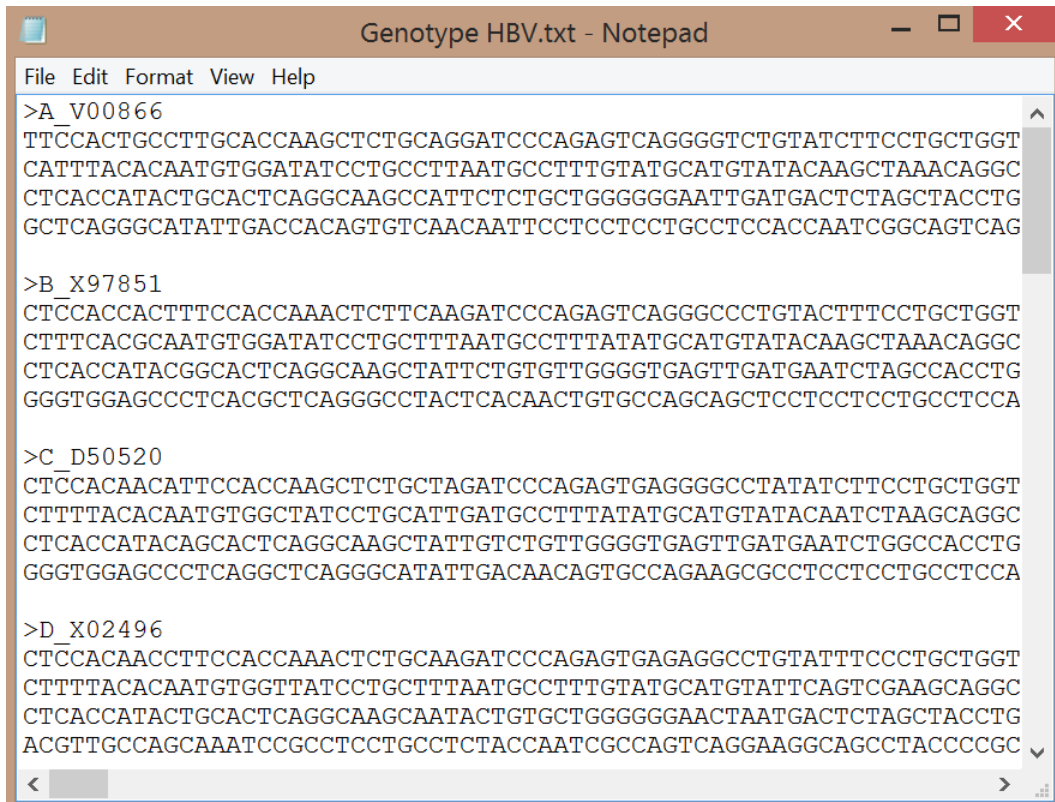


รูปที่ 3.23 ตัวอย่างลำดับนิวคลีโอไทด์ที่ผ่านการเชื่อมต่อเสร็จแล้ว ในโปรแกรม Notepad

6. ในกรณีที่ลำดับนิวคลีโอไทด์ไม่สามารถเชื่อมต่อกันได้ เช่น ไม่มีลำดับนิวคลีโอไทด์ที่ซ้อนทับกัน (overlapping sequences) หรือส่วนที่ซ้อนทับกันนั้นมีจำนวนนิวคลีโอไทด์น้อยกว่าค่า Minimum base overlap ที่กำหนดไว้ โปรแกรม CAP contig assembly จะแสดงผลการเชื่อมต่อเป็นหลาย contigs เช่น Contig-0, Contig-1, Contig-2 เป็นต้น

### ➤ การจัดเรียงลำดับนิวคลีโอไทด์ โดยใช้โปรแกรม BioEdit

1. เริ่มจากการเตรียมลำดับนิวคลีโอไทด์หลายๆ สายที่ต้องการจัดเรียงร่วมกันไว้ใน text file (Fasta format) ของโปรแกรม Notepad (รูปที่ 3.24)



```

File Edit Format View Help
>A_V00866
TTCCACTGCCTTGCACCAAGCTCTGCAGGATCCCAGAGTCAGGGGTCTGTATCTTCCTGCTGGT
CATTTACACAATGTGGATATCCTGCCTTAATGCCTTTGTATGCATGTATAACAAGCTAAACAGGC
CTCACCATACTGCACTCAGGCAAGCCATTCTCTGCTGGGGGAATTGATGACTCTAGCTACCTG
GCTCAGGGCATATTGACCACAGTGTCAACAATTCCTCCTCCTGCCTCCACCAATCGGCAGTCAG

>B_X97851
CTCCACCACCTTCCACCAAACCTCTTCAAGATCCCAGAGTCAGGGCCCTGTACTTTCCTGCTGGT
CTTTACGCAATGTGGATATCCTGCTTTAATGCCTTTATATGCATGTATAACAAGCTAAACAGGC
CTCACCATAAGCACTCAGGCAAGCTATTCTGTGTTGGGGTGAGTTGATGAATCTAGCCACCTG
GGGTGGAGCCCTCAGGCTCAGGGCCTACTCACAACCTGTGCCAGCAGCTCCTCCTCCTGCCTCCA

>C_D50520
CTCCACAACATTCCACCAAGCTCTGCTAGATCCCAGAGTGAGGGGCCTATATCTTCCTGCTGGT
CTTTTACACAATGTGGCTATCCTGCATTGATGCCTTTATATGCATGTATAACAATCTAAGCAGGC
CTCACCATAAGCACTCAGGCAAGCTATTGTCTGTTGGGGTGAGTTGATGAATCTGGCCACCTG
GGGTGGAGCCCTCAGGCTCAGGGCATATTGACAACAGTGCCAGAAGCGCCTCCTCCTGCCTCCA

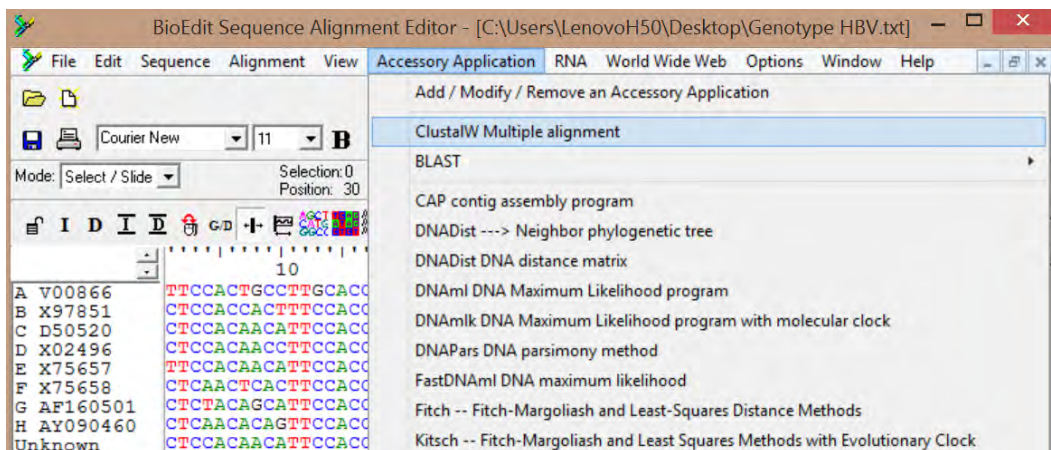
>D_X02496
CTCCACAACCTTCCACCAAACCTCTGCAAGATCCCAGAGTGAGAGGCCTGTATTTCCCTGCTGGT
CTTTTACACAATGTGGTTATCCTGCTTTAATGCCTTTGTATGCATGTATTCAGTCGAAGCAGGC
CTCACCATACTGCACTCAGGCAAGCAATACTGTGCTGGGGGAACCTAATGACTCTAGCTACCTG
ACGTTGCCAGCAAATCCGCCTCCTGCCTCTACCAATCGCCAGTCAGGAAGGCAGCCTACCCCGC

```

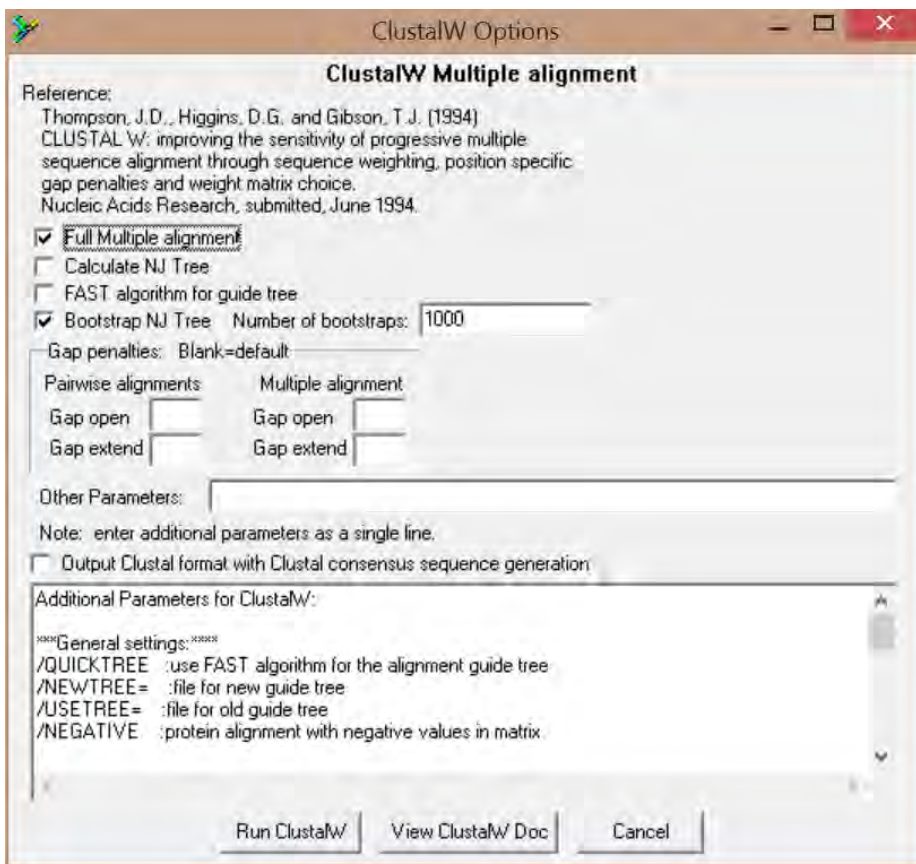
รูปที่ 3.24 ตัวอย่างการเตรียมลำดับนิวคลีโอไทด์ใน text file เพื่อใช้ในการ alignment

2. เปิดใช้งานโปรแกรม BioEdit แล้วเปิด text file ที่เตรียมไว้
3. คลิกที่คำสั่ง Accessory Application > ClustalW Multiple alignment (รูปที่ 3.25) จะมีหน้าต่างปรากฏขึ้นเพื่อให้ตั้งค่าที่จะใช้ในการจัดเรียงลำดับนิวคลีโอไทด์ (รูปที่ 3.26)
4. เมื่อตั้งค่าเสร็จแล้ว (นิยมใช้ค่า default) ให้คลิกที่คำว่า Run ClustalW จากนั้นกด OK จะปรากฏหน้าต่างการทำงานของโปรแกรม แล้วก็แสดงผลขึ้นมาอัตโนมัติ (รูปที่ 3.27)

การวิเคราะห์ข้อมูลลำดับนิวคลีโอไทด์

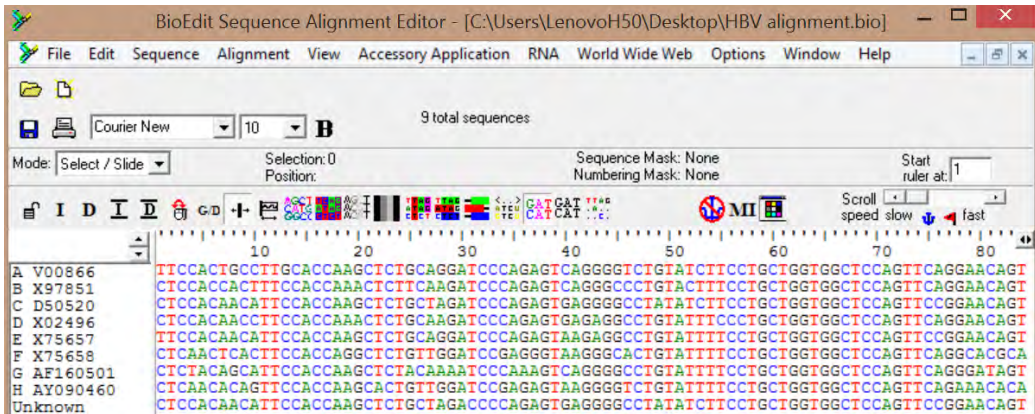


รูปที่ 3.25 การใช้คำสั่ง ClustalW Multiple alignment



รูปที่ 3.26 การตั้งค่าสำหรับจัดเรียงลำดับนิวคลีโอไทด์ด้วย ClustalW Multiple alignment

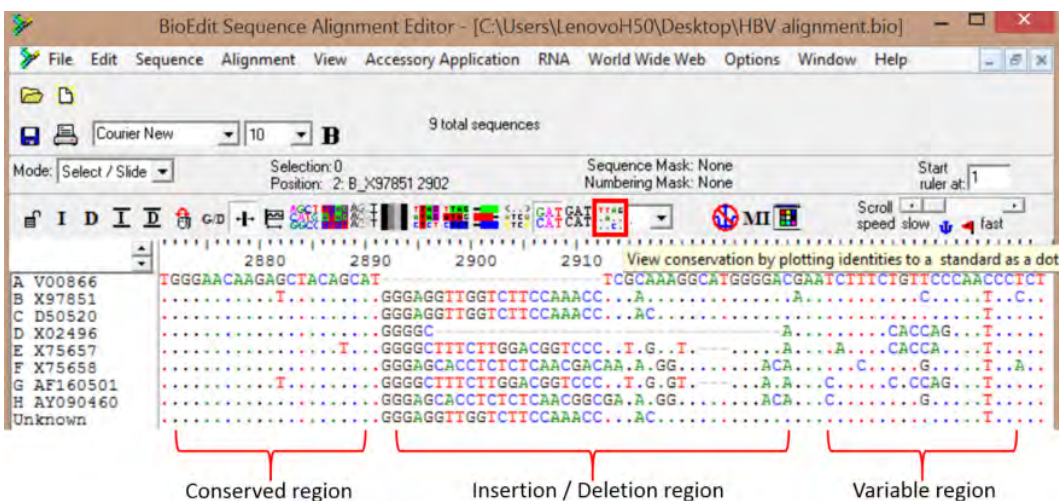




รูปที่ 3.27 ผลการจัดเรียงลำดับนิวคลีโอไทด์ด้วย ClustalW Multiple alignment

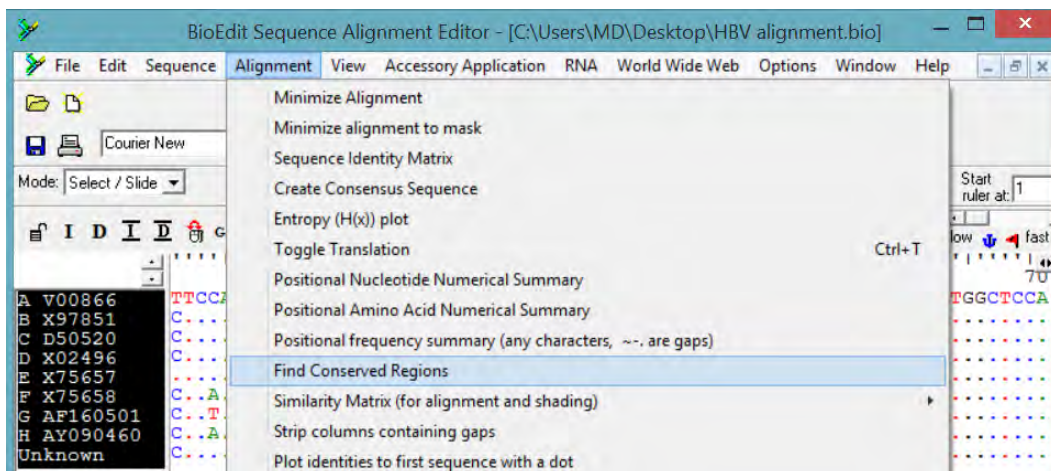
5. จากรูปที่ 3.27 แสดงลำดับนิวคลีโอไทด์ที่จัดเรียงเสร็จแล้ว (aligned sequences) หากต้องการจำแนกส่วนที่เหมือนกันและแตกต่างกันอย่างชัดเจน ให้คลิกที่ปุ่ม “View conservation by plotting identities to a standard as a dot” (รูปที่ 3.28) โปรแกรมจะใช้ลำดับนิวคลีโอไทด์สายแรก ในการเปรียบเทียบกับลำดับนิวคลีโอไทด์สายอื่นๆ โดยแสดงผล ดังนี้

- บริเวณอนุรักษ์ (conserved region) จะแสดงเป็นเครื่องหมาย (.)
- บริเวณแปรเปลี่ยน (variable region) จะแสดงเป็นเบสที่แตกต่างกัน (A/T/C/G)
- บริเวณที่ขาดหายไป (deletion region) จะแสดงเป็นเครื่องหมาย (-)

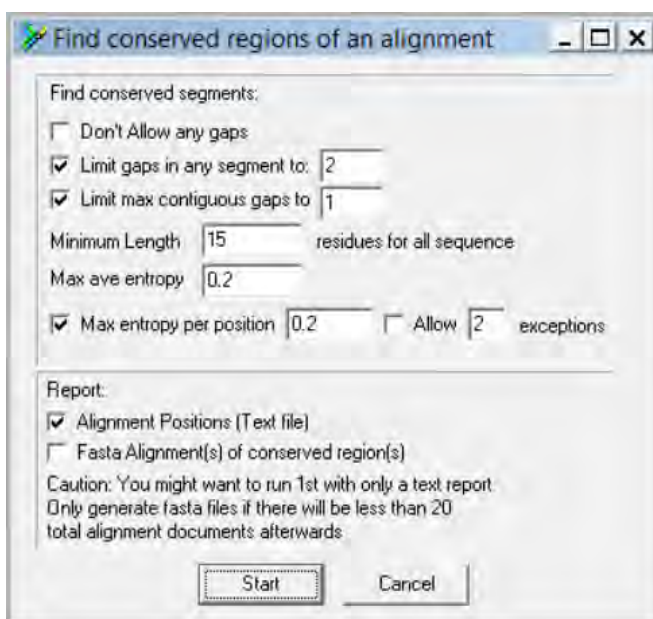


รูปที่ 3.28 การจำแนก conserved / variable / insertion / deletion regions

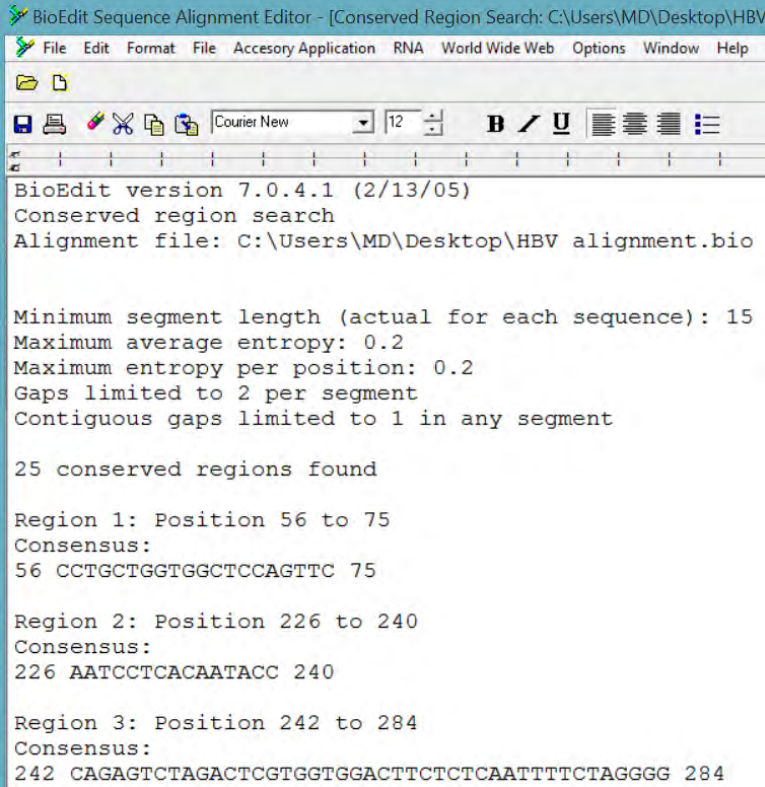
6. หากต้องการสั่งให้โปรแกรมค้นหาบริเวณอนุรักษ์ (conserved regions) แบบอัตโนมัติ สามารถคลิกที่คำสั่ง Alignment > Find Conserved Regions (รูปที่ 3.29) จะมีหน้าต่างปรากฏขึ้นเพื่อให้ตั้งค่าที่จะใช้ในการค้นหาบริเวณอนุรักษ์ (รูปที่ 3.30) จากนั้นคลิกที่คำว่า Start โปรแกรมก็จะแสดงผลการค้นหาบริเวณอนุรักษ์ (รูปที่ 3.31)



รูปที่ 3.29 การค้นหาบริเวณอนุรักษ์ โดยใช้คำสั่ง Find Conserved Regions



รูปที่ 3.30 การตั้งค่าสำหรับการค้นหาบริเวณอนุรักษ์ โดยใช้คำสั่ง Find Conserved Regions



```

BioEdit Sequence Alignment Editor - [Conserved Region Search: C:\Users\MD\Desktop\HBV
File Edit Format File Accessory Application RNA World Wide Web Options Window Help
Courier New 12 B / U
BioEdit version 7.0.4.1 (2/13/05)
Conserved region search
Alignment file: C:\Users\MD\Desktop\HBV alignment.bio

Minimum segment length (actual for each sequence): 15
Maximum average entropy: 0.2
Maximum entropy per position: 0.2
Gaps limited to 2 per segment
Contiguous gaps limited to 1 in any segment

25 conserved regions found

Region 1: Position 56 to 75
Consensus:
56 CCTGCTGGTGGCTCCAGTTC 75

Region 2: Position 226 to 240
Consensus:
226 AATCCTCACAATACC 240

Region 3: Position 242 to 284
Consensus:
242 CAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTTCTAGGGG 284

```

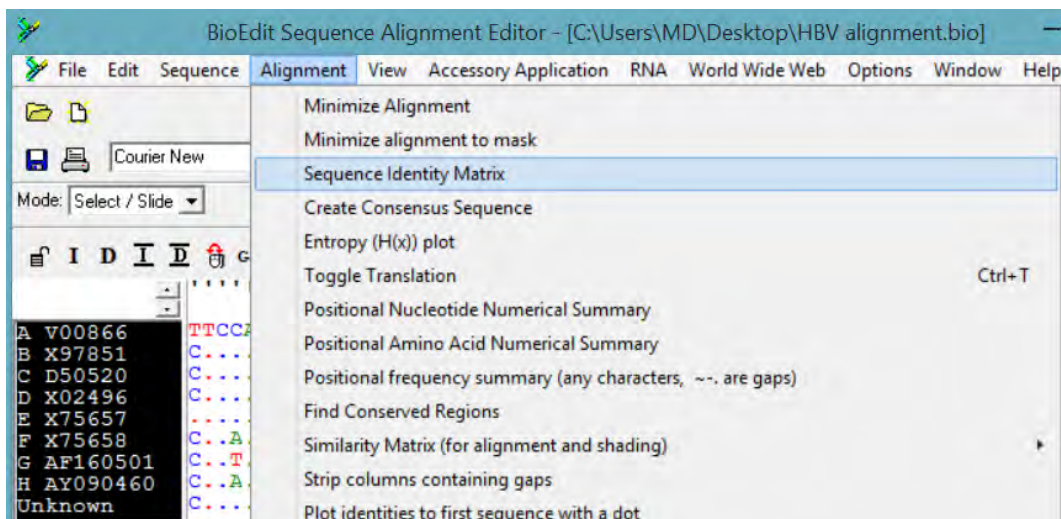
**รูปที่ 3.31** ตัวอย่างผลการค้นหาบริเวณอนุรักษ์ โดยใช้คำสั่ง Find Conserved Regions

7. จากรูปที่ 3.31 โปรแกรมจะสรุปจำนวน conserved regions ที่พบทั้งหมด พร้อมทั้งแสดงตำแหน่งและลำดับนิวคลีโอไทด์ที่พบในแต่ละบริเวณอนุรักษ์ ซึ่งสามารถนำไปประยุกต์ใช้ในการออกแบบ primers และ probes ได้ ซึ่งจะกล่าวรายละเอียดในบทต่อไป

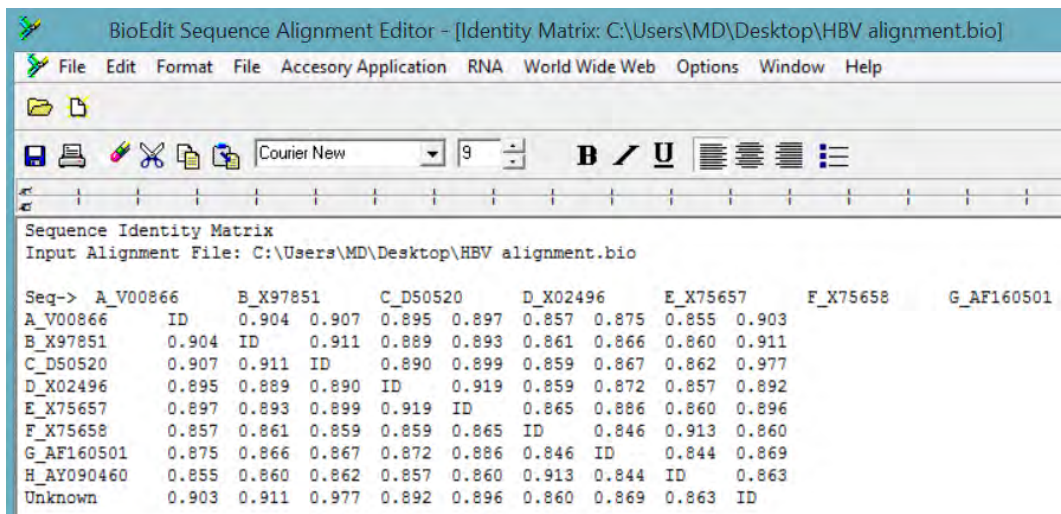
8. หากต้องการสั่งให้โปรแกรมคำนวณหาค่าความเหมือน (identity score) เปรียบเทียบกันระหว่างลำดับนิวคลีโอไทด์หลายๆ สาย สามารถคลิกที่คำสั่ง Alignment > Sequence Identity Matrix (รูปที่ 3.32) จะมีหน้าต่างปรากฏขึ้นเพื่อให้ตั้งชื่อ file ที่ต้องการบันทึก จากนั้นคลิกที่คำสั่ง Save โปรแกรมก็จะแสดงผลค่าความเหมือน (identity score) (รูปที่ 3.33) ซึ่งสามารถ copy ข้อมูลดังกล่าวแล้วไป paste ในโปรแกรม MS Excel เพื่อให้พิจารณาและวิเคราะห์ข้อมูลได้ง่ายขึ้น นอกจากนี้หากต้องการข้อมูลเป็นค่าเปอร์เซ็นต์ความเหมือน (% identity) ก็สามารถคูณตัวเลขดังกล่าวด้วยค่า 100 ก็จะได้เป็นค่า % identity (รูปที่ 3.34)



การวิเคราะห์ข้อมูลลำดับนิวคลีโอไทด์



รูปที่ 3.32 การคำนวณหาค่าความเหมือน โดยใช้คำสั่ง Sequence Identity Matrix



รูปที่ 3.33 ตัวอย่างการแสดงผลค่าความเหมือน (identity score) ในโปรแกรม BioEdit

	A	B	C	D	E	F	G	H	I	J
1	Sequence Identity Matrix									
2	Input Alignment File: C:\Users\MD\Desktop\HBV alignment.bio									
3										
4	Seq->	A_V00866	B_X97851	C_D50520	D_X02496	E_X75657	F_X75658	G_AF160501	H_AY090460	Unknown
5	A_V00866	ID	90.40%	90.70%	89.50%	89.70%	85.70%	87.50%	85.50%	90.30%
6	B_X97851	90.40%	ID	91.10%	88.90%	89.30%	86.10%	86.60%	86.00%	91.10%
7	C_D50520	90.70%	91.10%	ID	89.00%	89.90%	85.90%	86.70%	86.20%	97.70%
8	D_X02496	89.50%	88.90%	89.00%	ID	91.90%	85.90%	87.20%	85.70%	89.20%
9	E_X75657	89.70%	89.30%	89.90%	91.90%	ID	86.50%	88.60%	86.00%	89.60%
10	F_X75658	85.70%	86.10%	85.90%	85.90%	86.50%	ID	84.60%	91.30%	86.00%
11	G_AF160501	87.50%	86.60%	86.70%	87.20%	88.60%	84.60%	ID	84.40%	86.90%
12	H_AY090460	85.50%	86.00%	86.20%	85.70%	86.00%	91.30%	84.40%	ID	86.30%
13	Unknown	90.30%	91.10%	97.70%	89.20%	89.60%	86.00%	86.90%	86.30%	ID

รูปที่ 3.34 ตัวอย่างการแสดงผลเปอร์เซ็นต์ความเหมือน (% identity) ในโปรแกรม MS Excel

➤ การวิเคราะห์การกลายพันธุ์โดยใช้โปรแกรม BioEdit

1. เริ่มจากการเตรียมลำดับนิวคลีโอไทด์หลายๆ สายที่ต้องการจัดเรียงร่วมกันไว้ใน text file (Fasta format) ของโปรแกรม Notepad โดยให้ลำดับนิวคลีโอไทด์สายแรก เป็น Reference sequence (Ref Seq) หรือเป็น wild type sequence เพื่อใช้เป็นลำดับนิวคลีโอไทด์ตั้งต้นในการเปรียบเทียบกับลำดับนิวคลีโอไทด์สายอื่นๆ (รูปที่ 3.35)

```

Mutation analysis - Notepad
File Edit Format View Help
>Ref_seq
AAATTGAGCCCGCAGCCTCCCGCTTCGCTCTCTGCTCCTCCTGTTTCGACAGTTCAGCCGCATCTTCTTTGCGTCGCCAGCCGAG
AGCTCATTTCTCTGGTATGACAACGAATTTGGCTACAGCAACAGGGTGGTGGACCTCATGGCCACATGGCCTCCAAGGAGTAAG

>Unknown_1
AAATTGAGCCCGCAGCCTCCCGCTTCGCTCTCTGCTCCTCCTGTTTCGACAGTTCAGCCGCATCTTCTTTGCGTCGCCAGCCGAG
AGCTCATTTCTCTGGTATGACAACGAATTTGGCTACAGCAACAGGGTGGTGGACCTCATGGCCACATGGCCTCCAAGGAGTAAG

>Unknown_2
AAATTGAGCCCGCAGCCTCCCGCTTCGCTCTCTGCTCCTCCTGTTTCGACAGTTCAGCCGCATCTTCTTTGCGTCGCCAGCCGAG
AGCTCATTTCTCTGGTATGACAACGAATTTGGCTACAGCAACAGGGTGGTGGACCTCATGGCCACATGGCCTCCAAGGAGTAAG

>Unknown_3
AAATTGAGCCCGCAGCCTCCCGCTTCGCTCTCTGCTCCTCCTGTTTCGACAGTTCAGCCGCATCTTCTTTGCGTCGCCAGCCGAG
AGCTCATTTCTCTGGTATGACAACGAATTTGGCTACAGCAACAGGGTGGTGGACCTCATGGCCACATGGCCTCCAAGGAGTAAG

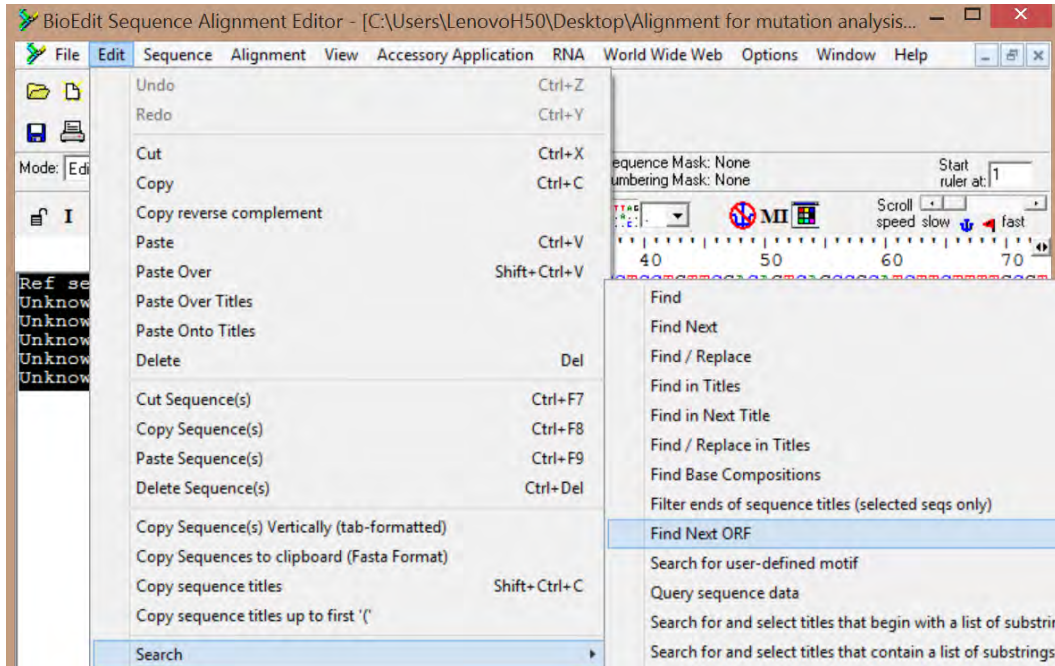
>Unknown_4
AAATTGAGCCCGCAGCCTCCCGCTTCGCTCTCTGCTCCTCCTGTTTCGACAGTTCAGCCGCATCTTCTTTGCGTCGCCAGCCGAG
AGCTCATTTCTCTGGTATGACAACGAATTTGGCTACAGCAACAGGGTGGTGGACCTCATGGCCACATGGCCTCCAAGGAGTAAG

>Unknown_5
AAATTGAGCCCGCAGCCTCCCGCTTCGCTCTCTGCTCCTCCTGTTTCGACAGTTCAGCCGCATCTTCTTTGCGTCGCCAGCCGAG
TTTCTCTGGTATGACAACGAATTTGGCTACAGCAACAGGGTGGTGGACCTCATGGCCACATGGCCTCCAAGGAGTAAGACCCCT
    
```

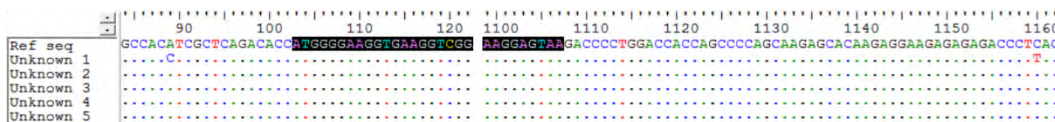
รูปที่ 3.35 ตัวอย่างการเตรียมลำดับนิวคลีโอไทด์ใน text file เพื่อใช้วิเคราะห์การกลายพันธุ์

2. เปิดใช้งานโปรแกรม BioEdit แล้วเปิด text file ที่เตรียมไว้
3. คลิกที่คำสั่ง Accessory Application > ClustalW Multiple alignment จะมีหน้าต่างปรากฏขึ้นเพื่อให้ตั้งค่าที่จะใช้ในการจัดเรียงลำดับนิวคลีโอไทด์ เมื่อตั้งค่าเสร็จแล้ว (นิยมใช้ค่า default) ให้คลิกที่คำว่า Run ClustalW จากนั้นกด OK จะปรากฏหน้าต่างการทำงานของโปรแกรม แล้วก็จะแสดงผลลำดับนิวคลีโอไทด์ที่จัดเรียงเสร็จแล้ว (aligned sequences) หากต้องการจำแนกส่วนที่เหมือนกันและแตกต่างกันอย่างชัดเจน ให้คลิกที่ปุ่ม “View conservation by plotting identities to a standard as a dot” โปรแกรมจะใช้ลำดับนิวคลีโอไทด์สายแรก (Ref Seq) เป็นตัวตั้งต้นในการเปรียบเทียบการกลายพันธุ์ในลำดับนิวคลีโอไทด์สายอื่นๆ
4. ในการวิเคราะห์การกลายพันธุ์ของยีนควรพิจารณาว่าตำแหน่งที่เกิดการกลายพันธุ์อยู่ในบริเวณส่วนใดของยีน ได้แก่
  - Coding sequence (CDS) เป็นลำดับนิวคลีโอไทด์ส่วนที่จะใช้ในการแปลรหัสเพื่อสร้างเป็นกรดอะมิโน ซึ่งค้นหาจาก open reading frame (ORF) เริ่มจาก start codon ถึง stop codon หากพบการกลายพันธุ์ในส่วนนี้ควรพิจารณาต่อว่าเป็นการกลายพันธุ์แบบใด
    - Synonymous (silent) mutation
    - Nonsynonymous mutation
      - Missense mutation
      - Nonsense mutation
  - Untranslated regions (UTR) เป็น non-coding sequence หรือลำดับนิวคลีโอไทด์ส่วนที่ไม่ได้ใช้ในการแปลรหัสเพื่อสร้างเป็นกรดอะมิโน โดยแบ่งเป็น
    - 5'-UTR เป็นบริเวณที่อยู่ก่อนหน้า start codon
    - 3'-UTR เป็นบริเวณที่อยู่หลังจาก stop codon
  - หากพบ insertion หรือ deletion ควรพิจารณาต่อว่าเกิด frameshift หรือไม่

5. หากต้องการให้โปรแกรมค้นหาส่วนที่เป็น open reading frame (ORF) ให้คลิกที่ชื่อของนิวคลีโอไทด์ที่ต้องการค้นหา จากนั้นใช้คำสั่ง Edit > Search > Find Next ORF (รูปที่ 3.36) จากนั้นโปรแกรมจะแสดงผลเป็นแถบสีดำบริเวณ ORF (start → stop codons) (รูปที่ 3.37)



รูปที่ 3.36 การค้นหา open reading frame โดยใช้คำสั่ง Find Next ORF

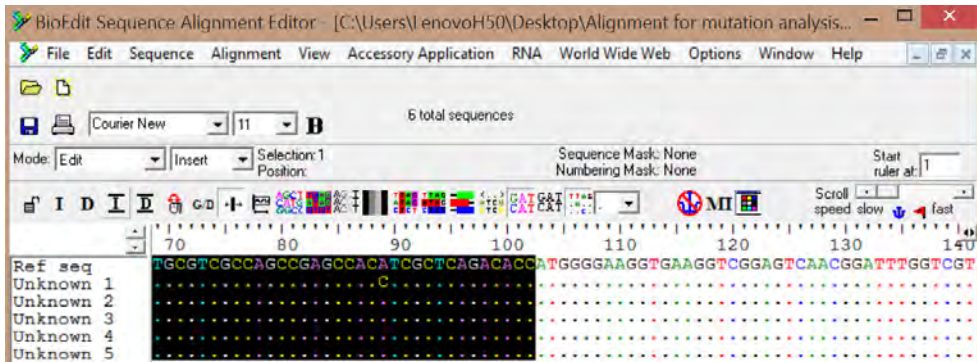


รูปที่ 3.37 ตัวอย่างการแสดงผล open reading frame ในโปรแกรม BioEdit

6. จากรูปที่ 3.37 จะพบว่าบริเวณที่เป็น ORF หรือ CDS จะเริ่มตั้งแต่ตำแหน่ง 103-1107 แสดงว่า 5'-UTR จะเริ่มตั้งแต่ตำแหน่งที่ 1-102 และ 3'-UTR จะเริ่มตั้งแต่ตำแหน่ง 1108 เป็นต้นไป นอกจากนี้พบว่าตัวอย่าง Unknown 1 มีการกลายพันธุ์ที่บริเวณ 5'-UTR ที่ตำแหน่ง 89 โดยเปลี่ยนแบบ transversion จาก A ไปเป็น C (A89C) และ 3'-UTR ที่ตำแหน่ง 1160 โดยเปลี่ยนแบบ transition จาก C ไปเป็น T (C1160T) เป็นต้น ส่วนตัวอย่างอื่นๆ (Unknown 2-5) ไม่พบการกลายพันธุ์ในบริเวณ UTR

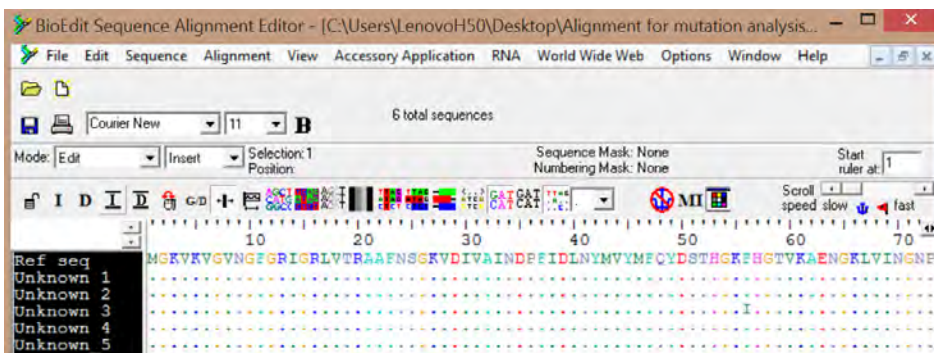


7. หากต้องการพิจารณาเฉพาะส่วนที่เป็น CDS และต้องการตัดลำดับนิวคลีโอไทด์ (nucleotide trimming) ในส่วนที่เป็น UTR ออกไป ให้เลือก Mode เป็น Edit จากนั้นเลือก ลำดับนิวคลีโอไทด์เฉพาะส่วนที่เป็น UTR โดยทำเป็นแถบสีดำ (รูปที่ 3.38) แล้วกดปุ่ม Del (บนแป้นพิมพ์) เพื่อทำการตัดลำดับนิวคลีโอไทด์บริเวณดังกล่าวออกไป เหลือแต่ลำดับนิวคลีโอไทด์ ส่วนที่เป็น CDS เท่านั้น



รูปที่ 3.38 การเลือกและตัดลำดับนิวคลีโอไทด์ ส่วนที่เป็น UTR โดยใช้โปรแกรม BioEdit

8. จากรูปที่ 3.38 เมื่อตัดส่วนที่เป็น UTR ออกไปแล้ว ลำดับนิวคลีโอไทด์จะเริ่มต้นที่ start codon (ATG) ซึ่งเป็นตำแหน่งเริ่มต้นของกระบวนการแปลรหัส หากต้องการสั่งให้โปรแกรมทำการแปลรหัสให้กลายเป็นกรดอะมิโน ให้คลิกที่ชื่อของลำดับนิวคลีโอไทด์ที่ต้องการ แล้วกด Ctrl + G (บนแป้นพิมพ์) โปรแกรมจะแสดงผลการแปลรหัสเป็นลำดับกรดอะมิโน (รูปที่ 3.39) หากต้องการเปลี่ยนกลับไปเป็นลำดับนิวคลีโอไทด์ สามารถทำได้โดยการกด Ctrl + G (บนแป้นพิมพ์) อีกครั้ง



รูปที่ 3.39 การแปลรหัสจากลำดับนิวคลีโอไทด์ ไปเป็นลำดับกรดอะมิโน ด้วย BioEdit

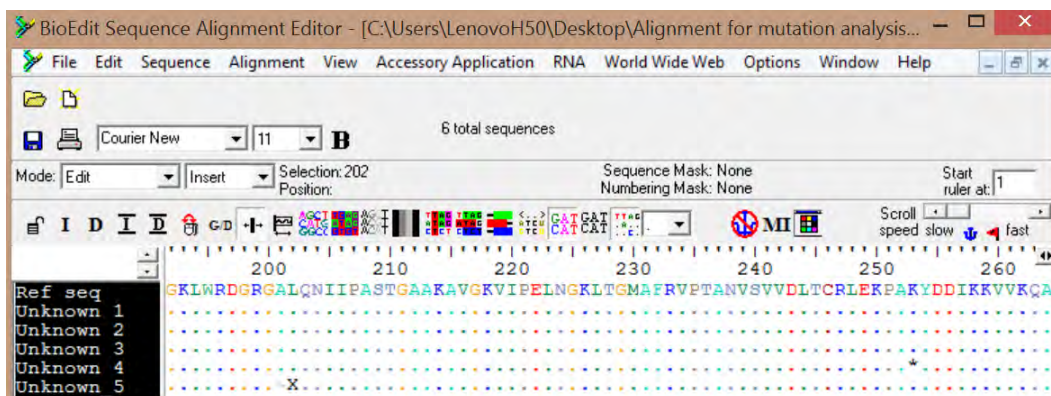
9. ในกรณีที่ต้องการตรวจสอบว่าการกลายพันธุ์ที่พบในลำดับนิวคลีโอไทด์ จะทำให้เกิดการเปลี่ยนแปลงกรดอะมิโนหรือไม่ สามารถทำได้โดยการทำแถบคำในตำแหน่งที่พบการกลายพันธุ์บนลำดับนิวคลีโอไทด์ จากนั้นกด Ctrl + G (บนแป้นพิมพ์) เพื่อเปลี่ยนเป็นลำดับกรดอะมิโน ซึ่งแถบคำจะยังคงปรากฏอยู่ ซึ่งแปลผลได้ ดังนี้

- หากกรดอะมิโนที่ตำแหน่งดังกล่าว มีความเหมือนกับ Ref Seq แสดงว่าเป็น synonymous (silent) mutation (รูปที่ 3.40) ตัวอย่างเช่น Unknown 2 พบการกลายพันธุ์ C108A แต่เมื่อแปลรหัสเป็นกรดอะมิโนก็ยังคงเป็น P ที่ตำแหน่ง 36 เช่นเดียวกับ Ref Seq
- หากกรดอะมิโนที่ตำแหน่งดังกล่าว มีความแตกต่างจาก Ref Seq แสดงว่าเป็น nonsynonymous mutation ชนิด missense mutation (รูปที่ 3.40) ตัวอย่างเช่น Unknown 3 พบการกลายพันธุ์ T166A ทำให้เปลี่ยนกรดอะมิโน จาก Phenylalanine ที่ตำแหน่ง 56 ไปเป็น Isoleucine (F56I)



รูปที่ 3.40 ตัวอย่างการจำแนกระหว่าง synonymous และ nonsynonymous mutations

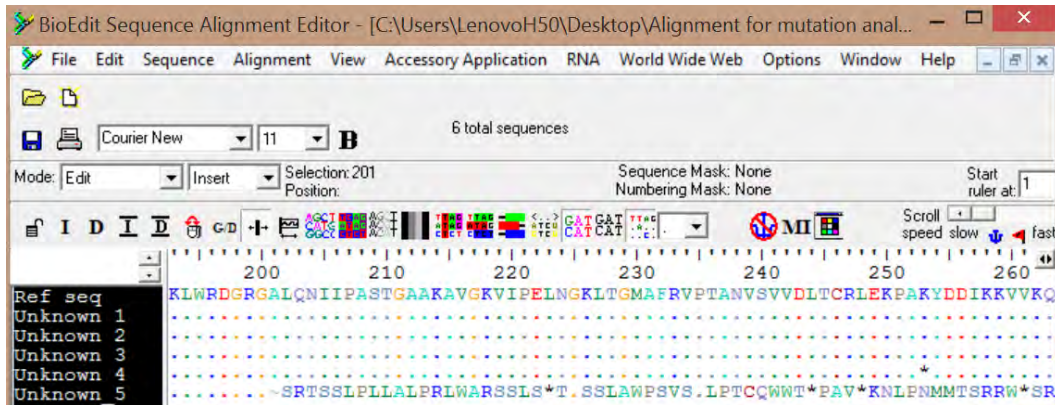
- หากกรดอะมิโนที่ตำแหน่งดังกล่าว กลายเป็นสัญลักษณ์ \* (stop codon) แสดงว่าเป็น nonsynonymous mutation ชนิด nonsense mutation (รูปที่ 3.41) ตัวอย่างเช่น Unknown 4 พบการกลายพันธุ์ ทำให้กลายเป็น stop codon ที่ตำแหน่ง 253 ของลำดับกรดอะมิโน ส่งผลให้ได้โปรตีนที่สั้นกว่าปกติ (truncated protein)
- หากพบสัญลักษณ์ ~ แสดงว่าเกิดการ deletion แบบ inframe คือขาดหายไป (จำนวนนิวคลีโอไทด์ที่หารด้วย 3 ลงตัว) (รูปที่ 3.41)
- หากพบสัญลักษณ์ X แสดงว่าเกิดการ deletion ที่ทำให้เกิด frameshift mutation (จำนวนนิวคลีโอไทด์ที่หารด้วย 3 ไม่ลงตัว) (รูปที่ 3.41)



รูปที่ 3.41 ตัวอย่างการพบ nonsense / inframe / frameshift mutations

10. หากต้องการแสดงการเกิด frameshift mutation ให้ชัดเจนมากยิ่งขึ้น ควรทำการแปลรหัสจากลำดับนิวคลีโอไทด์ไปเป็นลำดับกรดอะมิโน (กด Ctrl + G) ก่อน แล้วจึงทำการจัดเรียงลำดับกรดอะมิโน (amino acid sequences alignment) ด้วยคำสั่ง Accessory Application > ClustalW Multiple alignment โปรแกรมก็จะแสดงผลการเกิด frameshift mutation (รูปที่ 3.42) ตัวอย่างเช่น Unknown 5 พบการกลายพันธุ์แบบ deletion ที่ทำให้เกิด frameshift mutation ตั้งแต่ตำแหน่งที่ 202 จะได้รูปแบบลำดับกรดอะมิโนเปลี่ยนแปลงไป และเกิดเป็น stop codon ที่ตำแหน่ง 223 เป็นต้น





รูปที่ 3.42 ตัวอย่างการแสดงผล frameshift mutation ด้วยโปรแกรม BioEdit

### เอกสารอ้างอิง และแนะนำอ่านเพิ่มเติม

1. F. Sanger, S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74: 5463-5467.
2. Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41:95-98.
3. T. Hunkapiller, R.J. Kaiser, B.F. Koop, and L. Hood. 1991. Large-scale and automated DNA sequence determination. Science 254: 59-67.



# 4

เครื่องมือในการค้นหาเปรียบเทียบความคล้ายคลึงกันทางชีวภาพ

Basic Local Alignment Search Tool (BLAST)

## เครื่องมือในการค้นหาเปรียบเทียบความคล้ายคลึงกันทางชีวภาพ

### Basic Local Alignment Search Tool (BLAST)

BLAST (Basic Local Alignment Search Tool) เป็นเครื่องมือที่ใช้ในการค้นหาส่วนของยีนหรือโปรตีนที่มีความคล้ายคลึงกัน โดยโปรแกรมจะเปรียบเทียบลำดับนิวคลีโอไทด์หรือกรดอะมิโนกับฐานข้อมูลและคำนวณนัยสำคัญทางสถิติของความเหมือน จึงสามารถประยุกต์ใช้ในการระบุชนิดของยีน โปรตีน หรือจำแนกสายพันธุ์ของสิ่งมีชีวิตได้ นอกจากนี้สามารถอนุมานความสัมพันธ์ระหว่างบทบาทหน้าที่และวิวัฒนาการของยีนในสิ่งมีชีวิตได้ การใช้ BLAST สามารถเข้าได้จาก <https://blast.ncbi.nlm.nih.gov/Blast.cgi> (รูปที่ 4.1) โดยจำแนกได้ ดังนี้

- Nucleotide BLAST การใช้ลำดับนิวคลีโอไทด์ที่สนใจเปรียบเทียบกับลำดับนิวคลีโอไทด์ในฐานข้อมูล
- Protein BLAST การใช้ลำดับกรดอะมิโนที่สนใจเปรียบเทียบกับลำดับกรดอะมิโนในฐานข้อมูล
- Translated BLAST แบ่งออกเป็น
  - BLASTx การใช้ลำดับนิวคลีโอไทด์ที่สนใจมาแปลรหัส (translated nucleotides) ให้เป็นลำดับกรดอะมิโน แล้วนำไปเปรียบเทียบกับลำดับกรดอะมิโนในฐานข้อมูล
  - tBLASTn การใช้ลำดับกรดอะมิโนที่สนใจเปรียบเทียบกับลำดับนิวคลีโอไทด์ในฐานข้อมูลที่แปลรหัสให้เป็นลำดับกรดอะมิโน (translated nucleotides)
  - tBLASTx การใช้ลำดับนิวคลีโอไทด์ที่สนใจ มาแปลรหัส (translated nucleotides) ให้เป็นลำดับกรดอะมิโน แล้วนำไปเปรียบเทียบกับลำดับกรดอะมิโนที่แปลรหัสมาจากนิวคลีโอไทด์ (translated nucleotides) ในฐานข้อมูล

การเลือกใช้ BLAST แต่ละชนิดจะขึ้นกับ 3 ปัจจัย ดังสรุปในตารางที่ 4.1 ได้แก่

- Query sequence ข้อมูลที่เราต้องการวิเคราะห์ แบ่งออกเป็น
  - Nucleotide เป็นข้อมูลที่ได้จากเทคนิค nucleotide sequencing แล้วต้องการค้นหาชื่อยีนและชื่อสิ่งมีชีวิต โดยไม่จำเป็นต้องจะเป็น coding sequences (CDS) หรือ untranslated region (UTR) sequences

- Translated nucleotide เป็นข้อมูลที่ได้จากเทคนิค nucleotide sequencing แล้วต้องการค้นหาชื่อยีนและชื่อสิ่งมีชีวิต โดยเน้นช่วงยีนเฉพาะส่วนที่เป็น coding sequences (CDS) เท่านั้น
- Protein เป็นข้อมูลที่ได้จากเทคนิค mass spectrometry แล้วต้องการค้นหาชื่อโปรตีนและชื่อสิ่งมีชีวิต
- Databases ฐานข้อมูลที่ต้องการใช้ในการค้นหา ซึ่งจะแบ่งเป็นฐานข้อมูลสำหรับ nucleotide และ protein รายละเอียดเพิ่มเติมสรุป ดังตารางที่ 4.2 และ 4.3 ต่อไป
- Alignments การจัดเรียงข้อมูลเพื่อใช้ในการเปรียบเทียบระหว่าง query sequence และ sequences ในฐานข้อมูล ซึ่งแบ่งออกเป็น nucleotide vs nucleotide และ protein vs protein

The image shows the BLAST website interface. At the top, there are logos for NIH (U.S. National Library of Medicine) and NCBI (National Center for Biotechnology Information). The main heading is "BLAST" with navigation links for "Home", "Recent Results", and "Saved". A QR code is located on the right side. The main content area is titled "Basic Local Alignment Search Tool" and includes a brief description of BLAST's function: "BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance." Below this, there are four search options: "Nucleotide BLAST" (nucleotide to nucleotide), "blastx" (translated nucleotide to protein), "tblastn" (protein to translated nucleotide), and "Protein BLAST" (protein to protein). A "BLAST Genomes" section at the bottom allows users to search by organism name or select from a list including Human, Mouse, Rat, and Microbes.

รูปที่ 4.1 หน้าแรกของเว็บไซต์ BLAST [https://blast.ncbi.nlm.nih.gov/Blast.cgi]

#### ตารางที่ 4.1 สรุปเปรียบเทียบลักษณะพื้นฐานของการ BLAST แต่ละชนิด

BLAST types	Query sequences	Database	Alignments
Nucleotide BLAST	Nucleotide	Nucleotide	Nucleotide vs Nucleotide
Protein BLAST	Protein	Protein	Protein vs Protein
BLASTx	Translated nucleotide	Protein	Protein vs Protein
tBLASTn	Protein	Translated nucleotide	Protein vs Protein
tBLASTx	Translated nucleotide	Translated nucleotide	Protein vs Protein


ดังนั้นการใช้ BLAST แต่ละชนิดอาจได้ผลการวิเคราะห์แตกต่างกัน และมีผลต่อความถูกต้องของข้อมูล ดังนั้นควรเลือกใช้อย่างมีวิจารณญาณ รายละเอียดการใช้ BLAST แต่ละชนิดเป็นดังนี้

#### Nucleotide BLAST

Nucleotide BLAST เป็นวิธีมาตรฐานที่ใช้ในการระบุชื่อยีน และชนิดของสิ่งมีชีวิต โดยการใส่ข้อมูลลำดับนิวคลีโอไทด์ (รหัสพันธุกรรม) โดยสามารถคลิกที่คำว่า Nucleotide BLAST จากรูปที่ 4.1 ก็จะเข้าสู่หน้าสำหรับการใส่ข้อมูลของ Nucleotide BLAST (รูปที่ 4.2) ซึ่งแบ่งออกเป็น 3 ส่วน ได้แก่

- Enter Query Sequence เป็นส่วนช่องว่างที่ให้ใส่ข้อมูลลำดับนิวคลีโอไทด์ที่สนใจ (query) หรือจะ upload file ที่มีข้อมูลลำดับนิวคลีโอไทด์ก็ได้
- Choose Search Set เป็นการเลือกใช้ฐานข้อมูลที่จะใช้ในการศึกษาเปรียบเทียบ ในกรณีที่ไม่แน่ใจว่าควรใช้ฐานข้อมูลใด นิยมเลือกใช้ Others (Nucleotide collection) ในการศึกษา อย่างไรก็ตามสามารถปรับเปลี่ยนฐานข้อมูลได้ (ตารางที่ 4.2)

**Standard Nucleotide BLAST**



Enter Query Sequence BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

```

CTCCACACATTCCACCAAGCTCTGTAGACCCCGAGTGGGGGCTATATCTTCTGCTGGTGGC
TCCAGTCCGGACAGTAACCCCTGTCCGACTACTGCCTCACCCATATCGTCAATCTTCTCGAGGA
CTGGGGACCTGCACCGACATGGAGACACACATCAGGATTCTTAGGACCCCTGCTCGTGTACA
GGGGGGTTTTCTTGTGACACAGAACTCTCAATACCCACAGAGTCTAGACTCGTGGTGGACTCT
CTCAATTTCTAGGGGAGCACCCACGTCTGCTGGCCAAATTCGCAGTCCCAACCTCCAATCACT

```

Query subrange

From

To

Or, upload file

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database  Human genomic + transcript  Mouse genomic + transcript  Others

Nucleotide collection (nr/nt)

Organism Optional   Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude Optional  Models (XM/XP)  Uncultured/environmental sample sequences

Limit to Optional  Sequences from type material

Entrez Query Optional  [You Tube](#) [Create custom database](#)

Enter an Entrez query to limit search

Program Selection

Optimize for

Highly similar sequences (megablast)

More dissimilar sequences (discontiguous megablast)

Somewhat similar sequences (blastn)

Choose a BLAST algorithm

**BLAST** Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

Show results in a new window

**Genomic plus Transcript**  
Human genomic plus transcript (Human G+T)  
Mouse genomic plus transcript (Mouse G+T)

**Other Databases**  
**Nucleotide collection (nr/nt)**  
\*6S ribosomal RNA sequences (Bacteria and Archaea)  
Reference RNA sequences (refseq\_rna)  
RefSeq Representative genomes (refseq\_representative\_genomes)  
RefSeq Genome Database (refseq\_genomes)  
Whole-genome shotgun contigs (wgs)  
Expressed sequence tags (est)  
Sequence Read Archive (SRA)  
Transcriptome Shotgun Assembly (TSA)  
High throughput genomic sequences (HTGS)  
Patent sequences(pat)  
Protein Data Bank (pdb)  
Reference genomic sequences (refseq\_genomic)  
Human RefSeqGene sequences(RefSeq\_Gene)  
Genomic survey sequences (gss)  
Sequence tagged sites (dbsts)

รูปที่ 4.2 การใส่ข้อมูล และกำหนดค่าต่างๆ ในการวิเคราะห์ข้อมูลด้วย Nucleotide BLAST

- Program Selection เป็นการเลือกใช้โปรแกรมในการศึกษาเปรียบเทียบ ดังนี้
  - Highly similar sequences (megablast) ใช้สำหรับการค้นหาลำดับนิวคลีโอไทด์ที่มีความเหมือนกันมากๆ มาจากสิ่งมีชีวิตชนิดเดียวกัน
  - More dissimilar sequences (discontiguous megablast) ใช้สำหรับการค้นหาข้อมูลลำดับนิวคลีโอไทด์ที่มีความเหมือนกันไม่มาก มักมาจากสิ่งมีชีวิตต่างชนิดกัน
  - Somewhat similar sequences (blastn) ใช้สำหรับการค้นหาลำดับนิวคลีโอไทด์ที่มีขนาดสั้นๆ ใช้ค้นหาชิ้นที่มาจากสิ่งมีชีวิตต่างชนิดกัน



#### ตารางที่ 4.2 สรุปชื่อและนิยามของฐานข้อมูลนิวคลีโอไทด์ที่ใช้สำหรับ BLAST

ฐานข้อมูล	รายละเอียด
Nucleotide collection (default)	All GenBank + EMBL + DDBJ + PDB sequences, excluding sequences from PAT, EST, STS, GSS, WGS, TSA and phase 0, 1 or 2 HTGS sequences, mostly non-redundant.
16S ribosomal RNA sequences	16S Microbial rRNA sequences from Targeted Loci Project.
Reference RNA sequences	Curated (NM_, NR_) plus predicted (XM_, XR_) sequences from NCBI Reference Sequence Project.
RefSeq Representative genomes	NCBI RefSeq Reference and Representative genomes across broad taxonomy groups including eukaryotes, bacteria, archaea, viruses and viroids. These genomes are among the best quality genomes available with minimum redundancy - one genome per species for eukaryotes and diverse isolates for the same species for others.
RefSeq Genome Databases	Genomic sequences from NCBI Reference Sequence Project.
Whole-genome shotgun contigs	Assemblies of Whole Genome Shotgun sequences.
Expressed sequence tags	Database of GenBank + EMBL + DDBJ sequences from EST division.
Transcriptome Shotgun Assembly	Transcriptome Shotgun Assemblies, assembled from RNA-seq Sequence Read Archive (SRA) data.
High throughput genomic sequences	Unfinished High Throughput Genomic Sequences; Sequences: phases 0, 1 and 2.
Patent sequences	Nucleotides from the Patent division of GenBank.
Protein Data Bank	Nucleotide sequences from the 3-dimensional structure records from Protein Data Bank.
Ref genomic sequences	Complete genomes and complete chromosomes from the NCBI Reference Sequence project.
Human genomic + transcript	The genomic sequences plus curated and predicted RNAs from the current build of the human genome.
Mouse genomic + transcript	The genomic sequences plus curated and predicted RNAs from the current build of the mouse genome.

เมื่อกำหนดค่าต่างๆ เสร็จแล้ว ให้คลิกที่คำว่า BLAST โปรแกรมจะนำลำดับนิวคลีโอไทด์ที่เราใส่ไว้ ไปทำการเปรียบเทียบกับลำดับนิวคลีโอไทด์ทั้งหมดที่มีอยู่ในฐานข้อมูลตามที่เรากำหนด และคำนวณค่านัยสำคัญทางสถิติของความเหมือน พร้อมกับระบุชื่อยีน และชนิดของสิ่งมีชีวิตที่มีความคล้ายคลึงกับลำดับนิวคลีโอไทด์ที่สนใจ โดยแสดงผล ดังรูปที่ 4.3 โดยแบ่งออกเป็น 4 ส่วนได้แก่

- Job title เป็นส่วนที่ระบุค่าต่างๆ ตามที่กำหนดไว้ในการวิเคราะห์ข้อมูล ได้แก่ molecular type (ชนิดของสารชีวโมเลกุล), query length (ความยาว), database name (ชื่อฐานข้อมูล) และ program (โปรแกรมที่ใช้) เป็นต้น
- Graphic summary เป็นการสรุปผลการเปรียบเทียบลำดับนิวคลีโอไทด์ด้วยแถบสี ซึ่งไม่ค่อยมีประโยชน์ในการวิเคราะห์ข้อมูล
- Description เป็นการระบุรายละเอียดชื่อนิวคลีโอไทด์ต่างๆ ในฐานข้อมูลที่มีความคล้ายคลึงกับลำดับนิวคลีโอไทด์ที่เราใส่เข้าไปวิเคราะห์ โดยแสดงผล 100 อันดับแรกซึ่งเรียงลำดับตามค่าคะแนน และค่าทางสถิติ (ลำดับที่ 1 จะมีความเหมือนมากที่สุด) ซึ่งมีค่าที่ควรพิจารณา ดังนี้
  - Description เป็นรายละเอียดชื่อยีนและชื่อสิ่งมีชีวิต
  - Max score และ Total score ค่าคะแนนที่สูงจะมีความเหมือนมาก
  - Query coverage เป็นค่าที่บอกถึงความยาวของนิวคลีโอไทด์ที่ใช้ในการเปรียบเทียบกัน เช่น หากเราใส่นิวคลีโอไทด์ที่มีความยาว 630 bp ลงไป แล้วโปรแกรม BLAST ใช้ทั้ง 630 bp ในการเปรียบเทียบกับนิวคลีโอไทด์ในฐานข้อมูล ก็จะแสดงผลว่าเท่ากับ 100 %
  - Identity เป็นค่าที่บอกเปอร์เซ็นต์ความเหมือนระหว่างนิวคลีโอไทด์
  - E value เป็นค่าที่บอกความน่าเชื่อถือ ถ้าค่าเข้าใกล้ศูนย์ จะมีความแม่นยำสูง
  - Accession number สามารถคลิกเพื่อเชื่อมโยงไปสู่ฐานข้อมูล Nucleotide ได้
- Alignment เป็นส่วนที่แสดงการเปรียบเทียบเบสต่อเบส ระหว่างนิวคลีโอไทด์ที่เราใส่เข้าไป (query) และนิวคลีโอไทด์ที่พบในฐานข้อมูล (subject) หากมีเบสตำแหน่งใดแตกต่างกันก็จะสังเกตเห็นได้อย่างชัดเจน

เครื่องมือในการค้นหาเปรียบเทียบความคล้ายคลึงกันทางชีวภาพ

Edit and Resubmit Save Search Strategies Formatting options Download You Tube How to read this page Blast report de

**Job title: Nucleotide Sequence (630 letters)**

RID [D3DFBZHR01B](#) (Expires on 04-15 16:03 pm)

Query ID [lcl|Query\\_45293](#) Database Name [nr](#)

Description [None](#) Description [Nucleotide collection \(nt\)](#)

Molecule type [nucleic acid](#) Program [BLASTN 2.8.0+](#) [Citation](#)

Query Length [630](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [MSA viewer](#)

**Graphic Summary**

Distribution of the top 100 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments

**Color key for alignment scores**

<40 40-50 50-80 80-200 >=200

Query

1 100 200 300 400 500 600

**Descriptions**

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

Alignments [Download](#) [GenBank](#) [Graphics](#) [Distance tree of results](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Hepatitis B virus DNA, complete genome, clone: C_JPNAT</a>	1164	1164	100%	0.0	100%	<a href="#">AB246345.1</a>
<input type="checkbox"/>	<a href="#">Hepatitis B virus isolate 26_5 core protein (C) gene, complete cds</a>	1158	1158	100%	0.0	99%	<a href="#">KU964173.1</a>
<input type="checkbox"/>	<a href="#">Hepatitis B virus isolate cww1128, complete genome</a>	1158	1158	100%	0.0	99%	<a href="#">KC774276.1</a>

**Alignments**

[Download](#) [GenBank](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

**Hepatitis B virus DNA, complete genome, clone: C\_JPNAT**

Sequence ID: [AB246345.1](#) Length: 3215 Number of Matches: 1

Range 1: 1 to 630 [GenBank](#) [Graphics](#) [Next Hit/1](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand	
1164 bits(630)	0.0	630/630(100%)	0/630(0%)	Plus/Plus	
Query 1	CTCCACAACATTCACCAAGCTCTGCTAGACCCAGAGTGAGGGGCTATATCTTCTCTGC	60			
Sbjct 1	CTCCACAACATTCACCAAGCTCTGCTAGACCCAGAGTGAGGGGCTATATCTTCTCTGC	60			
Query 61	TGGTGGCTCCAGTTCGGGAACAGTAAACCCTGTTCCGACTACTGCCTCACCCATATCGTC	120			
Sbjct 61	TGGTGGCTCCAGTTCGGGAACAGTAAACCCTGTTCCGACTACTGCCTCACCCATATCGTC	120			
Query 121	AATCTTCTCGAGGACTGGGGACCCTGCACCGAACATGGAGAACAACATCAGGATTCCCT	180			
Sbjct 121	AATCTTCTCGAGGACTGGGGACCCTGCACCGAACATGGAGAACAACATCAGGATTCCCT	180			
Query 181	AGGACCCCTGCTCGTGTACAGGCGGGTTTTCTTGTGACAAGAATCCTCACAAATACC	240			
Sbjct 181	AGGACCCCTGCTCGTGTACAGGCGGGTTTTCTTGTGACAAGAATCCTCACAAATACC	240			
Query 241	ACAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTTCTAGGGGGAGCACCCAGTGTCC	300			
Sbjct 241	ACAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTTCTAGGGGGAGCACCCAGTGTCC	300			

รูปที่ 4.3 ตัวอย่างผลที่ได้จาก Nucleotide BLAST

## Protein BLAST

Protein BLAST เป็นวิธีมาตรฐานที่ใช้ในการระบุชื่อโปรตีนและชนิดของสิ่งมีชีวิต โดยการใส่ข้อมูลลำดับกรดอะมิโนของโปรตีนที่สนใจ โดยสามารถคลิกที่คำว่า Protein BLAST จากรูปที่ 4.1 ก็จะเข้าสู่หน้าสำหรับการใส่ข้อมูลของ Protein BLAST (รูปที่ 4.4) ซึ่งแบ่งออกเป็น 3 ส่วน ได้แก่

The screenshot shows the 'Standard Protein BLAST' web interface. It includes a 'Query subrange' section with 'From' and 'To' input fields. The 'Choose Search Set' section has a dropdown menu for 'Database' set to 'Non-redundant protein sequences (nr)'. A list of database options is shown, including 'Reference proteins (refseq\_protein)', 'Model Organisms (landmark)', 'UniProtKB/Swiss-Prot (swissprot)', 'Patented protein sequences (pat)', 'Protein Data Bank proteins (pdb)', 'Metagenomic proteins (env\_nr)', and 'Transcriptome Shotgun Assembly proteins (tsa\_nr)'. The 'Program Selection' section has radio buttons for 'Quick BLASTP (Accelerated protein-protein BLAST)', 'blastp (protein-protein BLAST)', 'PSI-BLAST (Position-Specific Iterated BLAST)', 'PHI-BLAST (Pattern Hit Initiated BLAST)', and 'DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)'. The 'blastp' option is selected. At the bottom, there is a 'BLAST' button and a search summary: 'Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)'. There is also a checkbox for 'Show results in a new window'.

รูปที่ 4.4 การใส่ข้อมูล และกำหนดค่าต่างๆ ในการวิเคราะห์ข้อมูลด้วย Protein BLAST

- Enter Query Sequence เป็นส่วนช่องว่างที่ให้ใส่ข้อมูลลำดับกรดอะมิโนที่สนใจ (query) หรือจะ upload file ที่มีข้อมูลลำดับกรดอะมิโนก็ได้
- Choose Search Set เป็นการเลือกใช้ฐานข้อมูลที่จะใช้ในการศึกษาเปรียบเทียบ ในกรณีที่ไม่แน่ใจว่าควรใช้ฐานข้อมูลใด นิยมเลือกใช้ Non-redundant protein sequences (nr) อย่างไรก็ตามสามารถปรับเปลี่ยนเป็นฐานข้อมูลอื่นได้ (ตารางที่ 4.3)

- Program Selection เป็นการเลือกใช้โปรแกรมในการศึกษาเปรียบเทียบ ดังนี้
  - BLASTp (protein-protein BLAST) ใช้สำหรับการค้นหาลำดับกรดอะมิโนที่มีความเหมือนกันมากๆ มาจากสิ่งมีชีวิตชนิดเดียวกัน
  - Quick BLASTp (Accelerated protein-protein BLAST) ใช้สำหรับการค้นหาลำดับกรดอะมิโนที่มีความเหมือนกันอย่างรวดเร็ว
  - PSI-BLAST (Position specific Iterated BLAST) ใช้สำหรับการค้นหาลำดับกรดอะมิโนเฉพาะตำแหน่ง ใช้สำหรับโปรตีนที่มีความเหมือนกันไม่มาก มาจากสิ่งมีชีวิตต่างชนิดกัน
  - PHI-BLAST (Protein Hit Initiated BLAST) ใช้สำหรับการค้นหาโปรตีนโดยอาศัยรูปแบบเฉพาะตัวของโปรตีนในการเปรียบเทียบ
  - DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST) ใช้สำหรับการค้นหาลำดับกรดอะมิโนที่มีความเหมือนกัน โดยเปรียบเทียบเป็นส่วนๆ (domain) ทำให้เพิ่มความไวในการค้นหา

ตารางที่ 4.3 สรุปชื่อและนิยามของฐานข้อมูลโปรตีนที่ใช้สำหรับ BLAST

ฐานข้อมูล	รายละเอียด
Non-redundant protein sequences (default)	Non-redundant GenBank CDS translations + RefSeq + PDB + SwissProt + PIR + PRF, excluding those in PAT, TSA, and env_nr.
Reference proteins	Protein sequences from NCBI Reference Sequence project.
Model Organisms	The landmark database includes proteomes from representative genomes spanning a wide taxonomic range.
UniProtKB/Swiss-Prot	Last major release of the UniProtKB/SWISS-PROT protein sequence database (no incremental updates).
Patented protein sequences	Proteins from the Patent division of GenBank.
Protein Data Bank	Protein sequences from the 3-dimensional structure records from the Protein Data Bank.
Metagenomic proteins	Protein sequences translated from the CDS annotation of metagenomic nucleotide sequences.
Transcriptome Shotgun Assembly proteins	Protein sequences translated from CDSs annotated on transcriptome shotgun assemblies.

เมื่อกำหนดค่าต่างๆ เสร็จแล้ว ให้คลิกที่คำว่า BLAST โปรแกรมจะนำลำดับกรดอะมิโนที่เราใส่ไว้ ไปทำการเปรียบเทียบกับลำดับกรดอะมิโนทั้งหมดที่มีอยู่ในฐานข้อมูลตามที่เรากำหนด และคำนวณค่านัยสำคัญทางสถิติของความเหมือน พร้อมกับระบุชื่อโปรตีน และชนิดของสิ่งมีชีวิตที่มีความคล้ายคลึงกับลำดับกรดอะมิโนที่สนใจ โดยแสดงผล ดังรูปที่ 4.5 โดยแบ่งออกเป็น 4 ส่วน ได้แก่

- Job title เป็นส่วนที่ระบุค่าต่างๆ ตามที่กำหนดไว้ในการวิเคราะห์ข้อมูล ได้แก่ molecular type (ชนิดของสารชีวโมเลกุล), query length (ความยาว), database name (ชื่อฐานข้อมูล) และ program (โปรแกรมที่ใช้) เป็นต้น
- Graphic summary เป็นการสรุปผลการเปรียบเทียบลำดับกรดอะมิโนด้วยแถบสี มีการแบ่งโครงสร้างของโปรตีนเป็นส่วน (conserved domains) ต่างๆ
- Description เป็นการระบุรายละเอียดชื่อโปรตีนต่างๆ ในฐานข้อมูลที่มีความคล้ายคลึงกับลำดับกรดอะมิโนที่เราใส่เข้าไปวิเคราะห์ โดยแสดงผล 100 อันดับแรกซึ่งเรียงลำดับตามค่าคะแนน และค่าทางสถิติ (ลำดับที่ 1 จะมีความเหมือนมากที่สุด) ซึ่งมีค่าที่ควรพิจารณา ดังนี้
  - Description เป็นรายละเอียดชื่อโปรตีนและชื่อสิ่งมีชีวิต
  - Max score และ Total score ค่าคะแนนที่สูงจะมีความเหมือนมาก
  - Query coverage เป็นค่าที่บอกถึงความยาวของโปรตีนที่ใช้ในการเปรียบเทียบกัน เช่น หากเราใส่โปรตีนที่ประกอบด้วยกรดอะมิโน 469 ตัวลงไป แล้วโปรแกรม BLAST ใช้กรดอะมิโนทั้ง 469 ตัว ในการเปรียบเทียบกับโปรตีนในฐานข้อมูล ก็จะแสดงผลว่าเท่ากับ 100 %
  - Identity เป็นค่าที่บอกเปอร์เซ็นต์ความเหมือนระหว่างโปรตีน
  - E value เป็นค่าที่บอกความน่าเชื่อถือ ถ้าค่าเข้าใกล้ศูนย์ จะมีความแม่นยำสูง
  - Accession number สามารถคลิก เพื่อเชื่อมโยงไปสู่ฐานข้อมูล Protein ได้
- Alignment เป็นส่วนที่แสดงการเปรียบเทียบกรดอะมิโนแต่ละตำแหน่ง ระหว่างโปรตีนที่เราใส่เข้าไป (query) และโปรตีนที่พบในฐานข้อมูล (subject) หากมีกรดอะมิโนตำแหน่งใดแตกต่างกันก็จะสังเกตเห็นได้อย่างชัดเจน



เครื่องมือในการค้นหาเปรียบเทียบความคล้ายคลึงกันทางชีวภาพ

### BLAST Results

[Edit and Resubmit](#) | [Save Search Strategies](#) | [Formatting options](#) | [Download](#) | [YouTube](#) | [How to read this page](#) | [Blast report de](#)

**Job title: Protein Sequence (469 letters)**

RID [D3JKBKJ4015](#) (Expires on 04-15 17:31 pm)

**Query ID** lc|Query\_49133  
**Description** None  
**Molecule type** amino acid  
**Query Length** 469

**Database Name** nr  
**Description** All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects  
**Program** BLASTP 2.8.0+ [Citation](#)

Other reports: [Search Summary](#) | [Taxonomy reports](#) | [Distance tree of results](#) | [Multiple alignment](#) | [MSA viewer](#)

#### Graphic Summary

Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed results.

Distribution of the top 100 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments

Color key for alignment scores

<40	40-50	50-80	80-200	>=200
-----	-------	-------	--------	-------

Query: 1 90 180 270 360 450

#### Descriptions

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

Alignments [Download](#) [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	neuraminidase [Influenza A virus (A/Ohio/07/2009(H1N1))]	956	956	100%	0.0	100%	ACP44181.1
<input type="checkbox"/>	neuraminidase [Influenza A virus (A/Poland/282/2009(H1N1))]	956	956	100%	0.0	99%	ACU68925.2

#### Alignments

[Download](#) [GenPept](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

neuraminidase, partial [Influenza A virus (A/Ohio/07/2009(H1N1))]  
 Sequence ID: [ACP44181.1](#) Length: 469 Number of Matches: 1  
[See 3405 more title\(s\)](#)

Range 1: 1 to 469 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
956 bits(2472)	0.0	Compositional matrix adjust.	469/469(100%)	469/469(100%)	0/469(0%)

Query 1 MNFNQKIITIGSVCMTIGMANLILQIGNIISIWISHSIQLGNQNQIETCNQSVITYENNT 60  
 MNFNQKIITIGSVCMTIGMANLILQIGNIISIWISHSIQLGNQNQIETCNQSVITYENNT  
 Sbjct 1 MNFNQKIITIGSVCMTIGMANLILQIGNIISIWISHSIQLGNQNQIETCNQSVITYENNT 60

**Related Information**  
[Identical Proteins - Identical proteins to ACP44181.1](#)

รูปที่ 4.5 ตัวอย่างผลที่ได้จาก Protein BLAST

## Translated BLAST: BLASTx

BLASTx เป็นวิธีมาตรฐานที่ใช้ในการระบุชื่อโปรตีน และชนิดของสิ่งมีชีวิต โดยการใส่ข้อมูลลำดับนิวคลีโอไทด์ (รหัสพันธุกรรม) โดยสามารถคลิกที่คำว่า BLASTx จากรูปที่ 4.1 ก็จะเข้าสู่หน้าสำหรับการใส่ข้อมูลของ BLASTx (รูปที่ 4.6) ซึ่งแบ่งออกเป็น 2 ส่วน ได้แก่

- Enter Query Sequence เป็นส่วนช่องว่างที่ให้ใส่ข้อมูลลำดับนิวคลีโอไทด์ที่สนใจ (query) หรือจะ upload file ที่มีข้อมูลลำดับนิวคลีโอไทด์ก็ได้ นอกจากนี้สามารถเลือก Genetic code ในการกำหนดรหัส codons ที่จะใช้ในการแปลรหัส (translation) จากลำดับนิวคลีโอไทด์ไปเป็นกรดอะมิโน ซึ่งโดยทั่วไปจะเลือกเป็น Standard อย่างไรก็ตามสามารถปรับเปลี่ยนให้เหมาะสมกับชนิดของสิ่งมีชีวิตที่สนใจได้ดังนี้ Vertebrate Mitochondrial, Yeast Mitochondrial, Mold Mitochondrial, Invertebrate Mitochondrial, Ciliate Nuclear, Echinoderm Mitochondrial, Euplotid Nuclear, Bacteria and Archaea, Alternative Yeast Nuclear, Ascidian Mitochondrial, Flatworm Mitochondrial และ Blepharisma Macronuclear เป็นต้น

รูปที่ 4.6 การใส่ข้อมูล และกำหนดค่าต่างๆ ในการวิเคราะห์ข้อมูลด้วย BLASTx

- Choose Search Set เป็นการเลือกใช้ฐานข้อมูลที่จะใช้ในการศึกษาเปรียบเทียบ ในกรณีที่ไม่แน่ใจว่าควรใช้ฐานข้อมูลใด นิยมเลือกใช้ Non-redundant protein sequences (nr) ในการศึกษา อย่างไรก็ตามสามารถปรับเปลี่ยนเป็นฐานข้อมูลอื่น ได้แก่ Reference proteins, Model Organisms, UniProtKB/Swiss-Prot, Patented protein sequences, Protein Data Bank, Metagenomic proteins และ Transcriptome Shotgun Assembly proteins เป็นต้น

เมื่อกำหนดค่าต่างๆ เสร็จแล้ว ให้คลิกที่คำว่า BLAST โปรแกรมจะนำลำดับนิวคลีโอไทด์ที่เราใส่ไว้ ไปทำการแปลรหัสเป็นกรดอะมิโน แล้วจึงนำไปเปรียบเทียบกับลำดับกรดอะมิโนทั้งหมดที่มีอยู่ในฐานข้อมูลตามที่เรากำหนด และคำนวณค่านัยสำคัญทางสถิติของความเหมือน พร้อมกับระบุชื่อโปรตีน และชนิดของสิ่งมีชีวิตที่มีความคล้ายคลึงกัน โดยแสดงผลดังรูปที่ 4.7 ซึ่งแบ่งออกเป็น 4 ส่วนได้แก่

- Job title เป็นส่วนที่ระบุค่าต่างๆ ตามที่กำหนดไว้ในการวิเคราะห์ข้อมูล ได้แก่ molecular type (ชนิดของสารชีวโมเลกุล), query length (ความยาว), database name (ชื่อฐานข้อมูล) และ program (โปรแกรมที่ใช้) เป็นต้น
- Graphic summary เป็นการสรุปผลการเปรียบเทียบลำดับนิวคลีโอไทด์และกรดอะมิโน ด้วยแถบสี มีการแบ่งโครงสร้างของโปรตีนเป็นส่วน (conserved domains) ต่างๆ
- Description เป็นการบรรยายละเอียดชื่อโปรตีนต่างๆ ในฐานข้อมูลที่มีความคล้ายคลึงกับลำดับกรดอะมิโนที่แปลรหัสมาจากลำดับนิวคลีโอไทด์ (translated nucleotides) ที่เราใส่เข้าไปวิเคราะห์ โดยแสดงผล 100 อันดับแรกซึ่งเรียงลำดับตามค่าคะแนน และค่าทางสถิติ ซึ่งมีค่าที่ควรพิจารณา ดังนี้
  - Description เป็นรายละเอียดชื่อโปรตีนและชื่อสิ่งมีชีวิต
  - Max score และ Total score ค่าคะแนนที่สูงจะมีความเหมือนมาก
  - Query coverage เป็นค่าที่บอกถึงความยาวของโปรตีนที่ใช้ในการเปรียบเทียบ
  - Identity เป็นค่าที่บอกเปอร์เซ็นต์ความเหมือนระหว่างโปรตีน
  - E value เป็นค่าที่บอกความน่าเชื่อถือ ถ้าค่าเข้าใกล้ศูนย์ จะมีความแม่นยำสูง
  - Accession number สามารถคลิก เพื่อเชื่อมโยงไปสู่ฐานข้อมูล Protein ได้

Edit and Resubmit Save Search Strategies Formatting options Download [YouTube](#) [How to read this page](#) [Blast report des](#)

**Job title: Nucleotide Sequence (630 letters)**

RID [D3TA4TH1015](#) (Expires on 04-15 19:25 pm)

Query ID [Id|Query\\_119717](#)

Description None

Molecule type nucleic acid

Query Length 630

Database Name nr

Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects

Program [BLASTX 2.8.0+](#) [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#)

**Graphic Summary**

Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed results.

Distribution of the top 100 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments

Color key for alignment scores

- <40
- 40-50
- 50-80
- 80-200
- >=200

**Descriptions**

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

Alignments [Download](#) [GenPept](#) [Graphics](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">polymerase [Hepatitis B virus]</a>	393	393	100%	5e-135	99%	<a href="#">ABQ43938.1</a>
<input type="checkbox"/>	<a href="#">polymerase [Hepatitis B virus]</a>	392	392	100%	9e-135	99%	<a href="#">AF119408.1</a>

**Alignments**

[Download](#) [GenPept](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

polymerase, partial [Hepatitis B virus]

Sequence ID: [ABQ43938.1](#) Length: 394 Number of Matches: 1

Range 1: 64 to 273 [GenPept](#) [Graphics](#) [View Full](#) [Previous](#) [Next](#)

Score	Expect	Method	Identities	Positives	Gaps	Frame
393 bits(1009)	5e-135	Compositional matrix adjust.	209/210(99%)	209/210(99%)	0/210(0%)	+1
Query 1	LHNI PPSSAR PQSEGPISSCWWLQFRNSKPCSDYCLTHIVNLLD WGPCTEHGEHNIRIP					180
Sbjct 64	LHNI PPSSAR PQSEGPI SCWWLQFRNSKPCSDYCLTHIVNLLD WGPCTEHGEHNIRIP					123
Query 181	RTPARVTGGVFLVDKNPHNTESRLVVD FSGQFSRGSTHVS WPKFAV Pnlgsitnlissn l					360
Sbjct 124	RTPARVTGGVFLVDKNPHNTESRLVVD FSGQFSRGSTHVS WPKFAV PNLQSLTNLLSSNL					183

**Related Information**

รูปที่ 4.7 ตัวอย่างผลที่ได้จาก Translated BLAST: BLASTx

- Alignment เป็นส่วนที่แสดงการเปรียบเทียบกรดอะมิโนแต่ละตำแหน่ง ระหว่างกรดอะมิโนที่แปลรหัสมาจากลำดับนิวคลีโอไทด์ (translated nucleotides) ที่เราใส่เข้าไป (query) และโปรตีนที่พบในฐานข้อมูล (subject) หากมีกรดอะมิโนตำแหน่งใดแตกต่างกันก็จะสังเกตเห็นได้อย่างชัดเจน

### Translated BLAST: tBLASTn


tBLASTn เป็นวิธีมาตรฐานที่ใช้ในการระบุชื่อยีน และชนิดของสิ่งมีชีวิต โดยการใส่ข้อมูลลำดับกรดอะมิโนที่สนใจ ไปเปรียบเทียบกับลำดับนิวคลีโอไทด์ในฐานข้อมูลที่แปลรหัสให้เป็นลำดับกรดอะมิโน (translated nucleotides) โดยสามารถคลิกที่คำว่า tBLASTn จากรูปที่ 4.1 ก็จะเข้าสู่หน้าสำหรับการใส่ข้อมูลของ tBLASTn (รูปที่ 4.8) ซึ่งแบ่งออกเป็น 2 ส่วน ได้แก่

- Enter Query Sequence เป็นส่วนช่องว่างที่ให้ใส่ข้อมูลลำดับนิวคลีโอไทด์ที่สนใจ (query) หรือจะ upload file ที่มีข้อมูลลำดับนิวคลีโอไทด์ก็ได้
- Choose Search Set เป็นการเลือกใช้ฐานข้อมูลในการศึกษาเปรียบเทียบ นิยมใช้ Nucleotide collection อย่างไรก็ตามสามารถปรับเปลี่ยนฐานข้อมูลได้ ดังนี้ 16S ribosomal RNA sequences, Reference RNA sequences, RefSeq Representative genomes, RefSeq Genome Databases, Whole-genome shotgun contigs, Expressed sequence tags, Sequence Read Archive, Transcriptome Shotgun Assembly, High throughput genomic sequences, Patent sequences, Protein Data Bank, Ref genomic sequences, Human RefSeqGene sequences, Genomic survey sequences และ Sequence tagged sites เป็นต้น

เมื่อกำหนดค่าต่างๆ เสร็จแล้ว ให้คลิกที่คำว่า BLAST โปรแกรมจะนำลำดับกรดอะมิโนที่เราใส่ไว้ไปเปรียบเทียบกับลำดับกรดอะมิโนที่ถูกแปลรหัสมาจากลำดับนิวคลีโอไทด์ (translated nucleotides) ทั้งหมดที่มีอยู่ในฐานข้อมูลตามที่เรากำหนด และคำนวณค่านัยสำคัญทางสถิติของความเหมือน พร้อมกับระบุชื่อยีน และชนิดของสิ่งมีชีวิตที่มีความคล้ายคลึงกันโดยแสดงผล ดังรูปที่ 4.9 โดยแบ่งออกเป็น 4 ส่วนได้แก่



Translated BLAST: tblastn



Enter Query Sequence Reset

TBLASTN search translated nucleotide databases using a protein query. more...

Enter accession number(s), gi(s), or FASTA sequence(s) Clear Query subrange

From   
To

Or, upload file Browse...

Job Title

Align two or more sequences

Choose Search Set

Database

Organism   Exclude

Exclude  Models (XM/XP)  Uncultured/environmental sample sequences

Limit to  Sequences from type material

Entrez Query  [YouTube](#) [Create custom database](#)

**BLAST** Search database **Nucleotide collection (nr/nt)** using **Tblastn** (search translated nucleotide databases using a protein query)

Show results in a new window

รูปที่ 4.8 การใส่ข้อมูล และกำหนดค่าต่างๆ ในการวิเคราะห์ข้อมูลด้วย tBLASTn

- Job title เป็นส่วนที่ระบุค่าต่างๆ ตามที่กำหนดไว้ในការวิเคราะห์ข้อมูล ได้แก่ molecular type (ชนิดของสารชีวโมเลกุล), query length (ความยาว), database name (ชื่อฐานข้อมูล) และ program (โปรแกรมที่ใช้) เป็นต้น
- Graphic summary เป็นการสรุปผลการเปรียบเทียบลำดับกรดอะมิโนด้วยแถบสี
- Description เป็นการบรรยายละเอียดชื่ออื่นต่างๆ ในฐานข้อมูลที่มีลำดับกรดอะมิโนที่แปลรหัสมาจากลำดับนิวคลีโอไทด์ (translated nucleotides) ความคล้ายคลึงกับลำดับกรดอะมิโนที่เราใส่เข้าไปวิเคราะห์ โดยแสดงผล 100 อันดับแรกซึ่งเรียงลำดับตามค่าคะแนน และค่าทางสถิติ ซึ่งมีค่าที่ควรพิจารณา ดังนี้
  - Description เป็นรายละเอียดชื่ออื่นและชื่อสิ่งมีชีวิต
  - Max score และ Total score ค่าคะแนนที่สูงจะมีความเหมือนมาก
  - Query coverage เป็นค่าที่บอกถึงความยาวของโปรตีนที่ใช้ในการเปรียบเทียบ
  - Identity เป็นค่าที่บอกเปอร์เซ็นต์ความเหมือนระหว่างโปรตีน
  - E value เป็นค่าที่บอกความน่าเชื่อถือ ถ้าค่าเข้าใกล้ศูนย์ จะมีความแม่นยำสูง
  - Accession number สามารถคลิกเพื่อเชื่อมโยงไปสู่ฐานข้อมูล Nucleotide ได้



เครื่องมือในการค้นหาเปรียบเทียบความคล้ายคลึงกันทางชีวภาพ

- Alignment เป็นส่วนที่แสดงการเปรียบเทียบกรดอะมิโนแต่ละตำแหน่ง ระหว่างกรดอะมิโนที่ที่เราใส่เข้าไป (query) และกรดอะมิโนที่แปลรหัสมาจากลำดับนิวคลีโอไทด์ (translated nucleotides) ที่พบในฐานข้อมูล (subject) หากมีกรดอะมิโนตำแหน่งใดแตกต่างกันก็จะสังเกตเห็นได้อย่างชัดเจน

**BLAST Results**

Edit and Resubmit Save Search Strategies Formatting options Download [YouTube](#) [How to read this page](#) [Blast report de](#)

**Job title: Protein Sequence (469 letters)**

RID [D3V6GNAR014](#) (Expires on 04-15 19:58 pm)

Query ID [Ic|Query\\_163877](#) Database Name [nr](#)

Description [None](#) Description [Nucleotide collection \(nt\)](#)

Molecule type [amino acid](#) Program [TBLASTN 2.8.0+ Citation](#)

Query Length [469](#)

Other reports: [Search Summary](#) [Taxonomy reports](#)

**Graphic Summary**

**Distribution of the top 100 Blast Hits on 100 subject sequences**

Mouse over to see the title, click to show alignments

**Color key for alignment scores**

<40    
  40-50    
  50-80    
  80-200    
  >=200

Query

1    
  90    
  180    
  270    
  360    
  450

**Descriptions**

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

[Alignments](#) [Download](#) [GenBank](#) [Graphics](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Influenza A virus (A/Singapore/443N/2009(H1N1)) segment 6 neuraminidase (NA) gene, complete cds</a>	957	957	100%	0.0	100%	<a href="#">KF695095.1</a>
<input type="checkbox"/>	<a href="#">Influenza A virus (A/Singapore/426N/2009(H1N1)) segment 6 neuraminidase (NA) gene, complete cds</a>	957	957	100%	0.0	100%	<a href="#">KF695091.1</a>

**Alignments**

[Download](#) [GenBank](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

**Influenza A virus (A/Singapore/443N/2009(H1N1)) segment 6 neuraminidase (NA) gene, complete cds**

Sequence ID: [KF695095.1](#) Length: 1453 Number of Matches: 1

Range 1: 16 to 1422 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps	Frame
957 bits(2474)	0.0	Compositional matrix adjust.	469/469(100%)	469/469(100%)	0/469(0%)	+1

```

Query 1      MNPNQKIITIGSVCMTIGMANLILQIGNIISIWIHSHSIQLGNQNQIETCNQSVITYENNT 60
Sbjct 16     MNPNQKIITIGSVCMTIGMANLILQIGNIISIWIHSHSIQLGNQNQIETCNQSVITYENNT 195
    
```

**Related Information**

รูปที่ 4.9 ตัวอย่างผลที่ได้จาก Translated BLAST: tBLASTn

## Translated BLAST: tBLASTx

tBLASTx เป็นวิธีมาตรฐานที่ใช้ในการระบุยีน และชนิดของสิ่งมีชีวิต โดยการใส่ข้อมูลลำดับนิวคลีโอไทด์ที่สนใจ แล้วแปลรหัส (translated nucleotides) ให้เป็นลำดับกรดอะมิโน จากนั้นจึงนำไปเปรียบเทียบกับลำดับกรดอะมิโนที่แปลรหัสมาจากนิวคลีโอไทด์ (translated nucleotides) ในฐานข้อมูล โดยสามารถคลิกที่คำว่า tBLASTx จากรูปที่ 4.2, 4, 6 หรือ 8 ก็จะเข้าสู่หน้าสำหรับการใส่ข้อมูลของ tBLASTx (รูปที่ 4.10) ซึ่งแบ่งออกเป็น 2 ส่วน ได้แก่

รูปที่ 4.10 การใส่ข้อมูล และกำหนดค่าต่างๆ ในการวิเคราะห์ข้อมูลด้วย tBLASTx

- Enter Query Sequence เป็นส่วนช่องว่างที่ให้ใส่ข้อมูลลำดับนิวคลีโอไทด์ที่สนใจ (query) หรือจะ upload file ที่มีข้อมูลลำดับนิวคลีโอไทด์ก็ได้ นอกจากนี้สามารถเลือก Genetic code ในการกำหนดรหัส codons ที่จะใช้ในการแปลรหัส (translation) จากลำดับนิวคลีโอไทด์ไปเป็นกรดอะมิโน ซึ่งโดยทั่วไปจะเลือกเป็น Standard อย่างไรก็ตามสามารถปรับเปลี่ยนให้เหมาะสมกับชนิดของสิ่งมีชีวิตที่สนใจได้
- Choose Search Set เป็นการเลือกใช้ฐานข้อมูลในการศึกษาเปรียบเทียบ นิยมใช้ Nucleotide collection อย่างไรก็ตามสามารถปรับเปลี่ยนฐานข้อมูลได้

เมื่อกำหนดค่าต่างๆ เสร็จแล้ว ให้คลิกที่คำว่า BLAST โปรแกรมจะนำลำดับนิวคลีโอไทด์ที่เราใส่ไว้ไปแปลรหัส (translated nucleotides) ให้เป็นลำดับกรดอะมิโน แล้วจึงไปเปรียบเทียบกับลำดับกรดอะมิโนที่ได้จากการแปลรหัสจากลำดับนิวคลีโอไทด์ (translated nucleotides) ทั้งหมดที่มีอยู่ในฐานข้อมูล ตามที่เรากำหนด จากนั้นคำนวณค่านัยสำคัญทางสถิติของความเหมือน พร้อมกับระบุชื่อยีน และชนิดของสิ่งมีชีวิตที่มีความคล้ายคลึงกันโดยแสดงผลดังรูปที่ 4.11 โดยแบ่งออกเป็น 4 ส่วน ได้แก่

- Job title เป็นส่วนที่ระบุค่าต่างๆ ตามที่กำหนดไว้ในการวิเคราะห์ข้อมูล ได้แก่ molecular type (ชนิดของสารชีวโมเลกุล), query length (ความยาว), database name (ชื่อฐานข้อมูล) และ program (โปรแกรมที่ใช้) เป็นต้น
- Graphic summary เป็นการสรุปผลการเปรียบเทียบลำดับนิวคลีโอไทด์ด้วยแถบสี
- Description เป็นการบรรยายละเอียดชื่อยีนต่างๆ ในฐานข้อมูลที่มีลำดับกรดอะมิโนที่แปลรหัสมาจากลำดับนิวคลีโอไทด์ (translated nucleotides) ความคล้ายคลึงกับลำดับกรดอะมิโนที่แปลรหัสมาจากลำดับนิวคลีโอไทด์ (translated nucleotides) ที่เราใส่เข้าไปวิเคราะห์ โดยแสดงผล 100 อันดับแรกซึ่งเรียงลำดับตามค่าคะแนน และค่าทางสถิติ ซึ่งมีค่าที่ควรพิจารณา ดังนี้
  - Description เป็นรายละเอียดชื่อยีนและชื่อสิ่งมีชีวิต
  - Max score และ Total score ค่าคะแนนที่สูงจะมีความเหมือนมาก
  - Query coverage เป็นค่าที่บอกถึงความยาวของนิวคลีโอไทด์ที่ใช้ในการเปรียบเทียบกัน
  - Identity เป็นค่าที่บอกเปอร์เซ็นต์ความเหมือนระหว่างโปรตีน
  - E value เป็นค่าที่บอกความน่าเชื่อถือ ถ้าค่าเข้าใกล้ศูนย์ จะมีความแม่นยำสูง
  - Accession number สามารถคลิกเพื่อเชื่อมโยงไปสู่ฐานข้อมูล Nucleotide ได้
- Alignment เป็นส่วนที่แสดงการเปรียบเทียบกรดอะมิโนแต่ละตำแหน่ง ระหว่างกรดอะมิโนที่ที่เราใส่เข้าไป (query) และกรดอะมิโนที่แปลรหัสมาจากลำดับนิวคลีโอไทด์ (translated nucleotides) ที่พบในฐานข้อมูล (subject) หากมีกรดอะมิโนตำแหน่งใดแตกต่างกันก็จะสังเกตเห็นได้อย่างชัดเจน

**BLAST Results**

[Edit and Resubmit](#)   [Save Search Strategies](#)   [Formatting options](#)   [Download](#)   [YouTube](#)   [How to read this page](#)   [Blast report de](#)

**Job title: Nucleotide Sequence (630 letters)**

RID [Q3VZAN41014](#) (Expires on 04-15 20:11 pm)

**Query ID** |cl|Query\_154459  
**Description** None  
**Molecule type** nucleic acid  
**Query Length** 630

**Database Name** nr  
**Description** Nucleotide collection (nt)  
**Program** TBLASTX 2.8.0+ [Citation](#)

Other reports: [Search Summary](#) [[Taxonomy reports](#)]

**Graphic Summary**

Distribution of the top 1087 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments

**Color key for alignment scores**

<40    
  40-50    
  50-80    
  80-200    
  >=200

**Descriptions**

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

[Alignments](#)   [Download](#)   [GenBank](#)   [Graphics](#)

	Description	Max score	Total score	Query cover	E value	N	Accession
<input type="checkbox"/>	<a href="#">Echinostoma caproni genome assembly E_caproni_Egypt_scaffold ECPE_scaffold0000044</a>	42.3	42.3	17%	2.9	1	<a href="#">LL233139.1</a>
<input type="checkbox"/>	<a href="#">Heligmosomoides polygyrus genome assembly H_bakeri_Edinburgh_scaffold HPBE_contig0002770</a>	30.4	83.8	25%	5.2	3	<a href="#">LL206793.1</a>
<input type="checkbox"/>	<a href="#">Hepatitis B virus isolate 26_5_core protein (C) gene, complete cds</a>	474	2873	100%	2e-130	2	<a href="#">KU964173.1</a>

**Alignments**

[Download](#)   [GenBank](#)   [Graphics](#)   [Next](#)   [Previous](#)   [Descriptions](#)

**Echinostoma caproni genome assembly E\_caproni\_Egypt, scaffold ECPE\_scaffold0000044**  
 Sequence ID: [LL233139.1](#) Length: 201150 Number of Matches: 1

Range 1: 90283 to 90390 [GenBank](#) [Graphics](#)   [Next Match](#)   [Previous Match](#)

Score	Expect	Identities	Positives	Gaps	Frame
42.3 bits(86)	2.9	16/36(44%)	21/36(58%)	0/36(0%)	+2/+1

Query 482 LPGTSTTSTGPKCTCTIPAQGTSMFSCCCTKPSDG 589  
 LPG +TT+ P C+IPA T + P C +P DG  
 Sbjct 90283 LPGVNTTNISPRMVCISIPATST\*LSPLCIAQQPYDG 90390

[Download](#)   [GenBank](#)   [Graphics](#)   Sort by: E value   [Next](#)   [Previous](#)   [Descriptions](#)

**Heligmosomoides polygyrus genome assembly H\_bakeri\_Edinburgh, scaffold HPBE\_contig0002770**  
 Sequence ID: [LL206793.1](#) Length: 6449 Number of Matches: 3

Range 1: 3919 to 3972 [GenBank](#) [Graphics](#)   [Next Match](#)   [Previous Match](#)

Score	Expect	Identities	Positives	Gaps	Frame
30.4 bits(60)	5.2	9/18(50%)	14/18(77%)	0/18(0%)	+1/-3

Query 1 LHNIPFSSARFPQSEGPIS 54  
 ++N+PPS RP +GP+S  
 Sbjct 3972 MNNLPPSELRENVDPVPS 3919

รูปที่ 4.11 ตัวอย่างผลที่ได้จาก Translated BLAST: tBLASTx



## การ BLAST แบบพิเศษ (Specialized BLAST)

นอกจากการ BLAST แบบพื้นฐานที่ใช้ในการค้นหาชื่อยีน หรือชื่อโปรตีนของสิ่งมีชีวิต ดังที่กล่าวไปแล้วนั้น ยังมีการ BLAST แบบพิเศษมีหน้าที่จำเพาะในการใช้งานหลากหลายด้าน (รูปที่ 4.12) สรุปดังต่อไปนี้



รูปที่ 4.12 ตัวเลือกสำหรับการ BLAST แบบพิเศษ

[<https://blast.ncbi.nlm.nih.gov/Blast.cgi>]

- **Smart BLAST** ค้นหาโปรตีนที่สนใจ โดยใช้โปรแกรม BLASTp เปรียบเทียบกับโปรตีนในฐานข้อมูล Landmark ซึ่งประกอบด้วย proteomes จาก 27 จีโนมของสิ่งมีชีวิตที่ครอบคลุมทั่วทุกอนุกรมวิธาน ผลที่ได้คือโปรตีนที่ไม่ซ้ำซ้อนที่มีความหลากหลายทางอนุกรมวิธาน นอกจากนี้ SmartBLAST ยังเสนอฐานข้อมูล Conserved Domain ที่เกี่ยวข้องด้วย
- **Primer-BLAST** ใช้ในการออกแบบ primers ที่จำเพาะต่อยีนที่เราต้องการ โดยใช้ primer3 algorithm จากนั้นตรวจสอบความจำเพาะกับ DNA template โดยการ BLAST เปรียบเทียบกับลำดับนิวคลีโอไทด์ในฐานข้อมูล

- **Global Align** ใช้โปรแกรม Needleman-Wunsch Global Align ในการจัดเรียงลำดับนิวคลีโอไทด์ หรือลำดับกรดอะมิโน ที่เราใส่เข้าไปเป็นคู่ โดยจัดเรียงและเปรียบเทียบตั้งแต่ต้นจนจบ ซึ่งไม่คำนึงว่ามีความเหมือนกันระหว่างลำดับหรือไม่
- **CD-search** การค้นหาบริเวณอนุรักษ์ (conserved domains) จากลำดับกรดอะมิโนที่ใช้ในการเปรียบเทียบกับฐานข้อมูลโปรตีน การค้นหาบริเวณอนุรักษ์เพื่อการวิเคราะห์เชิงหน้าที่ (functional analysis) ของโปรตีน
- **GEO** เป็นการค้นหานิวคลีโอไทด์ที่สนใจโดยใช้โปรแกรม BLASTn เปรียบเทียบกับนิวคลีโอไทด์ในฐานข้อมูล GEO (Gene Expression Omnibus) ซึ่งจะทำให้ได้ข้อมูลรูปแบบการแสดงออกของยีน
- **IgBLAST** เป็นการค้นหาโมเลกุลของ immunoglobulin (Ig) หรือ T cell receptor (TR) จากข้อมูลลำดับกรดอะมิโนที่สนใจ โดยเปรียบเทียบกับฐานข้อมูล Germline
- **VecScreen** เป็นการคัดกรองลำดับนิวคลีโอไทด์ที่สนใจว่ามีการปนเปื้อนลำดับนิวคลีโอไทด์ที่มาจาก DNA vector หรือไม่ โดยการใส่โปรแกรม BLASTn ทำการเปรียบเทียบกับลำดับนิวคลีโอไทด์กับฐานข้อมูลที่รวบรวมลำดับนิวคลีโอไทด์ของ vectors ชนิดต่างๆ
- **CDART** การระบุบริเวณอนุรักษ์ (conserved domains) ที่มีอยู่โปรตีนที่สนใจ แล้วตามด้วยการค้นหาโปรตีนชนิดอื่นๆ ที่มีบริเวณอนุรักษ์ (conserved domains) ลักษณะเดียวกัน
- **Targeted Loci** เป็นการระบุชนิดของเชื้อจุลชีพที่พบในสิ่งแวดล้อม โดยใช้ยีนเป้าหมาย ได้แก่ 16S, 18S, 28S และ ITS (internal transcribed spacer region) ของ ribosomal RNA แล้วใช้โปรแกรม BLASTn ในการเปรียบเทียบกับฐานข้อมูลต่างๆ ดังนี้
  - 16S ribosomal RNA sequences สำหรับ bacteria และ archaea
  - 18S ribosomal RNA sequences สำหรับ eukaryotic microbes
  - 28S ribosomal RNA sequences สำหรับ fungi
  - ITS (internal transcribed spacer region) สำหรับ fungi



- **Multiple alignments** การใช้ Constraint Based Protein Multiple Alignment Tool (COBALT) ในการจัดเรียงและเปรียบเทียบลำดับกรดอะมิโนของโปรตีนหลายๆ ชนิดพร้อมๆ กัน
- **BioAssay** เป็นการใช้โปรแกรม BLASTn ในการค้นหาสารเคมีจาก PubChem ที่สามารถจับกับลำดับนิวคลีโอไทด์ที่สนใจได้ หรือใช้โปรแกรม BLASTp ในการค้นหาสารเคมีจาก PubChem ที่สามารถจับกับโปรตีนที่สนใจได้
- **MOLE-BLAST** การระบุลำดับนิวคลีโอไทด์ที่มีความเหมือนหรือใกล้เคียง (neighbor sequences) กับลำดับนิวคลีโอไทด์ที่สนใจ โดยใช้โปรแกรม BLASTn จากนั้นทำการจัดเรียงและเปรียบเทียบลักษณะความเหมือน เพื่อใช้ในการจำแนกเป็นสายวิวัฒนาการ (phylogenetic) ตามลำดับความคล้ายคลึงกัน

#### เอกสารอ้างอิง และแนะนำอ่านเพิ่มเติม

- BLAST Homepage and Selected Search Pages  
[ftp://ftp.ncbi.nih.gov/pub/factsheets/HowTo\_BLASTGuide.pdf]
- The New BLAST® Results Page  
[ftp://ftp.ncbi.nlm.nih.gov/pub/factsheets/HowTo\_NewBLAST.pdf]

# 5

การวิเคราะห์ความสัมพันธ์เชิงวิวัฒนาการ

Phylogenetic analysis

## การวิเคราะห์ความสัมพันธ์เชิงวิวัฒนาการ

### Phylogenetic analysis

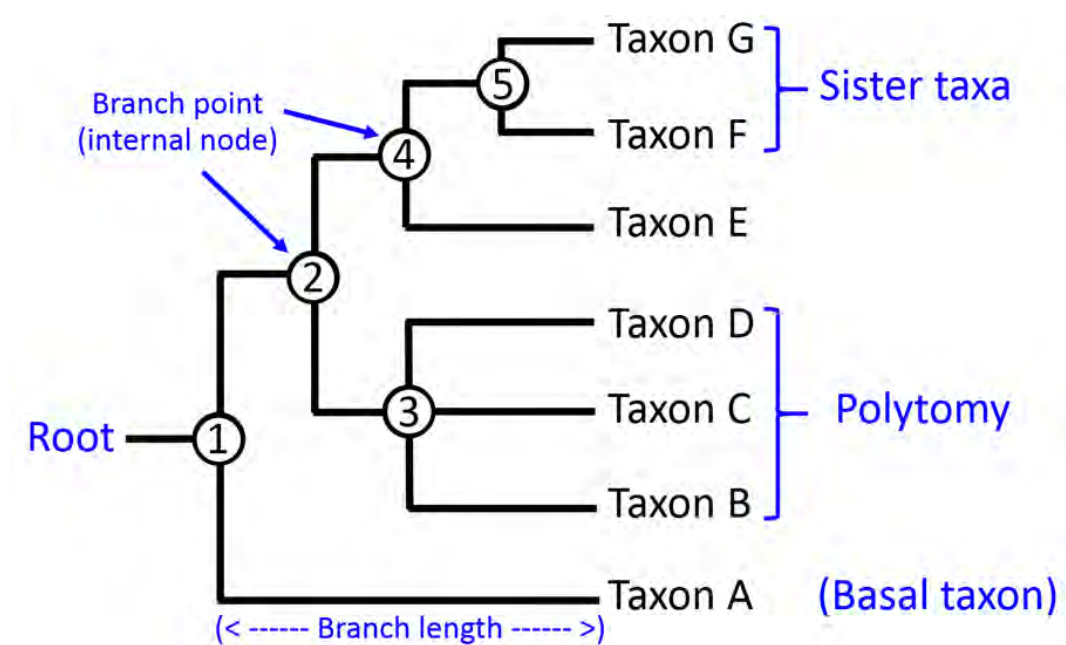
เชื้อไวรัสส่วนใหญ่มีอัตราการกลายพันธุ์ (mutation) สูง ทำให้มีการสะสมการกลายพันธุ์ระหว่างการเพิ่มจำนวน (replication) อย่างรวดเร็ว อันเนื่องมาจากหลายปัจจัย ได้แก่

- RNA virus อาศัยการทำงานของเอนไซม์ RNA-dependent RNA polymerase ในการเพิ่มจำนวนสารพันธุกรรม ซึ่งเอนไซม์ดังกล่าวมีอัตราการสร้างสารพันธุกรรมผิดพลาดค่อนข้างมาก ( $10^3$ - $10^4$  nucleotides/replication cycle)
- เชื้อไวรัสเพิ่มจำนวนภายในเซลล์เจ้าบ้าน (host cells) ซึ่งภายในเซลล์เจ้าบ้านจะมี RNA editing enzymes ที่สามารถปรับเปลี่ยนลำดับเบสบนสาย RNA ได้ ซึ่งมีรายงานวิจัยพบว่า การทำงานของเอนไซม์ดังกล่าว มีผลส่งเสริมให้เชื้อไวรัสมีการกลายพันธุ์มากขึ้น
- หลังจากเพิ่มจำนวนสารพันธุกรรมแล้ว ไวรัสไม่มีกลไก proof-reading เพื่อตรวจสอบความผิดพลาดของการสร้างสารพันธุกรรมของเชื้อไวรัส
- เชื้อไวรัสไม่มีกลไกการซ่อมแซม (repair mechanism) เพื่อแก้ไขรหัสพันธุกรรมที่ผิดพลาด
- Selective pressure เนื่องจากระบบภูมิคุ้มกันของร่างกาย มีการตอบสนองโดยต่อต้านและกำจัดเชื้อไวรัส เป็นผลทำให้เกิดภาวะกดดันให้เชื้อไวรัสมีอัตราการกลายพันธุ์อย่างรวดเร็ว เพื่อหลีกเลี่ยงการตอบสนองของระบบภูมิคุ้มกัน

ดังนั้นในการศึกษาหสพันธุกรรมหรืออนุชีววิทยาของเชื้อไวรัสชนิดต่างๆ จึงนิยมศึกษาเพื่อจำแนกสายพันธุ์ (genetic characterization) และวิเคราะห์ความสัมพันธ์เชิงวิวัฒนาการ (phylogenetic analysis) ของเชื้อไวรัส เพื่อให้เข้าใจลักษณะทางพันธุกรรมของเชื้อไวรัส ซึ่งมีประโยชน์ในด้านต่างๆ ได้แก่ การวิเคราะห์แหล่งที่มาเชื้อไวรัสว่าเริ่มต้นระบาดมาจากที่ใด การทำนายความรุนแรงในการก่อโรคของเชื้อไวรัส การเปรียบเทียบระหว่างสายพันธุ์ของวัคซีนและไวรัสที่กำลังระบาด และการทำนายสายพันธุ์ของเชื้อไวรัสที่กำลังจะระบาดในอนาคต เป็นต้น

### การวิเคราะห์ความสัมพันธ์เชิงวิวัฒนาการ (Phylogenetic analysis)

ในการศึกษาวิเคราะห์ความสัมพันธ์เชิงวิวัฒนาการนั้น จะแสดงผลเป็นแผนภาพต้นไม้เชิงวิวัฒนาการ (phylogenetic tree) ที่สะท้อนให้เห็นถึงสมมติฐาน (ไม่ใช่ข้อเท็จจริงที่แน่นอน) ของความสัมพันธ์เชิงวิวัฒนาการระหว่างสิ่งมีชีวิตจากบรรพบุรุษสืบเชื้อสายไปสู่ลูกหลาน โดยลักษณะกิ่งก้าน (branch) ของแผนภาพต้นไม้เชิงวิวัฒนาการจะบ่งชี้ให้เห็นว่าสิ่งมีชีวิตสายพันธุ์ใดมีความสัมพันธ์ใกล้ชิดกัน หรือมีบรรพบุรุษร่วมกัน (รูปที่ 5.1)



รูปที่ 5.1 ตัวอย่างแผนภาพต้นไม้เชิงวิวัฒนาการ (phylogenetic tree)

จากรูปที่ 5.1 มีคำศัพท์ที่ควรทราบเกี่ยวกับแผนภาพต้นไม้เชิงวิวัฒนาการ ดังนี้

- Root คือรากฐานหรือจุดเริ่มต้นของแผนภาพต้นไม้เชิงวิวัฒนาการ
- Taxon ใช้เรียกสิ่งมีชีวิตแต่ละตัวที่สนใจศึกษา ในกรณีที่ใช้เป็นพหูพจน์จะเรียกว่า Taxa
- Basal taxon ใช้เรียกสิ่งมีชีวิตที่คาดว่าจะ เป็นต้นกำเนิด หรือบรรพบุรุษเริ่มต้นของ taxa อื่นๆ
- Sister taxa ใช้เรียกสิ่งมีชีวิต 2 taxa ที่มีความสัมพันธ์เชิงวิวัฒนาการอย่างใกล้ชิด

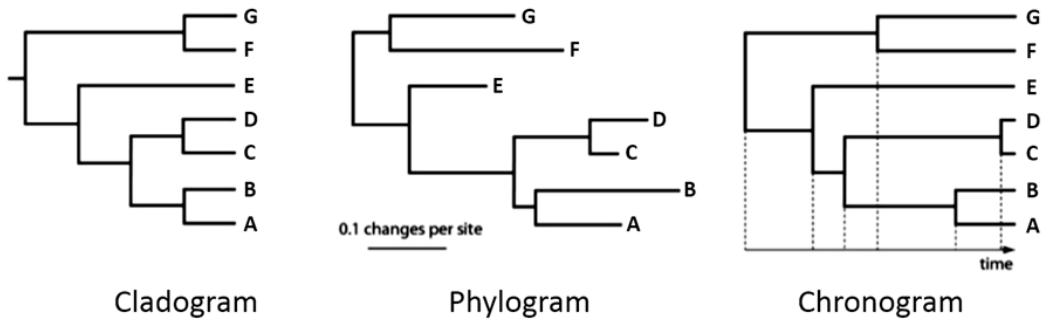
- Polytomy ใช้เรียกสิ่งมีชีวิตมากกว่า 2 taxa ที่มีความสัมพันธ์เชิงวิวัฒนาการอย่างใกล้ชิด แต่มีข้อมูลไม่เพียงพอในการจัดจำแนก taxa ออกเป็นกิ่งก้านตามลำดับ (branching order)
- Branch length ความยาวของกิ่ง ในกรณีของ Cladogram (รูปที่ 5.1) ความยาวของกิ่งจะไม่มี ความหมาย แต่ถ้าเป็น Phylogram ความยาวของกิ่งจะแสดงถึงจำนวนของการเปลี่ยนแปลงเชิงวิวัฒนาการ ซึ่งจะกล่าวรายละเอียดต่อไป
- Branch point จุดแตกกิ่ง เป็นการแยกความแตกต่างของ taxa ที่มีบรรพบุรุษร่วมกัน ออกจากกัน โดย taxa ที่มี branch point ร่วมกัน จะมีบรรพบุรุษล่าสุดร่วมกัน (most recent common ancestor) จึงมีความสัมพันธ์เชิงวิวัฒนาการอย่างใกล้ชิด ยกตัวอย่างเช่น
  - Branch point ① เป็นจุดที่มีบรรพบุรุษร่วมกันล่าสุดของ taxa A-G
  - Branch point ③ เป็นจุดที่มีบรรพบุรุษร่วมกันล่าสุดของ taxa B, C และ D
  - Branch point ⑤ เป็นจุดที่มีบรรพบุรุษร่วมกันล่าสุดของ taxa F และ G

### ประเภทของแผนภาพต้นไม้เชิงวิวัฒนาการ (Types of phylogenetic tree)

แผนภาพต้นไม้เชิงวิวัฒนาการ สามารถแบ่งเป็น 3 ประเภท (รูปที่ 5.2) ดังนี้

- Cladogram เป็นแผนภาพต้นไม้ที่เน้นแสดงเฉพาะรูปแบบของความสัมพันธ์เชิงวิวัฒนาการระหว่างกลุ่ม โดยจัดเรียงให้กิ่ง (branch) ของทุก taxa มาบรรจบที่ตำแหน่งเดียวกันเพื่อความสวยงามเท่านั้น ดังนั้นความยาวของกิ่ง (branch length) จะไม่มี ความหมายใดๆ สำหรับแผนภาพต้นไม้ประเภทนี้
- Phylogram เป็นแผนภาพต้นไม้ที่แสดงรูปแบบความสัมพันธ์เชิงวิวัฒนาการระหว่างกลุ่ม โดยให้ความยาวของกิ่ง (branch length) สະท้อนถึงจำนวนของการเปลี่ยนแปลงเชิงวิวัฒนาการด้วย ดังนั้นแผนภาพต้นไม้ประเภทนี้ จะมีความยาวของกิ่งแตกต่างกันไป ในแต่ taxa

- Chronogram เป็นแผนภาพต้นไม้ที่แสดงรูปแบบความสัมพันธ์เชิงวิวัฒนาการระหว่างกลุ่ม โดยให้ความยาวของกิ่ง (branch length) สະท้อนถึงเวลาที่เกิดขึ้นจริง เพื่อศึกษาความสัมพันธ์เชิงวิวัฒนาการเทียบกับช่วงเวลา



รูปที่ 5.2 ประเภทของแผนภาพต้นไม้เชิงวิวัฒนาการ

### วิธีการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ

ในอดีตการเขียนแผนภาพต้นไม้เชิงวิวัฒนาการจะอาศัยการเปรียบเทียบลักษณะ (traits) ภายนอกที่มองเห็น (phenotypes) เช่น มีหาง/ไม่มีหาง มีเขา/ไม่มีเขา เป็นต้น เพื่อใช้ในการจัดจำแนกความสัมพันธ์เชิงวิวัฒนาการของสิ่งมีชีวิต โดยมีหลักการว่าสิ่งมีชีวิตที่มีบรรพบุรุษร่วมกัน จะมีลักษณะคล้ายคลึงกัน แต่ในปัจจุบันนิยมใช้ลักษณะทางอณูชีววิทยา (molecular traits) ได้แก่ ลำดับนิวคลีโอไทด์ของยีน หรือลำดับกรดอะมิโนของโปรตีน ในการจัดจำแนกความสัมพันธ์เชิงวิวัฒนาการของสิ่งมีชีวิต โดยมีหลักการว่าสิ่งมีชีวิตที่มีวิวัฒนาการร่วมกันอย่างใกล้ชิด จะมีลักษณะทางพันธุกรรม (genotypes) ที่แตกต่างกัน (distance) เพียงเล็กน้อย ในขณะที่สิ่งมีชีวิตที่ไม่มี ความใกล้ชิดเชิงวิวัฒนาการ จะมีลักษณะทางพันธุกรรมแตกต่างกันมาก

การจัดจำแนกความสัมพันธ์เชิงวิวัฒนาการของสิ่งมีชีวิต โดยใช้ลักษณะทางอณูชีววิทยา ประกอบด้วยส่วนต่างๆ ดังนี้



## 1. การเลือกใช้ข้อมูลทางอนุชีววิทยา ในการจำแนกความสัมพันธ์เชิงวิวัฒนาการ ได้แก่

- ลำดับนิวคลีโอไทด์ส่วนที่เป็น non-coding regions ไม่ได้แปลรหัสเป็นกรดอะมิโน
- ลำดับนิวคลีโอไทด์ส่วนที่เป็น coding regions สามารถแปลรหัสเป็นกรดอะมิโนได้
- ลำดับกรดอะมิโนของโปรตีน สามารถใช้ลำดับกรดอะมิโนที่แปลรหัสมาจากลำดับนิวคลีโอไทด์ (translated nucleotide) ในส่วน coding regions ได้

งานวิจัยทางด้านไวรัสวิทยา นิยมเลือกใช้ลำดับนิวคลีโอไทด์ส่วนที่เป็น coding regions ของยีนที่ใช้ในการสร้างโปรตีนพื้นผิว (surface protein) ของอนุภาคไวรัส เนื่องจากยีนและโปรตีนดังกล่าวจะได้รับภาวะกดดันจากปัจจัยต่างๆ เช่น ระบบภูมิคุ้มกันของร่างกาย ยาต้านไวรัส และสิ่งแวดล้อม ทำให้มีอัตราการกลายพันธุ์มากกว่ายีนอื่นๆ นอกจากนี้การจำแนกสายพันธุ์ของเชื้อไวรัสบางชนิดในอดีตนิยมจำแนกด้วยการแบ่งเป็น serotypes โดยอาศัยการจับกันระหว่างแอนติบอดี (antibody) และแอนติเจน (antigen) ซึ่งมักจะเป็นโปรตีนพื้นผิว (surface protein) ของอนุภาคไวรัสนั่นเอง ดังนั้นการเลือกใช้ลำดับนิวคลีโอไทด์ส่วนที่เป็น coding regions ของยีนที่ใช้ในการสร้างโปรตีนพื้นผิว (surface protein) ของอนุภาคไวรัส มาใช้เป็นข้อมูลสำหรับการจำแนกความสัมพันธ์เชิงวิวัฒนาการ จะมีข้อดีหลายประการ ได้แก่

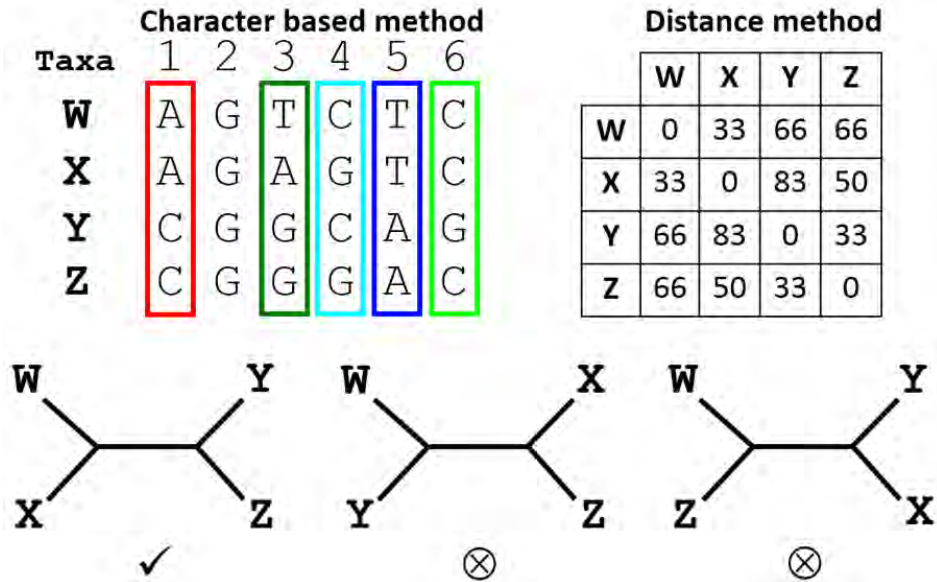
- ศึกษาความสัมพันธ์เชิงวิวัฒนาการที่สอดคล้องกับการจำแนกสายพันธุ์ด้วยระบบ serotypes
- ศึกษาความสัมพันธ์เชิงวิวัฒนาการที่สอดคล้องกับการปรับตัวของเชื้อไวรัส (viral adaptation) เพื่อให้อยู่รอด เช่น การหลบเลี่ยงระบบภูมิคุ้มกัน (immune escape) การดื้อต่อยาต้านไวรัส (anti-viral resistance) และ การคัดเลือกโดยธรรมชาติ (natural selection) เป็นต้น
- ลำดับนิวคลีโอไทด์ส่วนที่เป็น coding regions สามารถแปลรหัสไปเป็นกรดอะมิโนได้ ดังนั้นจึงสามารถใช้ทั้งลำดับนิวคลีโอไทด์ และกรดอะมิโนในการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการได้

## 2. การจัดเรียงลำดับข้อมูลร่วมกัน (Multiple sequences alignment)

ความยาวของลำดับนิวคลีโอไทด์ หรือความยาวของลำดับกรดอะมิโน จากแต่ละตัวอย่างที่ศึกษาควรจะต้องมีความยาวเท่ากัน ดังนั้นจึงควรปรับแต่งให้ตำแหน่งเริ่มต้นและตำแหน่งสิ้นสุดของลำดับนิวคลีโอไทด์ หรือลำดับกรดอะมิโน ในทุกๆ ตัวอย่างเป็นตำแหน่งที่ตรงกันเพื่อให้ได้ข้อมูลที่แม่นยำ ในกรณีที่เกิดการขาดหายไป (deletion) หรือมีส่วนที่เพิ่มขึ้น (insertion) ของลำดับนิวคลีโอไทด์หรือกรดอะมิโนในบางตัวอย่าง โปรแกรมสำหรับการทำ alignment จะเติม gap (-) ในส่วนที่ขาดหายไปให้อัตโนมัติ อย่างไรก็ตามควรตรวจสอบว่าการเติม gap (-) ของโปรแกรมถูกต้องเหมาะสมหรือไม่ โดยเฉพาะกรณีที่ต้องการแปลรหัสจากลำดับนิวคลีโอไทด์ไปเป็นลำดับกรดอะมิโน ตำแหน่ง gap (-) เหล่านี้ อาจทำให้เกิด frame shift ซึ่งจะทำให้ได้ข้อมูลลำดับกรดอะมิโนที่ผิดพลาด เมื่อนำไปสร้างแผนภาพต้นไม้เชิงวิวัฒนาการก็จะผิดพลาดด้วย

## 3. วิธีการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ (Phylogenetic method) แบ่งออกเป็น

- Character based method มีหลักการคือการเปรียบเทียบชนิดของนิวคลีโอไทด์ หรือชนิดของกรดอะมิโน ในแต่ละตำแหน่งของลำดับที่เกิดจาก multiple sequences alignment จากรูปที่ 5.3 เมื่อเปรียบเทียบลำดับนิวคลีโอไทด์จำนวน 6 ตำแหน่งระหว่างตัวอย่าง 4 กลุ่ม (taxa) จะพบว่าสามารถสร้าง unrooted tree ได้ 3 ลักษณะ จากนั้นเมื่อพิจารณาลำดับนิวคลีโอไทด์แต่ละตำแหน่ง จะพบว่า W และ X taxa ควรจัดเป็น sister taxa เนื่องจากมีความแตกต่างกันน้อย (แตกต่างในตำแหน่งที่ 3 และ 4) ส่วน Y และ Z taxa ควรจัดเป็น sister taxa เนื่องจากมีความแตกต่างกันน้อย (แตกต่างในตำแหน่งที่ 4 และ 6) ในขณะที่ W และ Y taxa (แตกต่างในตำแหน่งที่ 1, 3, 5 และ 6) หรือ W และ Z taxa (แตกต่างในตำแหน่งที่ 1, 3, 4 และ 5) ควรจัดอยู่ห่างกันออกไป ดังนั้นแผนภาพต้นไม้เชิงวิวัฒนาการที่น่าจะถูกต้องที่สุด คือ แผนภาพต้นไม้ด้านซ้ายมือในรูปที่ 5.3 ทั้งนี้หลักการ Character based method แบ่งออกเป็น 2 วิธี ได้แก่



รูปที่ 5.3 การสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ ด้วย character และ distance methods

- Maximum Parsimony เป็นวิธีในการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการหลายๆแบบ แล้วเลือกแสดงผลเฉพาะแผนภาพแบบที่มีจำนวนการกลายพันธุ์ (การเปลี่ยลลำดับเบส หรือลำดับกรดอะมิโน) น้อยที่สุด ซึ่งวิธีการนี้เหมาะสำหรับชุดข้อมูลที่มีความคล้ายคลึงกันมากๆ หากข้อมูลมีความหลายหลากและซับซ้อนมากๆ วิธีการนี้อาจจะสร้างแผนภาพต้นไม้เชิงวิวัฒนาการที่ผิดพลาดได้ นอกจากนี้ข้อจำกัดของวิธีนี้คือจะสามารถสร้างแผนภูมิต้นไม้ประเภท Cladogram ซึ่งเป็นแผนภาพต้นไม้ที่เน้นแสดงเฉพาะรูปแบบของความสัมพันธ์เชิงวิวัฒนาการเท่านั้น ไม่แสดงจำนวนของการเปลี่ยนแปลงเชิงวิวัฒนาการ
- Maximum Likelihood เป็นวิธีในการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการหลายๆแบบ แล้วเลือกแสดงผลเฉพาะแผนภาพแบบที่ข้อมูลมีความน่าจะเป็นมากที่สุด ซึ่งวิธีการนี้ใช้ข้อมูลในการประมวลผลสูงสุด ทำให้มีประสิทธิภาพและมีความน่าเชื่อถืออย่างมากในการจำแนกความสัมพันธ์เชิงวิวัฒนาการ จึงเป็นที่นิยมใช้ในการวิเคราะห์ข้อมูลงานวิจัยระดับนานาชาติ อย่างไรก็ตามข้อจำกัดของวิธีนี้ คือใช้เวลาในการประมวลผลค่อนข้างมาก และจำเป็นต้องใช้คอมพิวเตอร์ที่มีประสิทธิภาพสูง

- Distance method มีหลักการคือการเปรียบเทียบชนิดของนิวคลีโอไทด์ หรือชนิดของกรดอะมิโน โดยการนำ pair-wise sequences alignment จากนั้นคำนวณค่าความแตกต่าง (distance) ของแต่ละคู่ taxa แล้วจึงนำค่าความแตกต่างที่ได้นั้นมาสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ โดยจะจัด taxa ที่มีความแตกต่างกันเล็กน้อย ให้อยู่ใกล้ชิดกัน (sister taxa) ส่วน taxa ที่มีความแตกต่างกันมาก จะถูกจัดให้อยู่ห่างกัน จากรูปที่ 5.3 เมื่อเปรียบเทียบลำดับนิวคลีโอไทด์จำนวน 6 ตำแหน่ง ระหว่างตัวอย่าง 4 กลุ่ม (taxa) จะพบว่าสามารถสร้าง unrooted tree ได้ 3 ลักษณะ จากนั้นเมื่อพิจารณาค่าความแตกต่างกันระหว่าง taxa แต่ละคู่ แล้วสร้างเป็น distance matrix แล้วนำค่า distance ไปสร้างแผนภาพต้นไม้ จะพบว่า W และ X taxa ควรจัดเป็น sister taxa เนื่องจากมีความแตกต่างกันน้อย (distance = 33) ส่วน Y และ Z taxa ควรจัดเป็น sister taxa เนื่องจากมีความแตกต่างกันน้อย (distance = 33) ในขณะที่ W และ Y taxa (distance = 66) หรือ W และ Z taxa (distance = 66) ควรจัดอยู่ห่างกันออกไป ดังนั้นแผนภาพต้นไม้เชิงวิวัฒนาการที่น่าจะถูกต้องที่สุด คือ แผนภาพต้นไม้ด้านซ้ายมือในรูปที่ 5.3 โดยทั่วไป distance method มีข้อดีคือสามารถแผนภาพต้นไม้ได้อย่างรวดเร็ว จึงนิยมใช้ในการสร้างแผนภาพต้นไม้คร่าวๆ เพื่อใช้พิจารณาเบื้องต้นเท่านั้น ไม่นิยมใช้ในการวิเคราะห์ข้อมูลที่มีความหลากหลายและซับซ้อนมากๆ ทั้งนี้หลักการ distance method แบ่งออกเป็น 3 วิธี ได้แก่
  - UPGMA (Unweighted Pair Group Method with Arithmetic Mean) เป็นวิธีในการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ โดยใช้ข้อมูล distance มาจัดกลุ่ม taxa เป็นคู่ๆ ที่มีความใกล้ชิดกัน แล้วคำนวณหาค่าเฉลี่ย distance ในแต่ละคู่ taxa เพื่อใช้ในการกำหนด branch length ในแผนภาพต้นไม้
  - Neighbor-joining เป็นวิธีในการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ โดยใช้ข้อมูล distance มาจัดกลุ่มและสร้างแผนภาพอย่างรวดเร็ว

- Minimum evolution เป็นวิธีในการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ โดยใช้ข้อมูล distance มากำหนดความยาวของกิ่ง (branch length) จากนั้นคู่ taxa ที่มี branch length สั้นที่สุด (สัมพันธ์ใกล้เคียงกัน) จะถูกจำแนกเป็น sister taxa

#### 4. การกำหนดรูปแบบการกลายพันธุ์ (mutational models)

รูปแบบการกลายพันธุ์มีส่วนเกี่ยวข้องในการวิเคราะห์ความสัมพันธ์เชิงวิวัฒนาการ ดังนี้

- Jukes-Cantor (JC): การกลายพันธุ์ทุกๆ แบบมีโอกาสเกิดได้เท่าเทียมกัน
- Kimura 2-parameter (K2P): การกลายพันธุ์แบบ transitions ( $A \leftrightarrow G$  หรือ  $C \leftrightarrow T$ ) มีโอกาสเกิดได้มากกว่า transversions (purine  $\leftrightarrow$  pyrimidine)
- Felsenstein 84 (F84): การกลายพันธุ์แบบ transitions ( $A \leftrightarrow G$  หรือ  $C \leftrightarrow T$ ) มีโอกาสเกิดได้มากกว่า transversions (purine  $\leftrightarrow$  pyrimidine) และพิจารณา unequal base frequencies เพิ่มเติม
- Generalized Time Reversible (GTR): การกลายพันธุ์ไปตามกาลเวลา เป็นรูปแบบที่นิยมใช้ทั่วไป

#### 5. การวิเคราะห์ความเชื่อมั่นทางสถิติ (Analysis of statistical confidence)

เนื่องจากการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ เป็นการแสดงให้เห็นถึงสมมติฐาน (ไม่ใช่ข้อเท็จจริงที่แน่นอน) ดังนั้นจึงมีความเป็นไปได้ที่จะสร้างแผนภาพได้หลายๆ รูปแบบ ดังแสดงในตารางที่ 5.1 จะเห็นได้ว่าจำนวนแผนภาพต้นไม้เชิงวิวัฒนาการที่เป็นไปได้นั้นจะเพิ่มขึ้นตามจำนวนกลุ่ม (taxa) ที่ใช้ในการศึกษา ซึ่งถ้าข้อมูลมีความหลากหลายมากๆ จะทำให้มีความซับซ้อนในการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ

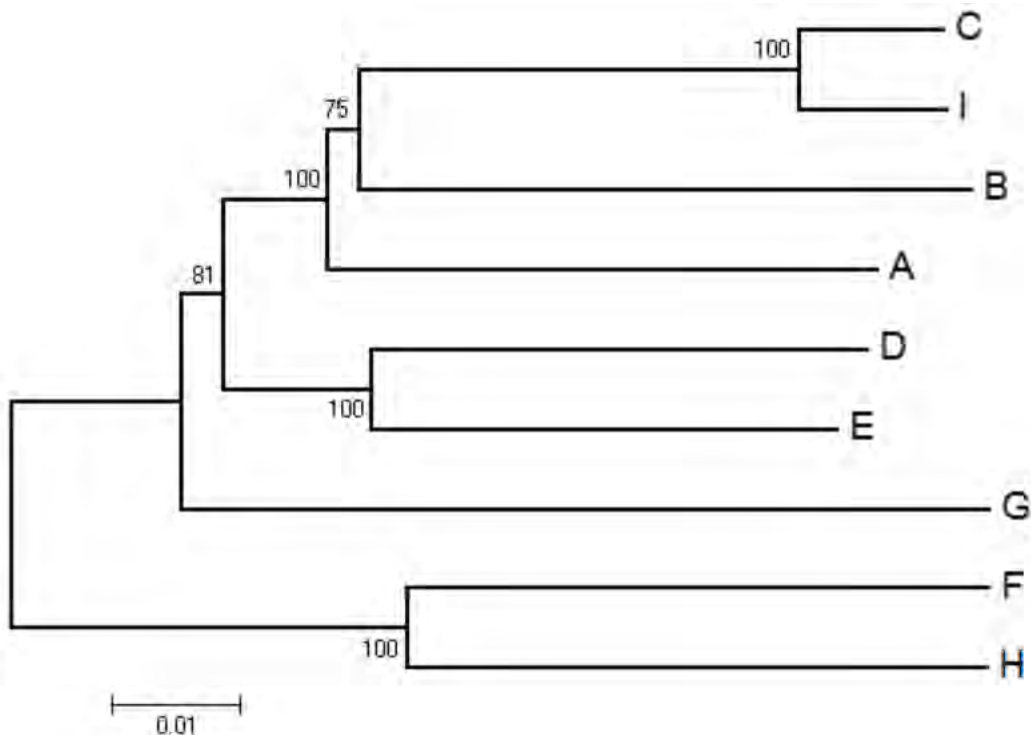
ตารางที่ 5.1 จำนวนรูปแบบที่เป็นไปได้ในการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ

Number of taxa	Possible number of trees	
	Unrooted tree	Rooted tree
3	1	3
4	3	15
5	15	105
6	105	945
7	945	10,395
8	10,395	135,135
9	135,135	2,027,025
10	2,027,025	34,459,425

ดังนั้นการวิเคราะห์ความเชื่อมั่นทางสถิติ สำหรับการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ จะอาศัยวิธีที่เรียกว่า bootstrap analysis ซึ่งมีหลักการในการจำลองการสร้างแผนภาพต้นไม้ จากข้อมูลชุดเดียวกัน (resample) โดยการทำซ้ำๆ (bootstrap replicates) จำนวนหลายๆ ครั้ง จากนั้นจึงนำแผนภาพต้นไม้ที่ได้จากการจำลองซ้ำๆ มาทำการเปรียบเทียบกัน หากพบลักษณะการจำแนกกิ่งแบบเดิมซ้ำๆ กันบ่อยๆ แสดงว่าการจำแนกรูปแบบดังกล่าวมีความเชื่อมั่นสูง โดยทั่วไปการกำหนด bootstrap replicates จะกำหนดค่าที่ 1,000 replicates (จำลองการสร้างแผนภาพต้นไม้ซ้ำๆ กันจำนวน 1,000 ครั้ง) อย่างไรก็ตามสามารถปรับเปลี่ยนค่า bootstrap replicates ได้ซึ่งถ้ากำหนดค่ามากๆ (>1,000) จะมีความถูกต้องสูงขึ้น แต่จะใช้เวลาในการวิเคราะห์ข้อมูลนาน ในทางกลับกันถ้ากำหนดค่าน้อยๆ จะวิเคราะห์ผลได้รวดเร็ว แต่ความน่าเชื่อถือจะลดลง หลังจากการประมวลผล bootstrap แล้วจะแสดงค่าเป็นเปอร์เซ็นต์ความน่าจะเป็นที่จะพบการจำแนกลักษณะกิ่ง ซึ่งจะระบุค่าความเชื่อมั่น (bootstrap support values) ไว้บริเวณจุดแตกกิ่ง (branch point หรือ node) หากระบุตัวเลขในช่วง 90-100 แสดงว่ามีความเชื่อมั่นทางสถิติสูงมาก (90-100%) ในการจัดจำแนกลักษณะกิ่งดังกล่าว



ตัวอย่างเช่นจากรูปที่ 5.4 จะเห็นได้ว่าการจำแนกกิ่งระหว่าง C และ I taxa หรือ D และ E หรือ F และ H จะมีความเชื่อมั่นเท่ากับ 100 แสดงว่าเชื่อถือได้แน่นอน ในขณะที่จุดแตกกิ่งบางตำแหน่งจะมีค่าน้อยกว่า 100 แสดงว่ามีความเชื่อมั่นทางสถิติลดลง ซึ่งหากมีค่าน้อยกว่า 60 จะประเมินว่าการจำแนกลักษณะกิ่งดังกล่าวไม่น่าเชื่อถือ



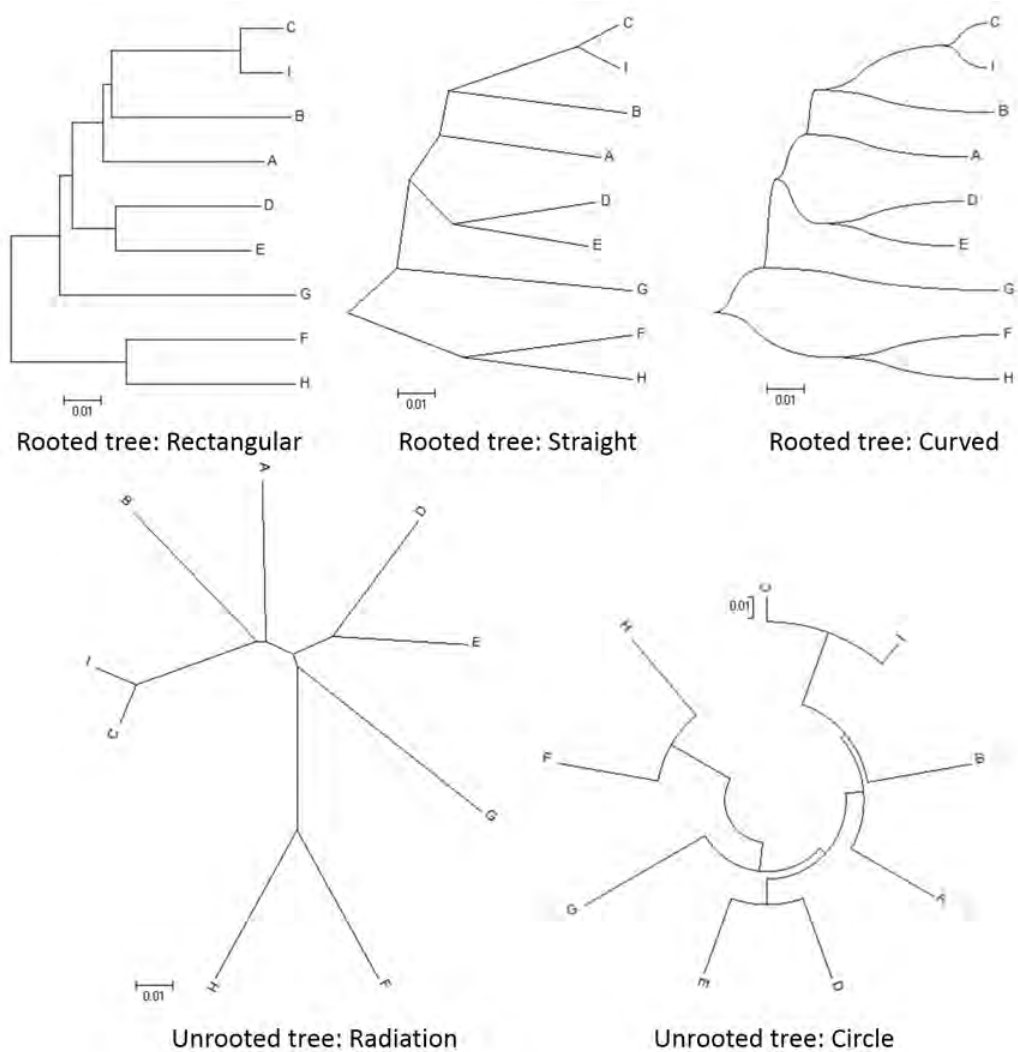
รูปที่ 5.4 การระบุค่า bootstrap support values ในแผนภาพต้นไม้เชิงวิวัฒนาการ

## 6. การจัดรูปแบบของแผนภาพต้นไม้เชิงวิวัฒนาการ (Styles of phylogenetic tree)

แผนภาพต้นไม้เชิงวิวัฒนาการ สามารถจำแนกได้หลายรูปแบบ ซึ่งสื่อถึงความสัมพันธ์เชิงวิวัฒนาการเหมือนกัน (รูปที่ 5.5) ดังนี้

- Rooted tree เป็นแผนภาพต้นไม้ที่มีรากฐานหรือจุดเริ่มต้นของวิวัฒนาการ ได้แก่
  - Rectangular ลักษณะการแตกกิ่งจะเป็นบล็อกเหมือนสี่เหลี่ยมผืนผ้า
  - Straight ลักษณะการแตกกิ่งจะเป็นเส้นทแยงมุม และรูปแบบกิ่งเป็นเส้นตรง
  - Curved ลักษณะการแตกกิ่งและรูปแบบกิ่งจะเป็นเส้นโค้ง

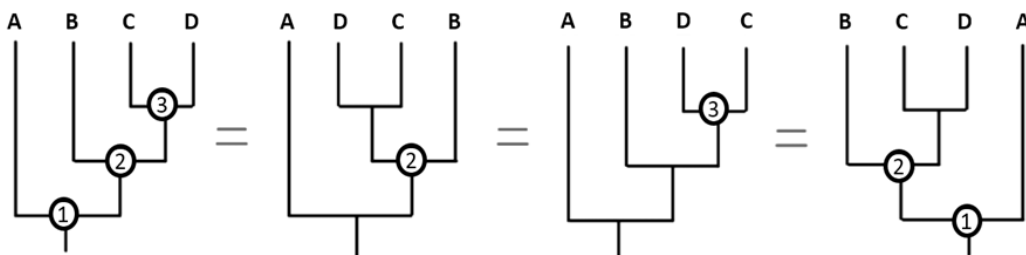
- Unrooted tree เป็นแผนภาพต้นไม้ที่ไม่มีรากฐานหรือจุดเริ่มต้นของวิวัฒนาการ ได้แก่
  - Radiation ลักษณะการแตกกิ่งจะเป็นเส้นตรงกระจายตัวจากจุดกึ่งกลาง
  - Circle ลักษณะการแตกกิ่งจะเป็นวงกลมกระจายตัวจากจุดกึ่งกลาง



รูปที่ 5.5 รูปแบบต่างๆ ของแผนภาพต้นไม้เชิงวิวัฒนาการ

### 7. การปรับหมุนที่ตำแหน่ง branch point ของแผนภาพต้นไม้เชิงวิวัฒนาการ

เนื่องจากแผนภาพต้นไม้แสดงความสัมพันธ์เชิงวิวัฒนาการ โดย taxa ที่มี branch point ร่วมกัน จะมีบรรพบุรุษล่าสุดร่วมกัน (most recent common ancestor) จึงมีความสัมพันธ์เชิงวิวัฒนาการอย่างใกล้ชิด ดังนั้นการปรับหมุนที่ตำแหน่ง branch point จะได้แผนภาพต้นไม้ที่มีรูปร่างเปลี่ยนแปลงไปเล็กน้อย อย่างไรก็ตามแผนภาพต้นไม้ดังกล่าวยังคงสะท้อนถึงความสัมพันธ์เชิงวิวัฒนาการดั้งเดิม ดังรูปที่ 5.6 จะเห็นได้ว่าไม่ว่าจะปรับหมุนที่ branch point ①, ② หรือ ③ ก็ยังคงความหมายเหมือนเดิม คือ A เป็น basal taxon ส่วน C และ D เป็น sister taxa

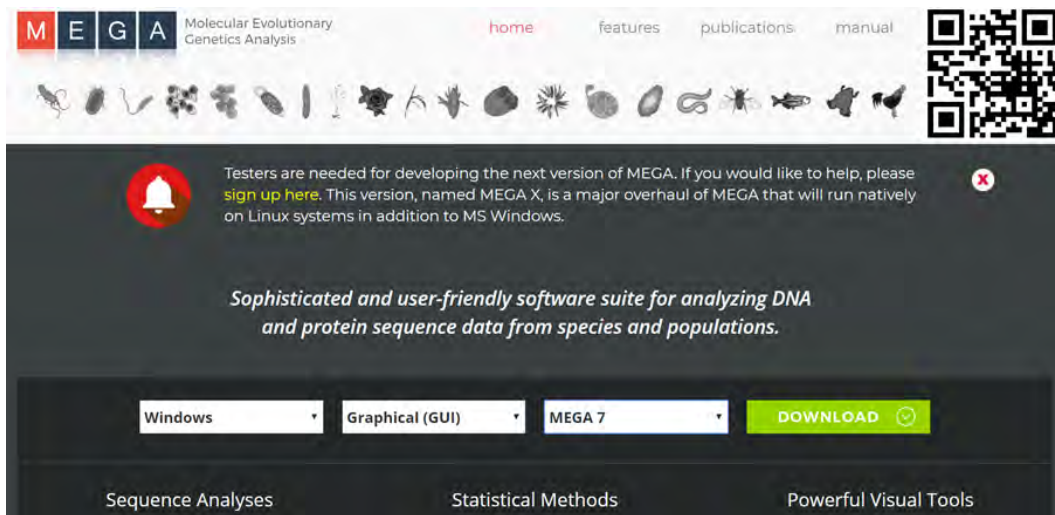


รูปที่ 5.6 ผลของการปรับหมุนที่ตำแหน่ง branch point ของแผนภาพต้นไม้เชิงวิวัฒนาการ

### การใช้โปรแกรมทางคอมพิวเตอร์ในการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ

ในปัจจุบันมีโปรแกรมสำหรับการวิเคราะห์ด้าน phylogenetic tree จำนวนหลายโปรแกรม ในที่นี้จะขอกล่าวถึงเฉพาะโปรแกรม MEGA (Molecular Evolutionary Genetics Analysis) version 7.0 เนื่องจากเป็นโปรแกรมที่สามารถใช้งานได้ฟรีโดยไม่เสียค่าใช้จ่าย (freeware) มีรูปแบบการใช้งานง่ายไม่ซับซ้อน เป็นโปรแกรมที่มีมาตรฐานเป็นที่ยอมรับสำหรับงานวิจัยในระดับนานาชาติ และผู้เขียนมีประสบการณ์ในการใช้งานโปรแกรมนี้มาหลายปี จึงขอสรุปขั้นตอนการใช้งานของโปรแกรม MEGA ดังนี้

1. การ download โปรแกรม MEGA ได้จาก <https://www.megasoftware.net/> (รูปที่ 5.7) โดยสามารถเลือกระบบปฏิบัติการที่ต้องการ ได้แก่ Windows, Mac OS X, Debian, RedHat และ Linux อื่นๆ นอกจากนี้ยังสามารถเลือกระบบ interface เป็นแบบ Graphical (GUI) หรือ Command Line (CC) ซึ่งในที่นี้จะถ่ายทอดวิธีการใช้โปรแกรม MEGA โดยใช้ระบบปฏิบัติการ Windows และใช้ interface เป็นแบบ Graphical (GUI) เมื่อ download เสร็จแล้ว ก็ดำเนินการติดตั้งโปรแกรมตามปกติ



รูปที่ 5.7 การ download โปรแกรม MEGA [<https://www.megasoftware.net/>]

2. การเตรียม file ที่เป็น FASTA format (นามสกุล .fas) ซึ่งประกอบไปด้วยลำดับนิวคลีโอไทด์ หรือลำดับกรดอะมิโนของตัวอย่างทั้งหมดที่ต้องการวิเคราะห์ ดังรูปที่ 5.8

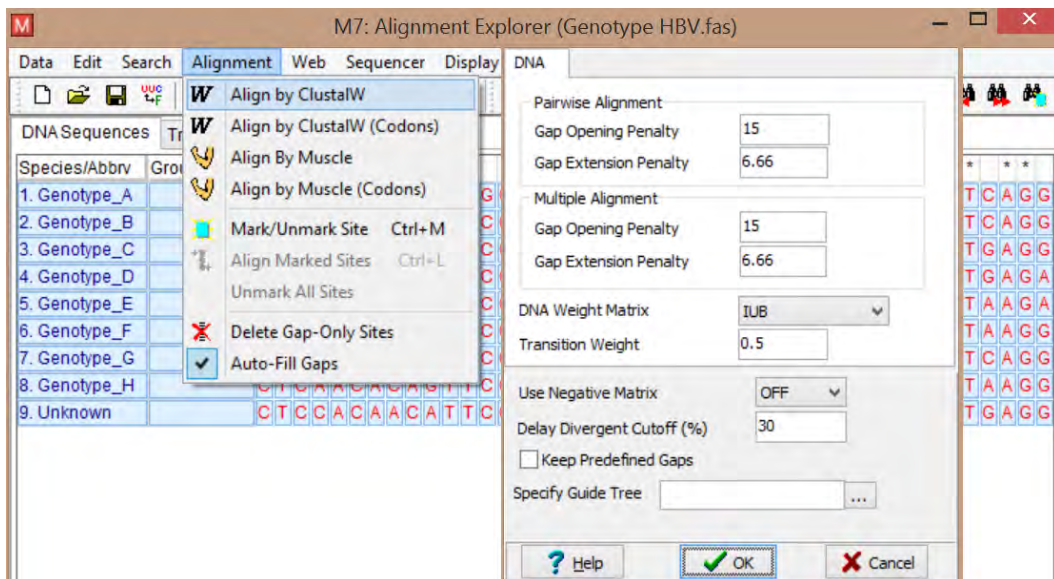
```
>Genotype_A
TTCCACTGCC~TTGCACCAAGCTCTGCAGGATCCCAGAGTCAGGGGTCTGTATC
CATTTACACAATGTGGATATCCTGCC~TTAATGCC~TTGTATGCATGTATACAA
CTCACCATACTGCACTCAGGCAAGCCATTCTCTGCTGGGGGAATTGATGACT
GCTCAGGGCATATTGACCACAGTGTCAACAATTCCTCCTCCTGCCTCCACCAA

>Genotype_B
CTCCACCAC~TTCCACAACTCTTCAAGATCCCAGAGTCAGGGCCTGTACT
CTTTCACGCAATGTGGATATCCTGCTT~TAATGCC~TTATATGCATGTATACAA
CTCACCATACGGCACTCAGGCAAGCTATTCTGTGTGGGGT~GAGTTGATGAAT
GGGTGGAGCCCTCACGCTCAGGGCTACTCACAAC~TGTGCCAGCAGCTCCTCC
```

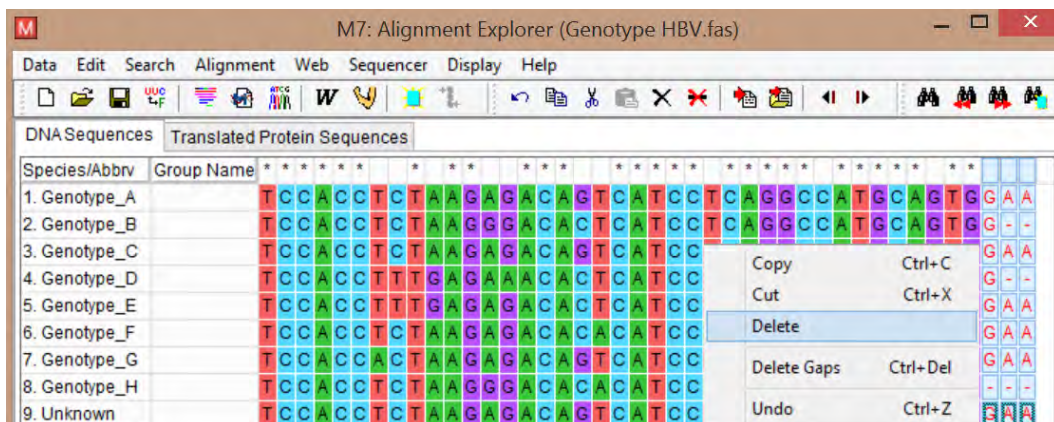
รูปที่ 5.8 ตัวอย่างลำดับนิวคลีโอไทด์ในรูปแบบ FASTA format



4. ทำการจัดเรียง (Alignment) ลำดับนิวคลีโอไทด์ หรือลำดับกรดอะมิโน โดยเลือกที่คำสั่ง Alignment และเลือกที่คำว่า Align by ClustalW ดังรูปที่ 5.11 จากนั้นจะมีหน้าต่างขึ้นมาให้ปรับเปลี่ยนค่า parameter ในการวิเคราะห์ ซึ่งไม่จำเป็นต้องปรับเปลี่ยนสามารถใช้ค่าเริ่มต้น (default) ต่างๆ ของโปรแกรมได้ โดยกดที่ปุ่ม OK เมื่อดำเนินการ alignment เสร็จแล้ว โปรแกรมจะแสดงผลข้อมูลดังรูปที่ 5.12



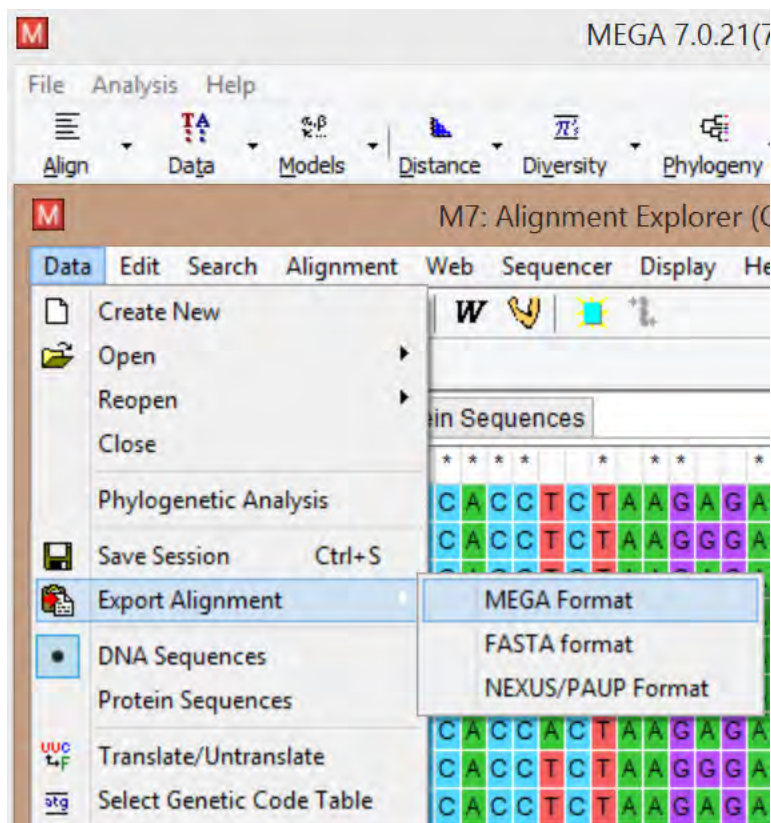
รูปที่ 5.11 การเลือกคำสั่ง Alignment ในโปรแกรม MEGA



รูปที่ 5.12 ผลจากการ Alignment โดยโปรแกรม MEGA

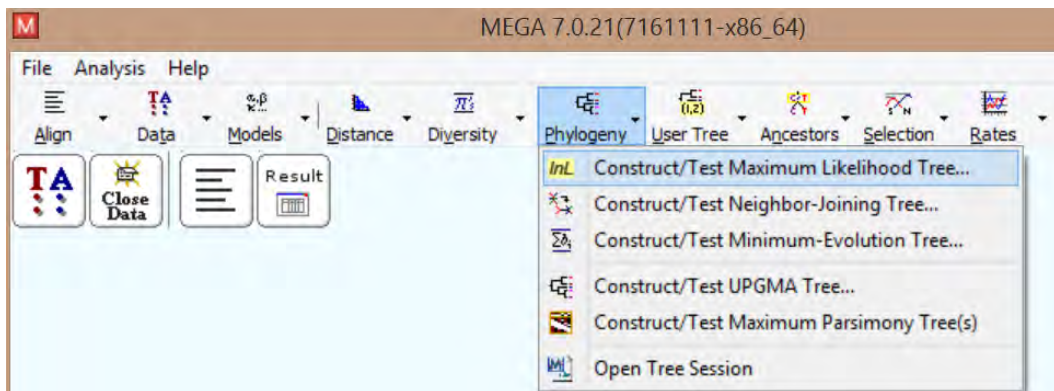


5. จากผลการ alignment ให้ตรวจสอบเบื้องต้นก่อนว่าลำดับนิวคลีโอไทด์ หรือลำดับกรดอะมิโน ของทุกตัวอย่างที่จะใช้ในการวิเคราะห์ มีจุดเริ่มต้นและจุดสิ้นสุดตรงกันหรือไม่ จากรูปที่ 5.12 จะเห็นได้ว่าช่วงปลายของลำดับนิวคลีโอไทด์มีจุดสิ้นสุดต่างกัน ดังนั้นจึงควรปรับแต่งให้ตำแหน่งเริ่มต้นและตำแหน่งสิ้นสุดของลำดับนิวคลีโอไทด์ หรือลำดับกรดอะมิโน ในทุกๆ ตัวอย่างเป็นตำแหน่งที่ตรงกันเพื่อให้ได้ข้อมูลที่แม่นยำ โดยเลือกแถบบริเวณลำดับนิวคลีโอไทด์ที่ต้องการตัดออก จากนั้นคลิกขวาแล้วเลือกคำสั่ง delete
6. เมื่อปรับแต่งความยาวของลำดับที่ได้จากการ alignment แล้ว ขั้นตอนต่อไปคือการ export ข้อมูล โดยเลือกที่คำสั่ง Data แล้วเลือกที่ Export Alignment จากนั้นเลือกที่ MEGA format ดังรูปที่ 5.13 ซึ่งจะเป็นการแปลงข้อมูลให้เป็นนามสกุล (.meg)



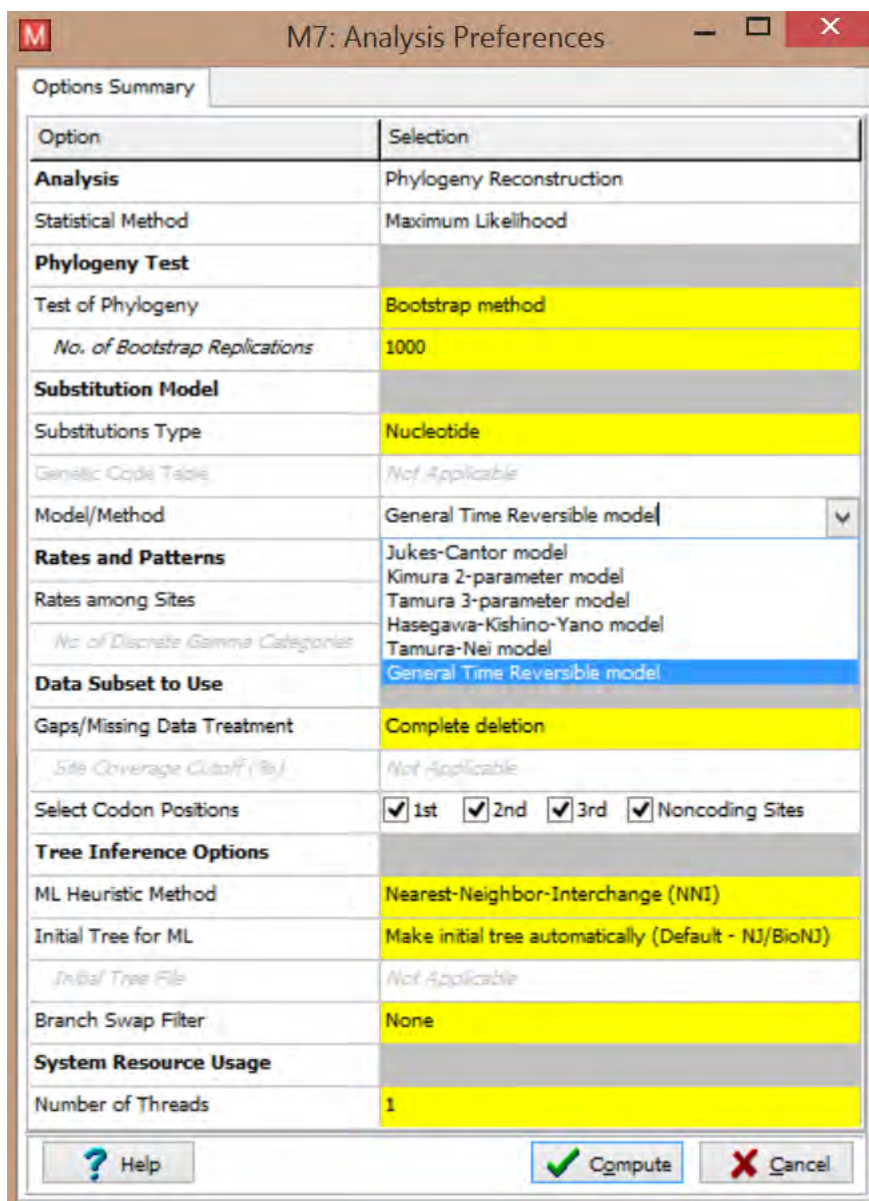
รูปที่ 5.13 การแปลงข้อมูล alignment ให้เป็น MEGA format

7. สร้างแผนภาพต้นไม้เชิงวิวัฒนาการ โดยเลือกที่คำสั่ง Phylogeny (รูปที่ 5.14) จะมีวิธีที่จะใช้ในการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ (Phylogenetic method) ต่างๆ ได้แก่
- Construct/Test Maximum Likelihood Tree
  - Construct/Test Neighbor-Joining Tree
  - Construct/Test Minimum-Evolution Tree
  - Construct/Test UPGMA Tree
  - Construct/Test Maximum Parsimony Tree



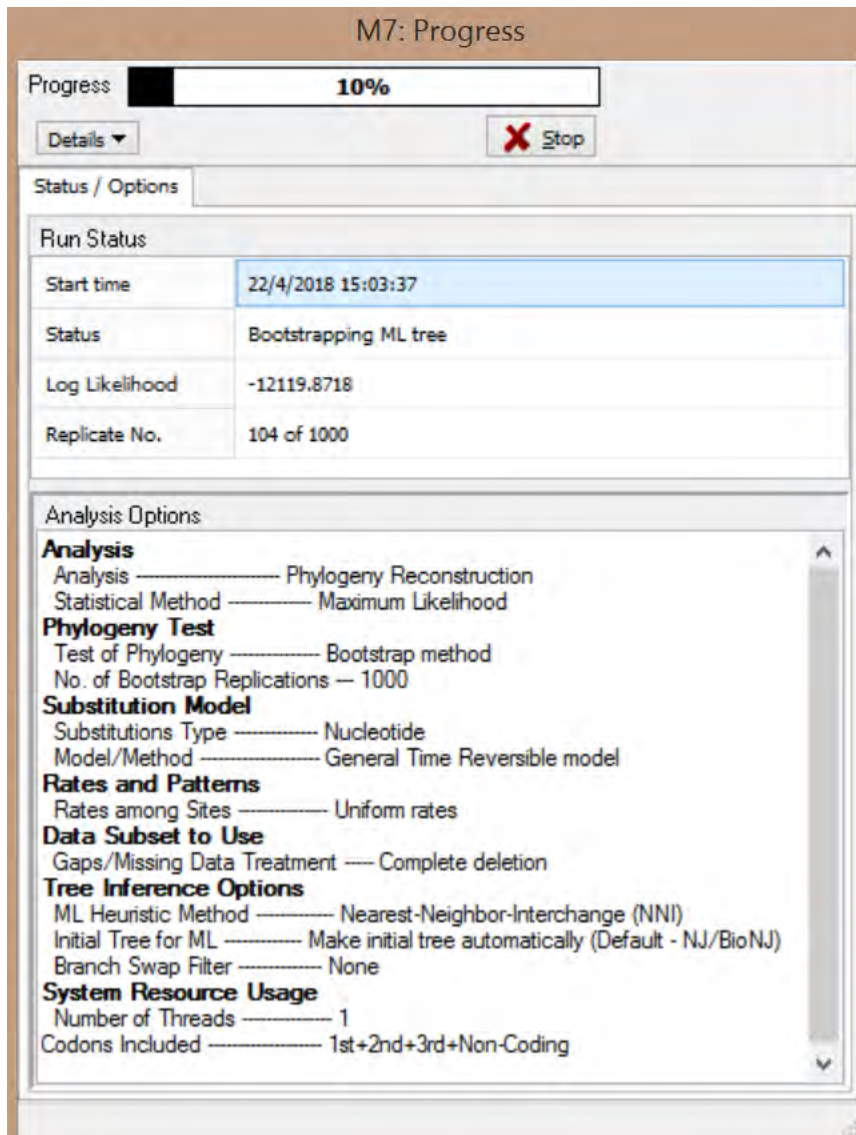
รูปที่ 5.14 การสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ (Phylogenetic method) ด้วยวิธี ต่างๆ

8. เมื่อเลือกวิธีที่จะใช้ในการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการแล้ว (ในที่นี้แสดงการใช้วิธี Maximum Likelihood) จะมีหน้าต่างปรากฏให้ปรับเปลี่ยนค่าต่างๆ (รูปที่ 5.15) ดังนี้
- Test of Phylogeny: เลือกว่าจะใช้ Bootstrap method หรือไม่
  - No. of Bootstrap replications: ใส่ค่าตัวเลขที่ต้องการ (โดยทั่วไปใช้ค่า 1000)
  - Substitution site: เลือกระหว่าง Nucleotide หรือ Amino acid ขึ้นกับข้อมูลที่ใช้
  - Model/Method: การกำหนดรูปแบบการกลายพันธุ์ โดยทั่วไปจะเลือกใช้ General Time Reversible model สำหรับข้อมูลลำดับนิวคลีโอไทด์ และ ใช้ Jones-Taylor-Thornton (JTT) model สำหรับข้อมูลลำดับกรดอะมิโน เป็นต้น อย่างไรก็ตามสามารถปรับเปลี่ยนได้ตามความเหมาะสม



รูปที่ 5.15 การตั้งค่า parameter ต่างๆ ในการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ

- จากรูปที่ 5.15 เมื่อคลิกที่คำว่า Compute โปรแกรมจะทำการคำนวณ ซึ่งอาจจะต้องใช้เวลานานในการประมวลผล ขึ้นอยู่กับความซับซ้อนของข้อมูล จำนวนตัวอย่างทั้งหมด วิธีที่ใช้ในการสร้างแผนภาพ และประสิทธิภาพของคอมพิวเตอร์ที่ใช้ (รูปที่ 5.16)

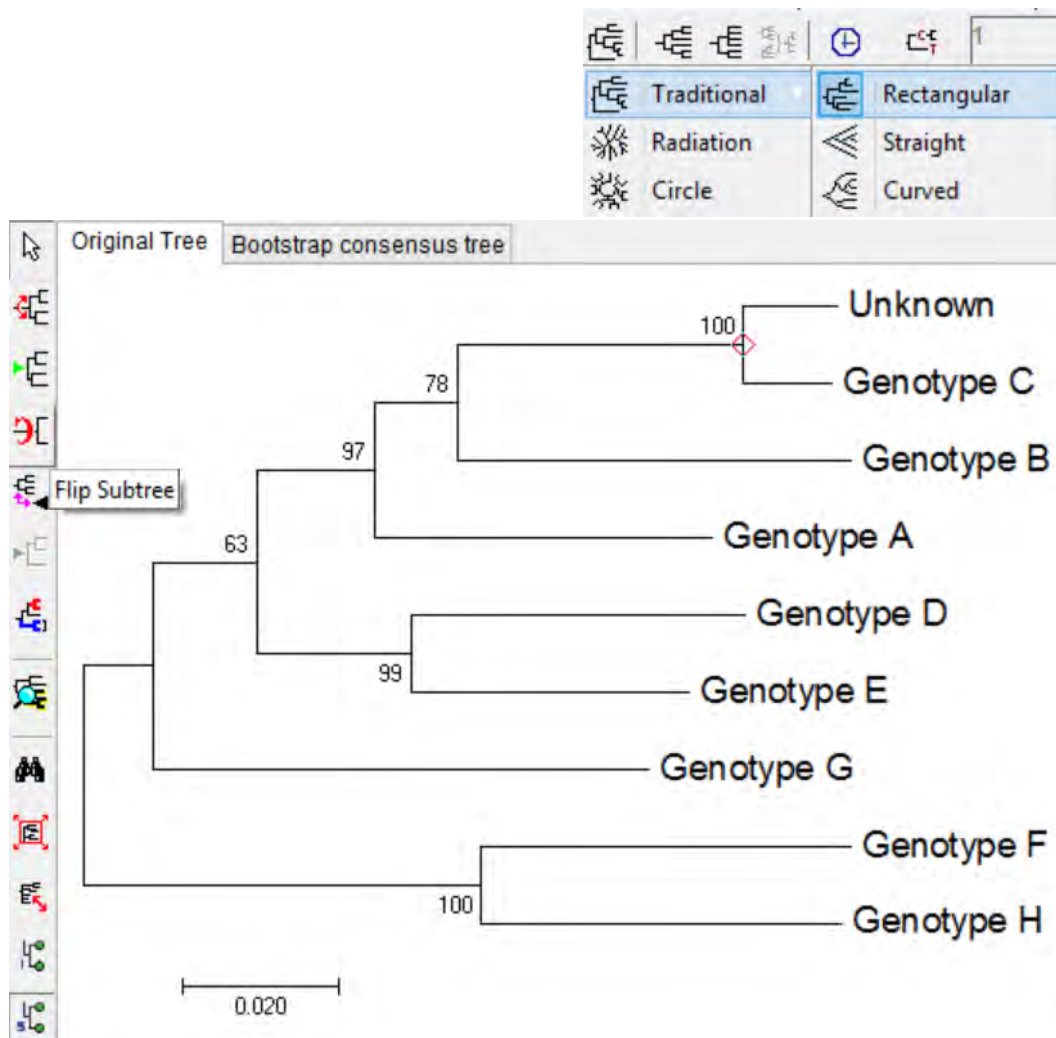


รูปที่ 5.16 การประมวลผลของโปรแกรม MEGA สำหรับการสร้างแผนภาพต้นไม้เชิงวิวัฒนาการ

10. เมื่อประมวลผลเสร็จ โปรแกรมจะแสดงแผนภาพต้นไม้เชิงวิวัฒนาการ ดังรูปที่ 5.17 ซึ่งสามารถปรับเปลี่ยนได้ เช่น

- การปรับหมุนที่ตำแหน่ง branch point โดยคลิกที่ปุ่ม Flip Subtree
- การจัดรูปแบบ (style) ของแผนภาพต้นไม้เชิงวิวัฒนาการ

- Traditional: Rectangular (rooted tree)
- Traditional: Straight (rooted tree)
- Traditional: Curved (rooted tree)
- Radiation (unrooted tree)
- Circle (unrooted tree)

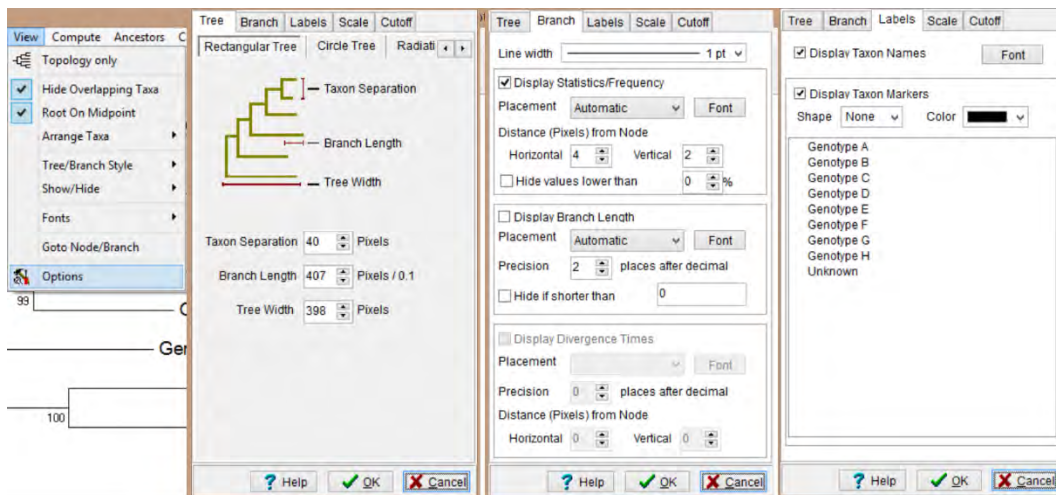


รูปที่ 5.17 การสร้างแผนภาพต้นไม้เชิงวิวัฒนาการด้วยวิธี Maximum Likelihood method



11. นอกจากนี้ยังสามารถปรับเปลี่ยนลักษณะต่างๆ ของแผนภาพต้นไม้เชิงวิวัฒนาการได้ โดยเลือกที่คำสั่ง View แล้วเลือก Options จะมีหน้าต่างปรากฏขึ้น (รูปที่ 5.18) ดังนี้

- Tree ปรับเปลี่ยนขนาดของแผนภาพต้นไม้ และความยาวของกิ่ง
- Branch ปรับเปลี่ยนความเข้มของเส้นในแผนภาพต้นไม้
- Label ปรับเปลี่ยนชื่อ สี และ font ของ taxa

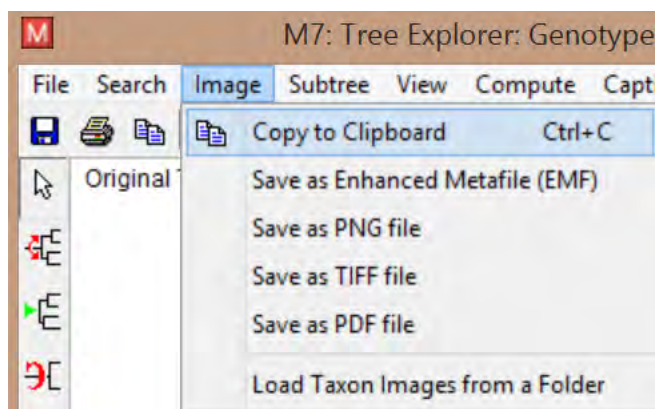


รูปที่ 5.18 ปรับเปลี่ยนลักษณะต่างๆ ของแผนภาพต้นไม้เชิงวิวัฒนาการ

12. การบันทึกภาพแผนภาพต้นไม้เชิงวิวัฒนาการ โดยคลิกที่คำว่า Image แล้วจะมีตัวเลือกในการบันทึกภาพแบบต่างๆ (รูปที่ 5.19) ดังนี้

- Copy to Clipboard
- Save as Enhanced Metafile (EMF)
- Save as PNG file
- Save as TIFF file
- Save as PDF file





รูปที่ 5.19 การบันทึกรูปภาพต้นไม้เชิงวิวัฒนาการ เป็น files ชนิดต่างๆ

13. การอ้างอิงถึงการใช้โปรแกรม MEGA หากทำการวิเคราะห์ข้อมูลหรือสร้างแผนภาพต้นไม้เชิงวิวัฒนาการด้วยโปรแกรม MEGA เพื่อเป็นการให้เครดิตแก่ผู้พัฒนาโปรแกรมดังกล่าว ควรข้อมูลในเอกสารอ้างอิง ดังนี้ “Kumar S., Stecher G., and Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870-1874.”

**เอกสารอ้างอิง และแนะนำอ่านเพิ่มเติม**

1. Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.
2. Felsenstein, J (1988) Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics* 22: 521-565.
3. Kumar S., Stecher G., and Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870-1874.
4. Saitou N. and Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.
5. Zuckerkandl E. and Pauling L. (1965). Evolutionary divergence and convergence in proteins. Edited in *Evolving Genes and Proteins*. In V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York.



# 6

เทคนิคที่เกี่ยวข้องกับพีซีอาร์ในด้านอณูไวรัสวิทยา

PCR-related techniques in molecular virology

## เทคนิคที่เกี่ยวข้องกับพีซีอาร์ในด้านอณูไวรัสวิทยา

### PCR-related techniques in molecular virology

#### การเพิ่มจำนวนสารพันธุกรรมด้วยเทคนิค Polymerase chain reaction (PCR)

หลักการของเทคนิค PCR เป็นเทคนิคที่ใช้ในการเพิ่มจำนวน DNA ที่ต้องการศึกษาอย่างจำเพาะในหลอดทดลอง โดยมีลักษณะเป็นการเพิ่มจำนวนแบบทวีคูณ ซึ่งจำเป็นต้องมีองค์ประกอบของปฏิกิริยา ดังนี้

- DNA template ที่สกัดได้จากเซลล์หรือเนื้อเยื่อที่ต้องการศึกษา ความเข้มข้นของ DNA template ที่เหมาะสม สำหรับปฏิกิริยา PCR ควรอยู่ในช่วง 100-500 ng/ $\mu$ L
- Primers เป็น oligonucleotide สายสั้น ๆ ขนาดประมาณ 17-25 bp ประกอบด้วย forward primer และ reverse primer ซึ่งจะมีทิศทางสวนทางกัน และจะจับกับ DNA template คนละ strand (forward primer จับกับ DNA template ที่เป็น anti-sense strand ส่วน reverse primer จับกับ DNA template ที่เป็น sense strand) นอกจากนี้ primers ควรมีส่วนเบส GC ประมาณ 40-60% และควรมีค่า melting temperature ( $T_m$ ) (อุณหภูมิที่ทำให้ 50% ของ DNA แยกสายเป็นสายเดี่ยว) ใกล้เคียงกัน ซึ่งสามารถคำนวณได้ ดังนี้  $T_m = [(A+T) \times 2] + [(G + C) \times 4]$  ความเข้มข้นของ primers ที่เหมาะสมสำหรับ PCR ควรอยู่ในช่วง 0.1-1  $\mu$ M (final concentration)
- Deoxynucleotide triphosphates (dNTPs) เป็น nucleotide หน่วยย่อยที่จะใช้ในการเติมเข้าไปในสาย DNA ที่สร้างขึ้นใหม่ ประกอบด้วย 4 ชนิด ได้แก่ dATP, dTTP, dCTP และ dGTP ความเข้มข้นของ dNTPs แต่ละชนิดที่เหมาะสมสำหรับ PCR ควรอยู่ในช่วง 200  $\mu$ M

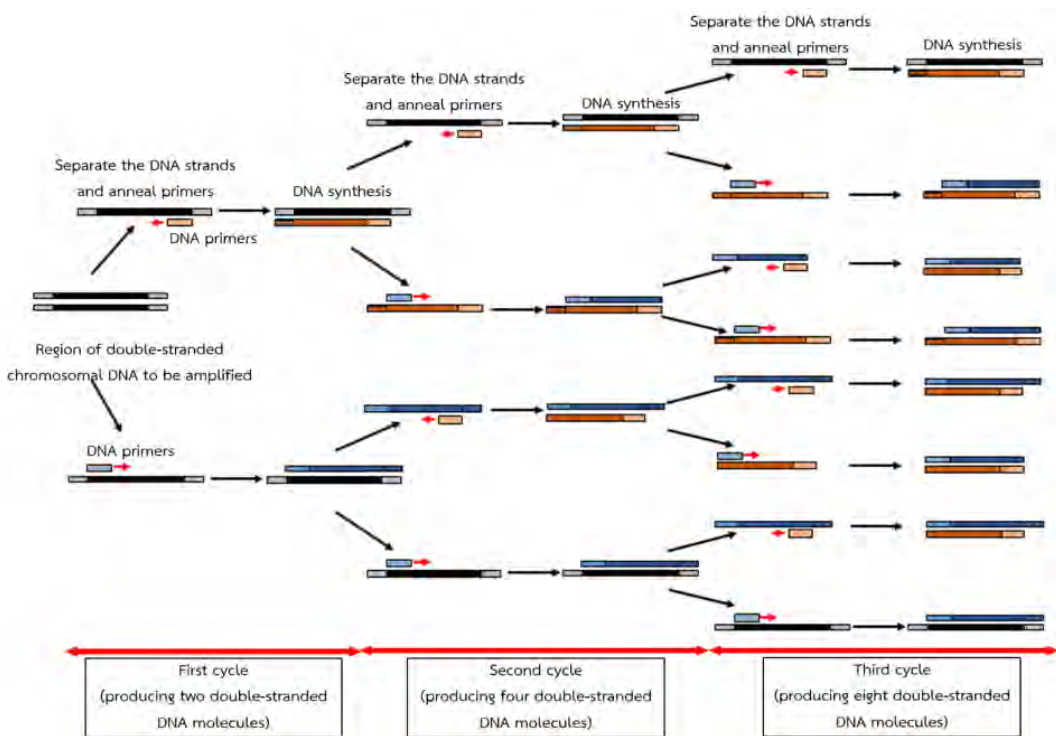
- *Taq* DNA Polymerase เป็นเอนไซม์ DNA polymerase ที่สามารถสร้างสาย DNA สายใหม่ในทิศทาง 5' → 3' โดยมีคุณสมบัติพิเศษสามารถทนความร้อนได้สูงมาก เนื่องจากเป็นเอนไซม์ที่สกัดมาจากแบคทีเรียที่มีชื่อว่า *Thermus aquaticus* ซึ่งสามารถอาศัยอยู่ในบริเวณน้ำพุร้อนได้ ดังนั้นเอนไซม์ *Taq* DNA Polymerase จึงเหมาะสมกับการใช้เป็นเอนไซม์ DNA polymerase ในปฏิกิริยา PCR ซึ่งมีขั้นตอน denaturation ที่อุณหภูมิ 94°C โดยความเข้มข้นของเอนไซม์ที่เหมาะสมสำหรับ PCR ควรอยู่ในช่วง 25 units/ml (final concentration)
- Magnesium ion เป็น co-factor ที่สำคัญของเอนไซม์ DNA polymerase ความเข้มข้นของ magnesium ion ที่เหมาะสมสำหรับ PCR ควรอยู่ในช่วง 1.5–2.0 mM (final concentration)
- Buffer เพื่อควบคุม pH ของปฏิกิริยาให้เหมาะสมต่อการทำงานของเอนไซม์

ในการเกิดปฏิกิริยา PCR จำเป็นต้องมีการควบคุมอุณหภูมิ (Thermal profile) เพื่อให้ปฏิกิริยาเกิดขึ้นได้อย่างเหมาะสม ซึ่งสามารถแบ่งออกเป็น 3 ขั้นตอน (รูปที่ 6.1) ได้แก่

- Denaturation เป็นขั้นตอนที่ใช้ความร้อนสูง (อุณหภูมิ 94°C ประมาณ 15-30 วินาที) เพื่อให้โมเลกุลของ DNA ต้นแบบมีการแยกสายจากสายคู่ (double strand DNA; dsDNA) เป็นสายเดี่ยว (single strand DNA; ssDNA)
- Annealing เป็นขั้นตอนที่ใช้อุณหภูมิ ในช่วง 50-60°C ประมาณ 15-30 วินาที เพื่อให้โมเลกุลของ primers สามารถจับแบบเบสคู่สม (complementary) กับ DNA template อย่างจำเพาะ โดยอุณหภูมิในขั้นตอน annealing จะขึ้นกับค่า Melting temperature ( $T_m$ ) ของ primers ที่ใช้ในปฏิกิริยา สามารถคำนวณได้จาก [Annealing temperature =  $T_m$  of primers – 5]

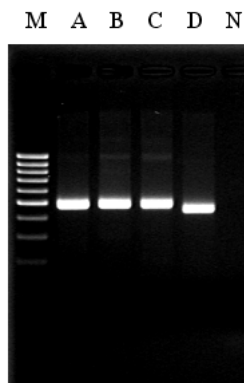


- Extension เป็นขั้นตอนที่ใช้ในการสร้าง DNA สายใหม่ในทิศทางจาก 5' → 3' ให้มีลำดับเบส complementary กับ DNA ต้นแบบ ซึ่งในขั้นนี้ใช้อุณหภูมิในช่วง 68-74°C (ขึ้นกับ Optimal temperature ของ DNA polymerase แต่ละชนิดที่ใช้ในปฏิกิริยา ซึ่งโดยทั่วไป *Taq* DNA polymerase จะมีอุณหภูมิเหมาะสมอยู่ที่ 72°C) ระยะเวลาในขั้นตอนนี้ประมาณ 30-180 วินาที ขึ้นกับขนาดความยาวของ PCR product ซึ่งจะถูกกำหนดโดยระยะห่างระหว่าง Forward primer และ Reverse primer ที่ใช้ในปฏิกิริยา โดยปกติสามารถประมาณได้ว่า PCR product ขนาด 1,000 bp ใช้เวลาในขั้น extension 1 นาที



รูปที่ 6.1 หลักการของเทคนิค Polymerase Chain Reaction (PCR)

ปฏิกิริยาจะดำเนินไปในแต่ละรอบซึ่งประกอบด้วย denaturation, annealing และ extension จำนวนทั้งสิ้นประมาณ 35-40 รอบ ดังนั้นหลังจากเสร็จปฏิกิริยาจะได้ปริมาณ DNA เพิ่มขึ้นแบบทวีคูณ  $2^n$  (โดยที่  $n$  = จำนวนรอบที่ใช้ในปฏิกิริยา) ปฏิกิริยาดังกล่าวเป็นการเพิ่มจำนวน DNA (amplification) โดยเกิดขึ้นเป็นปฏิกิริยาลูกโซ่ (chain reaction) จึงเรียกว่า polymerase chain reaction (PCR) ซึ่ง PCR แบบมาตรฐานจะเรียกว่า “conventional PCR” การประยุกต์ใช้เทคนิค conventional PCR มักใช้ร่วมกับเทคนิค agarose gel electrophoresis (รูปที่ 6.2) เพื่อใช้ในการตรวจวินิจฉัยแบบ screening โรคติดเชื้อต่างๆ และการตรวจสอบ insertion หรือ deletion ที่มีจำนวนเบสเปลี่ยนแปลงมากกว่า 30 bp เป็นต้น จากรูปที่ 6.2 เป็นการเพิ่มจำนวนยีนในส่วน *PreS1* ของเชื้อ hepatitis B virus (HBV) genotypes A-D ซึ่งจะพบว่า HBV genotype D จะเกิด deletion (ประมาณ 33 bp) ซึ่งสามารถตรวจสอบและจำแนกได้ด้วยเทคนิค conventional PCR และ agarose gel electrophoresis



รูปที่ 6.2 ตัวอย่างผลการตรวจสอบ PCR product ด้วย agarose gel electrophoresis

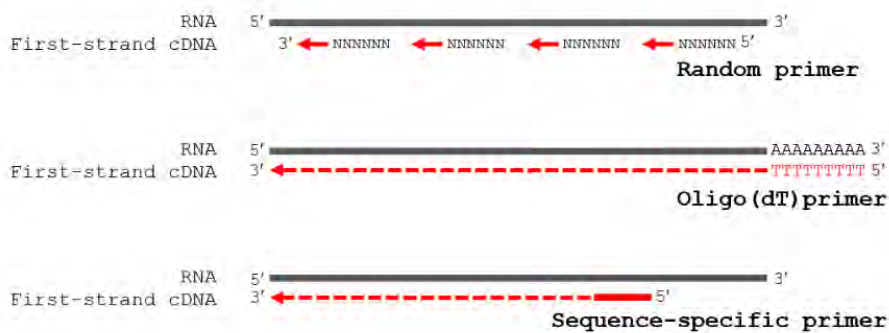
[จากผลงานวิจัยของ รศ.ดร.สัญญา พยุงกร <https://www.ncbi.nlm.nih.gov/pubmed/15288955>]

### เทคนิคที่มีพื้นฐานมาจาก PCR (PCR-based assay)

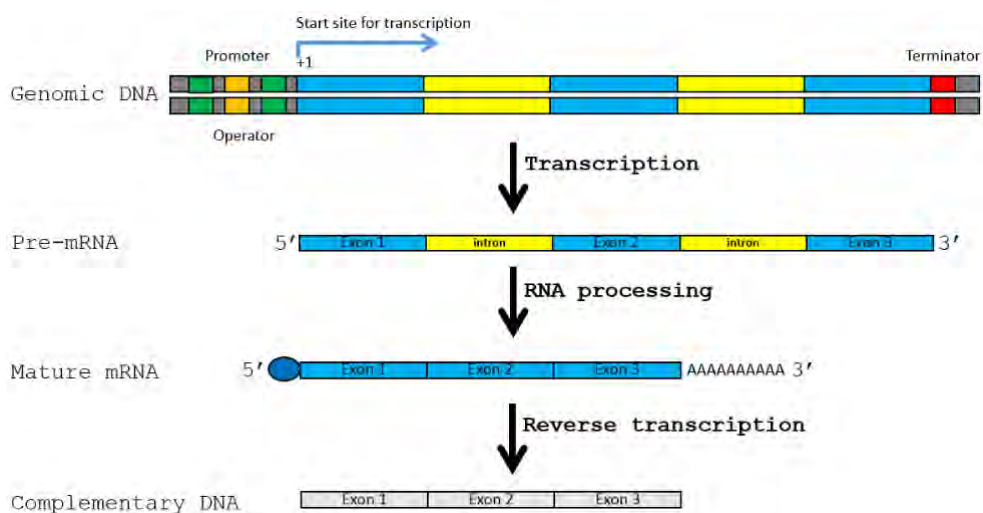
เนื่องจากเทคนิค PCR มีประโยชน์ในการเพิ่มจำนวนยีนที่ต้องการศึกษา จึงมีบทบาทสำคัญอย่างมากในการศึกษาด้านอณูชีววิทยา ในปัจจุบันจึงมีเทคนิคต่าง ๆ ที่ถูกพัฒนาขึ้นโดยมีพื้นฐานมาจาก PCR ดังนี้

### Reverse Transcription PCR (RT-PCR)

การทำ RT-PCR อาศัยหลักการสร้างสาย cDNA จาก mRNA โดยใช้เอนไซม์ reverse transcriptase และ primer ซึ่งแบ่งได้เป็น 3 ประเภท ได้แก่ random primer, oligo (dT) primer และ specific primer (รูปที่ 6.3) เมื่อได้เป็น cDNA แล้วจึงนำมาใช้เป็น template ในการทำ PCR ข้อดีของเทคนิคนี้คือเป็นการเพิ่มจำนวนยีนเฉพาะส่วนที่เป็น exons เท่านั้น เนื่องจาก mRNA จะมีการ splicing เอาส่วน introns ออกไป ซึ่งจะแตกต่างจากการเพิ่มจำนวนยีนในส่วน genomic DNA ที่มีทั้ง exons และ introns ต่อกัน (รูปที่ 6.4) นอกจากนี้นิยมใช้เทคนิค RT-PCR ในงานวิจัยทางด้านไวรัส เนื่องจากเชื้อไวรัสหลายชนิดมีสารพันธุกรรมเป็น RNA



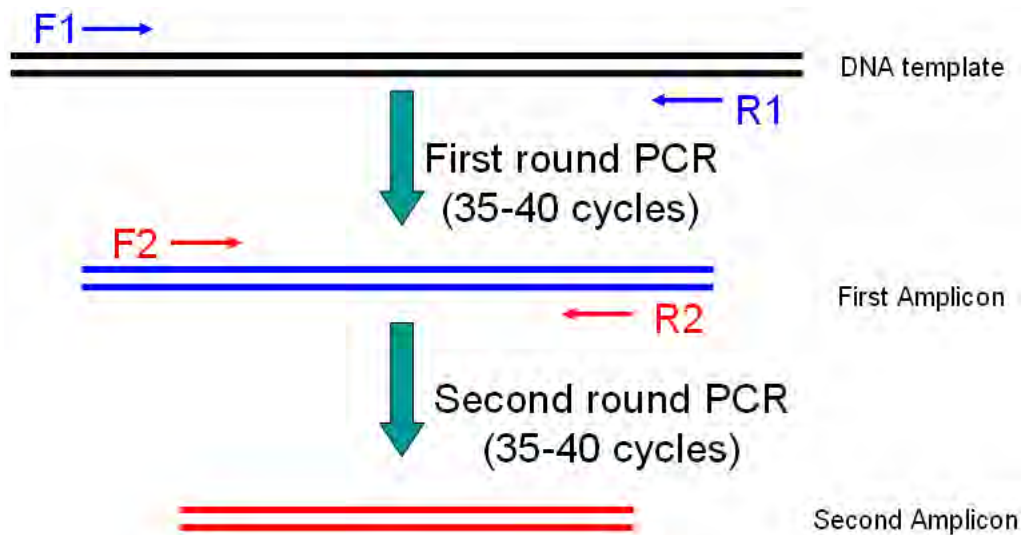
รูปที่ 6.3 Primer ประเภทต่าง ๆ สำหรับการทำให้ reverse transcription



รูปที่ 6.4 การเปรียบเทียบระหว่าง genomic DNA และ complementary DNA

## Nested PCR

เป็นเทคนิคการเพิ่มจำนวนยีนที่ต้องการด้วย PCR 2 ครั้ง ซึ่งใช้ primers 2 คู่ที่จำเพาะต่อยีนเดียวกัน แต่ primer แต่ละตัวนั้นจับกับ DNA template ในบริเวณที่แตกต่างกัน กล่าวคือ การทำ PCR ครั้งที่ 1 (first round PCR) จะใช้ primer จำนวน 1 คู่ที่จำเพาะต่อยีนที่ต้องการศึกษา เมื่อได้ PCR products จากการทำ PCR ครั้งที่ 1 แล้วจะนำมาใช้เป็น DNA template สำหรับการทำให้ PCR ครั้งที่ 2 ต่อไป (second round PCR) ซึ่งจะเปลี่ยน primers ที่ใช้ให้จำเพาะต่อยีนเดิมแต่มีการเปลี่ยนตำแหน่งในการจับกับ DNA template (รูปที่ 6.5) ซึ่งถ้าเปลี่ยนทั้ง forward primer และ reverse primer จะเรียกว่า nested PCR แต่ถ้าเปลี่ยน primer เพียงตัวเดียว (forward primer หรือ reverse primer) จะเรียกว่า semi-nested PCR หรือ hemi-nested PCR ดังนั้นจุดเด่นของเทคนิคนี้คือเป็นการเพิ่มความจำเพาะ (specificity) และความไว (sensitivity) ในการเพิ่มจำนวนยีน และการตรวจสอบยีนที่ต้องการศึกษา จึงนิยมใช้ในการตรวจวินิจฉัยโรคติดเชื้อที่เกิดจากไวรัส เนื่องจากตัวอย่างสิ่งส่งตรวจจากผู้ป่วย อาจมีสารพันธุกรรมของเชื้อไวรัสในปริมาณที่น้อยมากๆ

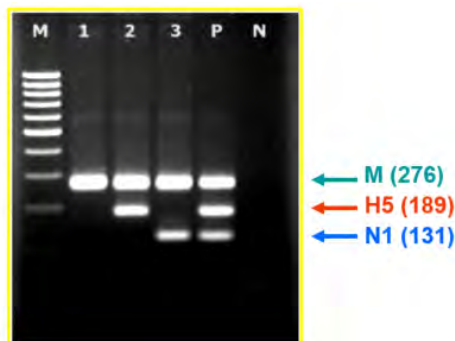


รูปที่ 6.5 หลักการของเทคนิค nested PCR

## Multiplex PCR

เป็นเทคนิคที่มีพื้นฐานมาจาก PCR โดยมีการใช้ primers มากกว่า 1 คู่ ในการเพิ่มจำนวนยีนพร้อมกันหลายๆ ท่อนในปฏิกิริยาหลอดเดียวกัน (รูปที่ 6.6) โดยมีจุดเด่นในการใช้เพื่อตรวจวินิจฉัยโรค เนื่องจากการทำปฏิกิริยา PCR โดยใช้ primers หลายคู่พร้อมกันในหลอดเดียวกัน จะทำให้ประหยัดทั้งเวลา และค่าใช้จ่ายสำหรับน้ำยาที่เป็นส่วนผสมของปฏิกิริยา PCR แต่จะมีข้อจำกัดในแง่ความไว (sensitivity) ในการทดสอบจะลดลง จากรูปที่ 6.6 เป็นตัวอย่างการใช้เทคนิค multiplex PCR ในการตรวจวินิจฉัยการติดเชื้อไวรัสไข้หวัดนก (avian influenza A virus) สายพันธุ์ H5N1 โดยใช้ primers จำนวน 3 คู่ ซึ่งจำเพาะกับยีน matrix (M), hemagglutinin (HA) และ neuraminidase (NA) ตามลำดับ ซึ่งสามารถแปลผลได้ ดังนี้

- หากขึ้น band ขนาด 276 bp ของ M gene เพียงอย่างเดียว แสดงว่าติดเชื้อไวรัสไข้หวัดใหญ่ สายพันธุ์อื่นๆ ซึ่งไม่ใช่ H5N1
- หากขึ้น band ขนาด 276 bp ของ M gene และ 189 bp ของ H5 แสดงว่าติดเชื้อไวรัสไข้หวัดใหญ่ สายพันธุ์ H5N? แต่ไม่ใช่ N1 ซึ่งอาจเป็น H5N2 เป็นต้น
- หากขึ้น band ขนาด 276 bp ของ M gene และ 131 bp ของ N1 แสดงว่าติดเชื้อไวรัสไข้หวัดใหญ่ สายพันธุ์ H?N1 แต่ไม่ใช่ H5 ซึ่งอาจเป็น H1N1 เป็นต้น
- หากขึ้นทั้ง 3 bands แสดงว่าติดเชื้อไวรัสไข้หวัดนก สายพันธุ์ H5N1



รูปที่ 6.6 ตัวอย่างผลการตรวจเชื้อไวรัสไข้หวัดนก สายพันธุ์ H5N1 ด้วยเทคนิค multiplex PCR

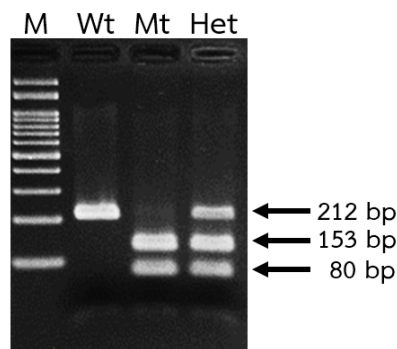
[จากผลงานวิจัยของ รศ.ดร.สัญญา พยุภกร <https://www.ncbi.nlm.nih.gov/pubmed/15671756>]

### PCR-Restriction Fragment Length Polymorphism (PCR-RFLP)

เป็นเทคนิคที่อาศัยการเพิ่มจำนวนยีนที่สนใจด้วย PCR ก่อน จากนั้นจึงใช้เอนไซม์ตัดจำเพาะ (restriction enzyme) ซึ่งเป็น endonuclease จากแบคทีเรียที่สามารถตัด DNA บริเวณตัดจำเพาะได้ ซึ่งเอนไซม์ตัดจำเพาะมีหลายชนิด แต่ละชนิดจะมีบริเวณตัดจำเพาะที่แตกต่างกันออกไป ข้อดีของเทคนิค PCR-RFLP คือสามารถใช้ตรวจสอบเพื่อจำแนก alleles หรือ polymorphisms หรือ mutations ได้ เช่น จำแนกระหว่าง wide type และ mutant โดยเลือกเอนไซม์ตัดจำเพาะชนิดที่สามารถตัดเพื่อแยกความแตกต่างได้ เมื่อนำ DNA มาแยกตามขนาดด้วยวิธี gel electrophoresis จะทำให้สามารถจำแนกได้ว่าตัวอย่างที่สนใจเป็น wide type หรือ mutant โดยดูจากขนาดของแถบ DNA (รูปที่ 6.7) ดังนี้

- ไม่สามารถตัด DNA ชนิด wide type จึงได้แถบ DNA ขนาด 212 bp
- สามารถตัด DNA ชนิด mutant จึงได้แถบ DNA ขนาด 153 และ 80 bp
- กรณีที่เป็น heterozygous จะมีทั้ง alleles ที่เป็น wild type และ mutant ผสมกัน จะได้แถบ DNA ขนาด 212, 153 และ 80 bp

อย่างไรก็ตามวิธีนี้มีข้อจำกัดคือ ในบางกรณีจะไม่สามารถหาเอนไซม์ตัดจำเพาะเพื่อจำแนกตำแหน่งที่พบการกลายพันธุ์ได้ และแถบ DNA ที่ได้จากการตัดบางครั้งไม่ชัดเจนทำให้การแปลผลค่อนข้างยาก



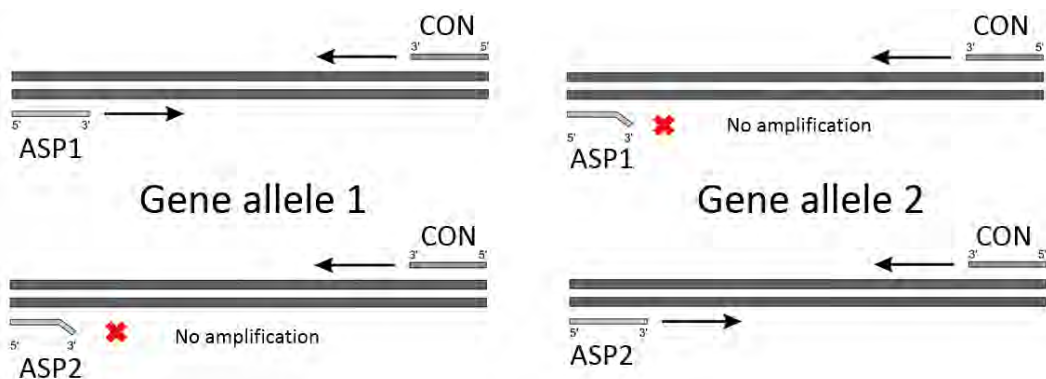
รูปที่ 6.7 ตัวอย่างผลการตรวจการกลายพันธุ์ของยีนด้วยเทคนิค PCR-RFLP

[จากผลงานวิจัยของ รศ.ดร.สัณชัย พยุงกร <https://www.ncbi.nlm.nih.gov/pubmed/26745093>]

### Allele specific PCR

เป็นเทคนิคการเพิ่มจำนวนยีนที่ต้องการตรวจหา single nucleotide polymorphisms (SNPs) หรือการกลายพันธุ์แบบ point mutation โดยใช้ allele specific primers 2 ตัว ซึ่งจะแตกต่างกันที่ลำดับเบสตัวสุดท้ายทางด้านปลาย 3' เพื่อให้จำเพาะกับแต่ละ allele (รูปที่ 6.8)

- Allele specific primer 1 (ASP1) จำเพาะกับ allele ที่ 1 หรือ wild type
- Allele specific primer 2 (ASP2) จำเพาะกับ allele ที่ 2 หรือ mutant



รูปที่ 6.8 หลักการของ allele specific PCR

การทำ allele specific PCR จะแยกปฏิกิริยาออกเป็น 2 หลอด โดยหลอดที่ 1 ใส่เฉพาะ ASP1 ส่วนหลอดที่ 2 ใส่เฉพาะ ASP2 เมื่อ primer สามารถจับได้อย่างจำเพาะกับ DNA template จะทำให้มีการเพิ่มจำนวน DNA จึงจะเกิด PCR product ขึ้น แต่ในกรณีที่เกิด mismatch ที่ปลาย 3' ของ primer จะไม่สามารถเพิ่มจำนวน DNA จึงไม่เกิด PCR product เกิดขึ้น ดังนั้นมีแนวทางในการแปลผล (ตารางที่ 6.1) ดังนี้

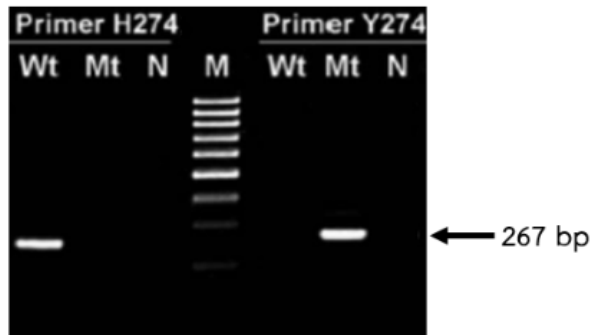
ตารางที่ 6.1 ตัวอย่างการแปลผล genotypes ด้วยเทคนิค allele specific PCR

Sample	PCR reactions		Interpretation of genotypes
	ASP1	ASP2	
A	+	-	Homozygous allele 1 (wild type)
B	-	+	Homozygous allele 2 (mutant)
C	+	+	Heterozygous alleles



จากรูปที่ 6.9 เป็นตัวอย่างการตรวจวินิจฉัยการกลายพันธุ์ที่ตำแหน่ง H274Y บนยีน neuraminidase ของเชื้อไวรัสไข้หวัดใหญ่ ซึ่งจะส่งผลทำให้ดื้อต่อยาต้านไวรัส oseltamivir โดยใช้เทคนิค allele specific PCR จากการทดสอบพบว่า

- primer H274 จะสามารถเพิ่มจำนวนยีนที่เป็น wild type (H274) ได้เท่านั้น
- primer Y274 จะสามารถเพิ่มจำนวนยีนที่เป็น mutant (Y274) ได้เท่านั้น



รูปที่ 6.9 ตัวอย่างผลการตรวจวินิจฉัยการกลายพันธุ์ ด้วยเทคนิค allele specific PCR

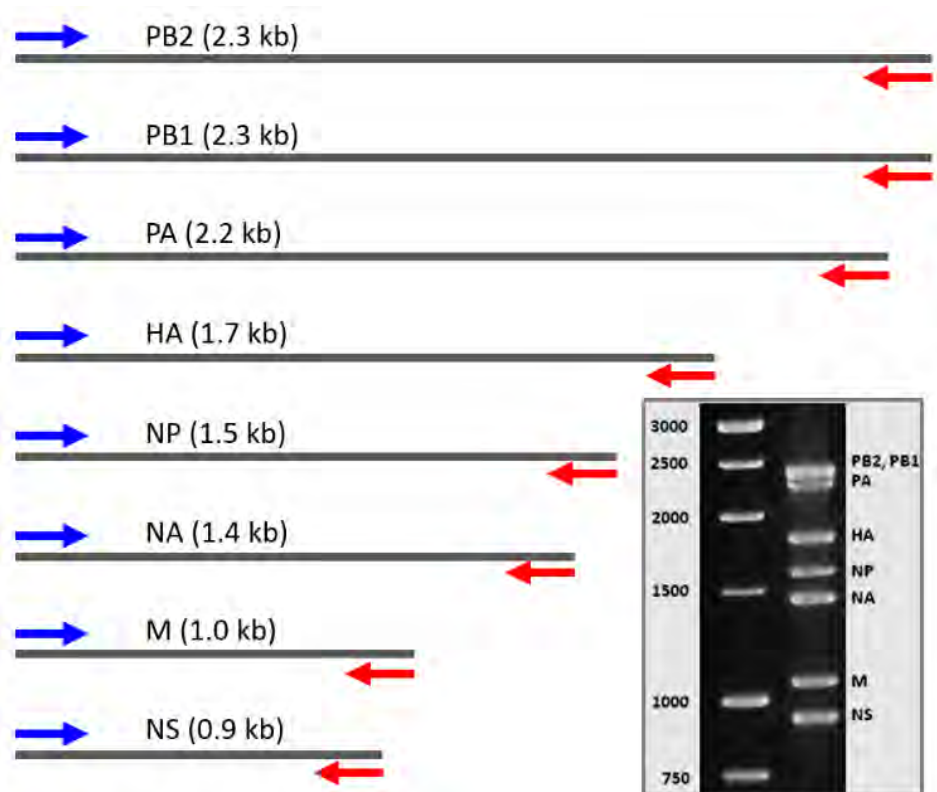
[จากผลงานวิจัยของ รศ.ดร.สัญญา พยุงภร <https://www.ncbi.nlm.nih.gov/pubmed/17354342>]

### Universal PCR

เป็นเทคนิคการเพิ่มจำนวนยีนโดยใช้ primers ที่จับกับบริเวณอนุรักษ์ (conserved region) ของยีน ทำให้ primers เพียง 1 คู่นี้มีความสามารถในการเพิ่มจำนวน DNA templates หลากๆ รูปแบบที่มีลักษณะคล้ายคลึงกันได้ นิยมใช้ในงานวิจัยทางด้านไวรัสวิทยา และการตรวจวินิจฉัยโรคติดเชื้อจากไวรัส ดังนี้

- การเพิ่มจำนวนยีนของไวรัสในตระกูลเดียวกัน แต่เป็นไวรัสต่างชนิดกัน เช่น Dengue virus, Yellow fever virus, Japanese encephalitis virus, West Nile virus และ Zika virus จัดเป็นไวรัสที่มียุงเป็นพาหะ อยู่ในตระกูล *Flaviviridae* เหมือนกัน ซึ่งมีลักษณะยีนและจีโนมบางส่วนคล้ายคลึงกัน จึงนิยมใช้ universal PCR ในการตรวจกรองเบื้องต้นว่าผู้ป่วยติดเชื้อไวรัสในกลุ่ม *Flavivirus* หรือไม่

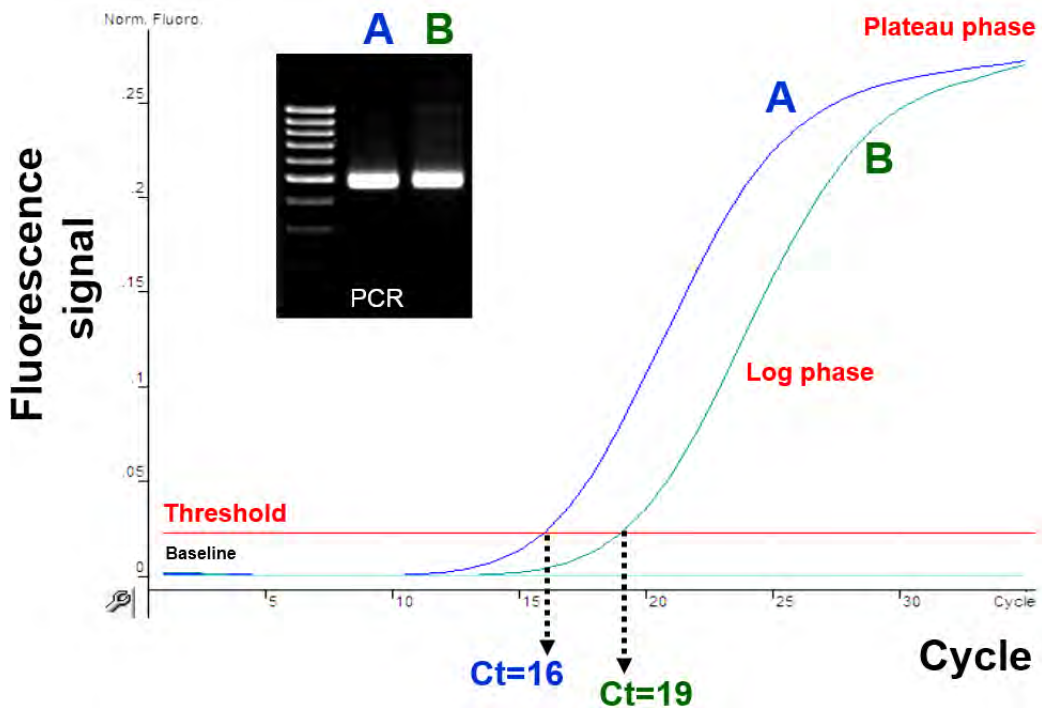
- การเพิ่มจำนวนยีนของไวรัสชนิดเดียวกัน แต่มีหลายสายพันธุ์ (subtypes, genotypes, serotypes) ในกรณีที่ไวรัสมีการกลายพันธุ์และถูกจำแนกออกเป็นหลากหลายสายพันธุ์ เช่น hepatitis C virus (HCV), human immunodeficiency virus (HIV) จะนิยมใช้ universal PCR ในการตรวจวินิจฉัยว่าผู้ป่วยติดเชื้อไวรัสชนิดที่สงสัยหรือไม่ โดยยังไม่จำแนกสายพันธุ์
- การเพิ่มจำนวนยีนพร้อมกันหลายๆ ยีนของไวรัสชนิดเดียวกัน ใช้ในกรณีที่พบบริเวณอนุรักษ์ (conserved regions) ในหลายๆ ยีนของเชื้อไวรัส เช่น Influenza A virus จะมี conserved regions ที่บริเวณปลาย 5' และ 3' ของยีนทั้ง 8 ท่อน ทำให้สามารถใช้ universal PCR ในการเพิ่มจำนวนทั้ง 8 ยีนได้พร้อมกัน (รูปที่ 6.10)



รูปที่ 6.10 การเพิ่มจำนวนยีนทั้ง 8 ท่อนของเชื้อไวรัสไข้หวัดใหญ่ ด้วยเทคนิค universal PCR

## Real-time PCR

Real-time PCR เป็นเทคนิคที่ใช้ในการเพิ่มจำนวน DNA ที่ต้องการศึกษาอย่างจำเพาะ และสามารถติดตามวัดปริมาณการเพิ่มจำนวนของ DNA เป้าหมายได้ในทุกๆ รอบของการเพิ่มจำนวน DNA ตั้งแต่เริ่มต้นจนกระทั่งสิ้นสุดปฏิกิริยา (real-time detection) โดยอาศัยการตรวจวัดสัญญาณเรืองแสงที่เพิ่มขึ้น ซึ่งเป็นสัดส่วนโดยตรงกับปริมาณ DNA ที่เพิ่มขึ้นจากปฏิกิริยาในแต่ละรอบ ดังนั้นเมื่อสร้างกราฟการเพิ่มจำนวน (amplification plot) ที่มีแกน Y แสดงสัญญาณการเรืองแสง และแกน X แสดงจำนวนรอบของการเพิ่มจำนวน จะได้กราฟที่มีลักษณะเป็นรูป S (sigmoid หรือ exponential curve) ซึ่งมีค่านิยามต่าง ๆ เกี่ยวกับ amplification plot (รูปที่ 6.11) ดังนี้



รูปที่ 6.11 กราฟการเพิ่มจำนวน DNA (Amplification plot) ด้วยเทคนิค real-time PCR  
 [จากผลงานวิจัยของ รศ.ดร.สัญญาชัย พยุงกร <https://www.ncbi.nlm.nih.gov/pubmed/21820468>]

- Baseline คือ ช่วงเริ่มต้นของการเพิ่มจำนวน DNA ซึ่งมีการสะสมของสารเรืองแสง แต่สัญญาณเรืองแสงยังต่ำกว่าระดับที่เครื่องจะตรวจวัดได้
- Threshold คือ ระดับสัญญาณสารเรืองแสงที่ใช้เป็นค่า cut-off ในการตรวจวัดสารเรืองแสง
- Threshold cycle (Ct) คือ จำนวนรอบของปฏิกิริยาในการเพิ่มจำนวนที่ทำให้สัญญาณสารเรืองแสงสูงกว่าระดับ Threshold
- Log/exponential phase คือช่วงที่ระดับสัญญาณสารเรืองแสงสูงเหนือระดับ Threshold และมีแนวโน้มการเพิ่มจำนวนแบบทวีคูณ
- Plateau phase คือช่วงที่ระดับสัญญาณสารเรืองแสงคงที่ ไม่มีการเพิ่มขึ้น เนื่องจากสารต่าง ๆ ที่จำเป็นในปฏิกิริยา เช่น primers และ dNTP หมดไป และ เอนไซม์ DNA polymerase เริ่มหมด activity

#### ➤ ความแตกต่างระหว่างเทคนิค real-time PCR และ conventional PCR

ดังที่กล่าวไปแล้วข้างต้นว่าการตรวจวัดการเพิ่มจำนวนของ DNA โดยเทคนิค Real-time PCR นั้นอาศัยการวัดปริมาณสารเรืองแสงที่เพิ่มขึ้นในแต่ละรอบของปฏิกิริยาที่ดำเนินไปตั้งแต่ต้นจนจบ (real-time detection) แต่การตรวจวัดการเพิ่มจำนวนของ DNA โดยใช้เทคนิค PCR นั้นจะเรียกว่าเป็นแบบ end-point detection เนื่องจากต้องปล่อยให้ปฏิกิริยาจนถึงรอบสุดท้ายก่อน จึงจะสามารถนำสารละลายจากปฏิกิริยามาตรวจสอบ ซึ่งอาจตรวจสอบได้ด้วยเทคนิค agarose Gel electrophoresis แล้วย้อมด้วยสาร ethidium bromide จากนั้นจึงตรวจสอบการเรืองแสงของแถบ DNA ภายใต้แสง UV ซึ่งเมื่อเปรียบเทียบการเพิ่มจำนวน DNA ระหว่าง 2 ตัวอย่าง (A และ B) ที่มีปริมาณ DNA ตั้งต้นแตกต่างกันโดยใช้ real-time detection และ end-point detection จะให้ผลแตกต่างกัน (รูปที่ 6.11) ดังนี้

- End-point detection จะพบว่าทั้งตัวอย่าง A และ B มีความเข้มข้นของแถบ DNA ใกล้เคียงกันมาก เนื่องจากเมื่อปฏิกิริยาดำเนินไปถึงช่วง plateau phase (รอบที่มากกว่า 30) ลักษณะการเพิ่มจำนวนของ DNA จะไม่เป็นแบบทวีคูณ ทำให้ในช่วงหลังของปฏิกิริยา ในตัวอย่าง A และ B มีปริมาณ DNA ใกล้เคียงกัน เมื่อตรวจสอบด้วยวิธี end-point detection จึงได้ความเข้มข้นของแถบ DNA ใกล้เคียงกัน
- Real-time detection อาศัยการวัดปริมาณสารเรืองแสงที่เพิ่มขึ้นในแต่ละรอบของปฏิกิริยา และใช้ค่า Ct เป็นตัวบ่งชี้ปริมาณ DNA ที่มีในสารตั้งต้น กล่าวคือ
  - ตัวอย่างที่มีปริมาณ DNA ตั้งต้นมาก จะมีค่า Ct ต่ำ เนื่องจากมีปริมาณ DNA ตั้งต้นมาก จึงใช้จำนวนรอบของปฏิกิริยาเพียงไม่กี่รอบก็สามารถเพิ่มปริมาณสารเรืองแสงให้สูงกว่าระดับ threshold ได้
  - ตัวอย่างที่มีปริมาณ DNA ตั้งต้นน้อย จะมีค่า Ct สูง เนื่องจากมีปริมาณ DNA ตั้งต้นน้อย จึงใช้จำนวนรอบของปฏิกิริยามากเพื่อให้สามารถเพิ่มปริมาณสารเรืองแสงให้สูงกว่าระดับ threshold ได้

ดังนั้นจะเห็นได้ว่าเทคนิค PCR แบบปกติ (conventional PCR) นั้นไม่เหมาะกับการตรวจวัด DNA เชิงปริมาณ (quantitative analysis) เช่นการวัดอัตราการแสดงออกของยีน หรือวัดปริมาณเชื้อไวรัสในผู้ป่วย เป็นต้น เนื่องจากผลที่ได้จาก end-point detection ไม่มีความถูกต้องและแม่นยำ แต่เทคนิค real-time PCR จะมีความเหมาะสมกับการตรวจวัด DNA เชิงปริมาณ (quantitative analysis) เนื่องจากให้ผลที่ถูกต้องและแม่นยำกว่า นอกจากนี้ยังสามารถประยุกต์ใช้ในการวัดปริมาณเชิงเปรียบเทียบ (relative quantitation) และการวัดปริมาณสัมบูรณ์ (absolute quantitation) ได้ด้วย

➤ การวัดปริมาณเชิงเปรียบเทียบ (relative quantitation) ด้วย delta Ct method

เป็นการประยุกต์ใช้เทคนิค Real-time PCR ในการคำนวณหาปริมาณเชิงเปรียบเทียบของ DNA ระหว่างตัวอย่างที่สนใจกับตัวอย่างควบคุม โดยผลการเปรียบเทียบจะแสดงออกมาในรูปจำนวนเท่าที่แตกต่างกัน โดยมีสูตรการคำนวณดังนี้

$$\Delta Ct (\text{sample}) = Ct (\text{target gene of sample}) - Ct (\text{reference gene of sample})$$

$$\Delta Ct (\text{calibrator}) = Ct (\text{target gene of calibrator}) - Ct (\text{reference gene of calibrator})$$

$$\Delta\Delta Ct = \Delta Ct (\text{sample}) - \Delta Ct (\text{calibrator})$$

$$\text{Ratio of sample: calibrator} = 2^{-\Delta\Delta Ct}$$

ตัวอย่างการคำนวณ relative quantitation by delta Ct method

การทดสอบผลของสารสกัดจากสมุนไพร ในการยับยั้งการเพิ่มจำนวนของเชื้อไวรัส (viral replication) โดยเซลล์ออกแบ่งเป็น 2 กลุ่ม คือ กลุ่มตัวอย่างที่ได้รับสารสกัดจากสมุนไพร (treated group) และกลุ่มควบคุมซึ่งไม่ได้รับสารสกัดจากสมุนไพร (untreated group) โดยทั้ง 2 กลุ่มได้รับเชื้อไวรัสเริ่มต้นเท่ากัน หลังจากนั้นเลี้ยงเซลล์ที่อุณหภูมิ 37°C เป็นเวลา 48 ชั่วโมง แล้วจึงทำการสกัด total RNA จากตัวอย่างทั้ง 2 กลุ่ม แล้วทำ reverse transcription สร้างเป็น cDNA จากนั้นจึงทำการวัดปริมาณยีนของเชื้อไวรัส (target gene) และ GAPDH (reference gene) โดยใช้เทคนิค real-time PCR ได้ผลค่า Ct ดังตารางที่ 6.2

ตารางที่ 6.2 ตัวอย่างค่า Ct ในการคำนวณ relative quantitation by Delta Ct method

Group	Ct of viral target gene	Ct of GAPDH (ref gene)
Treated (sample)	32	20
Untreated (calibrator)	26	22

จากตารางสามารถคำนวณได้ดังนี้

$$\Delta Ct (\text{sample}) = Ct (\text{target gene of sample}) - Ct (\text{reference gene of sample})$$

$$\text{แทนค่า } \Delta Ct (\text{treated}) = 32 - 20 = 12$$

$$\Delta C_t (\text{calibrator}) = C_t (\text{target gene of calibrator}) - C_t (\text{reference gene of calibrator})$$

$$\text{แทนค่า } \Delta C_t (\text{untreated}) = 26 - 22 = 4$$

$$\Delta \Delta C_t = \Delta C_t (\text{sample}) - \Delta C_t (\text{calibrator})$$

$$= \Delta C_t (\text{treated}) - \Delta C_t (\text{untreated}) = 12 - 4 = 8$$

$$\text{Ratio of sample: calibrator} = 2^{(-\Delta \Delta C_t)}$$

$$\text{Ratio of treated group: untreated group} = 2^{(-\Delta \Delta C_t)} = 2^{-8} = 1 / 256 \text{ folds}$$

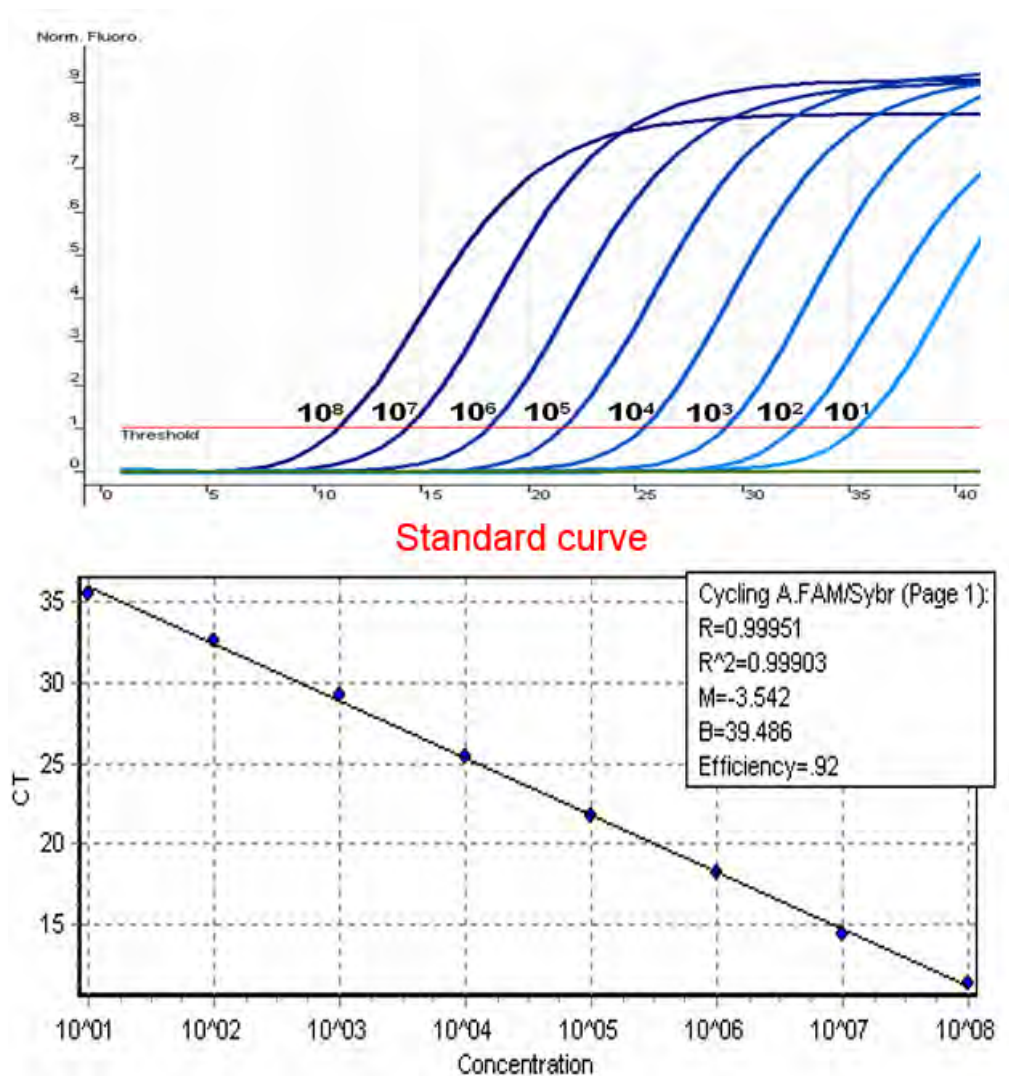
ดังนั้นจึงสามารถแปลผลได้ว่า กลุ่มที่ได้รับสารสกัดจากสมุนไพรมะขาม (treated group) มีจำนวนของเชื้อไวรัส น้อยกว่ากลุ่มที่ไม่ได้รับสารสกัดจากสมุนไพรมะขาม (untreated group) ประมาณ 256 เท่า ซึ่งจะสรุปได้ว่าสารสกัดจากสมุนไพรมะขามดังกล่าว มีผลในการยับยั้งการเพิ่มจำนวนของเชื้อไวรัส เป็นต้น

### ➤ การวัดปริมาณสัมบูรณ์ (absolute quantitation)

เป็นการประยุกต์ใช้เทคนิค Real-time PCR ในการคำนวณหาปริมาณสัมบูรณ์ของ DNA ในตัวอย่างที่สนใจโดยการเปรียบเทียบกับ DNA มาตรฐานที่ทราบปริมาณแน่นอนแล้ว ซึ่งจะต้องมีการเจือจาง DNA มาตรฐานเป็นความเข้มข้นที่ลดลงครั้งละ 10 เท่า (10-fold serial dilutions) เช่น จาก  $10^7$  เจือจางเป็น  $10^6$  copies/ $\mu\text{L}$  จาก  $10^6$  เจือจางเป็น  $10^5$  copies/ $\mu\text{L}$  และเจือจางต่อไปจนได้ 10 copies/ $\mu\text{L}$  ดังนั้นจะได้ DNA มาตรฐานที่มีความเข้มข้นลดหลั่นลงไปจาก  $10^7$  ไปสู่ 10 copies/ $\mu\text{L}$  แล้วนำเอา DNA มาตรฐานดังกล่าวมาใช้เป็น DNA template สำหรับทำ real-time PCR และบันทึกค่า Ct จาก amplification plot จากนั้นสามารถสร้างกราฟมาตรฐาน (standard curve) โดยที่แกน Y แสดงค่า Ct และแกน X แสดงค่า Log concentration ซึ่งจะได้กราฟที่มีลักษณะเป็นเส้นตรง (รูปที่ 6.12) เมื่อต้องการหาปริมาณ DNA ของยีนที่จำเพาะจากตัวอย่างที่สนใจ สามารถทำ real-time PCR และบันทึกค่า Ct เพื่อนำไปเปรียบเทียบกับ standard curve ก็จะสามารถระบุปริมาณสัมบูรณ์ของ DNA จากตัวอย่างที่สนใจได้



เทคนิคที่เกี่ยวข้องกับพีซีอาร์ในด้านอณูไวรัสวิทยา



รูปที่ 6.12 ตัวอย่าง Standard curve สำหรับ absolute quantitation

[จากผลงานวิจัยของ รศ.ดร.สัญญา พงษ์กร <https://www.ncbi.nlm.nih.gov/pubmed/21820468>]

การประยุกต์ใช้เทคนิค real-time PCR ทางด้านไวรัสวิทยา นิยมนำมาใช้ในการตรวจวัดปริมาณเชื้อไวรัส (viral load) เพื่อช่วยในการตรวจวินิจฉัยและพยากรณ์โรค เช่น การตรวจวัด viral load ก่อนและหลังการให้ยาต้านไวรัส เพื่อพิจารณาผลของการรักษา และประสิทธิภาพของยาต้านไวรัส ตลอดจนปรับเปลี่ยนแนวทางการรักษา เช่น การเพิ่มปริมาณยา หรือการเปลี่ยนชนิดของยาต้านไวรัส เป็นต้น

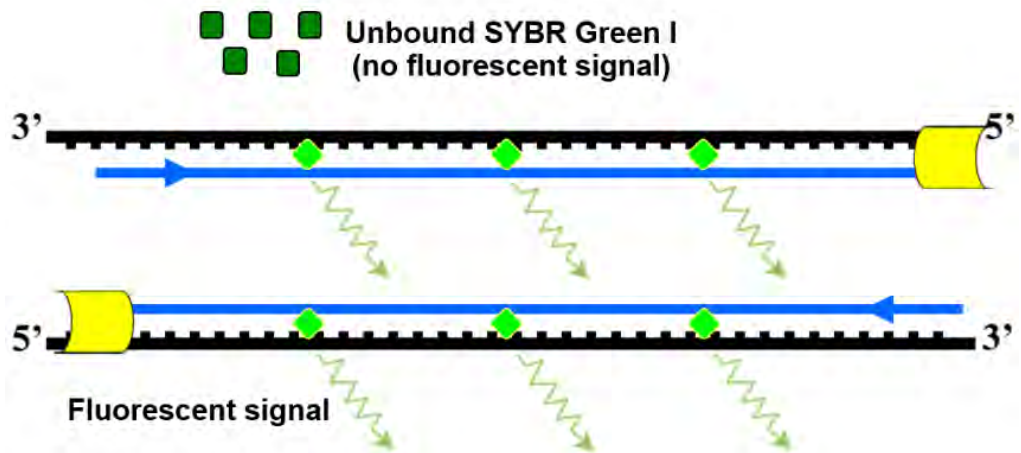
➤ รูปแบบการตรวจวัด (detection format) สำหรับ real-time PCR

โดยทั่วไปการตรวจติดตามวัดปริมาณการเพิ่มจำนวนของ DNA ด้วยสารเรืองแสงในเทคนิค real-time PCR สามารถแบ่งได้เป็น 2 รูปแบบ (ตารางที่ 6.3) ดังนี้

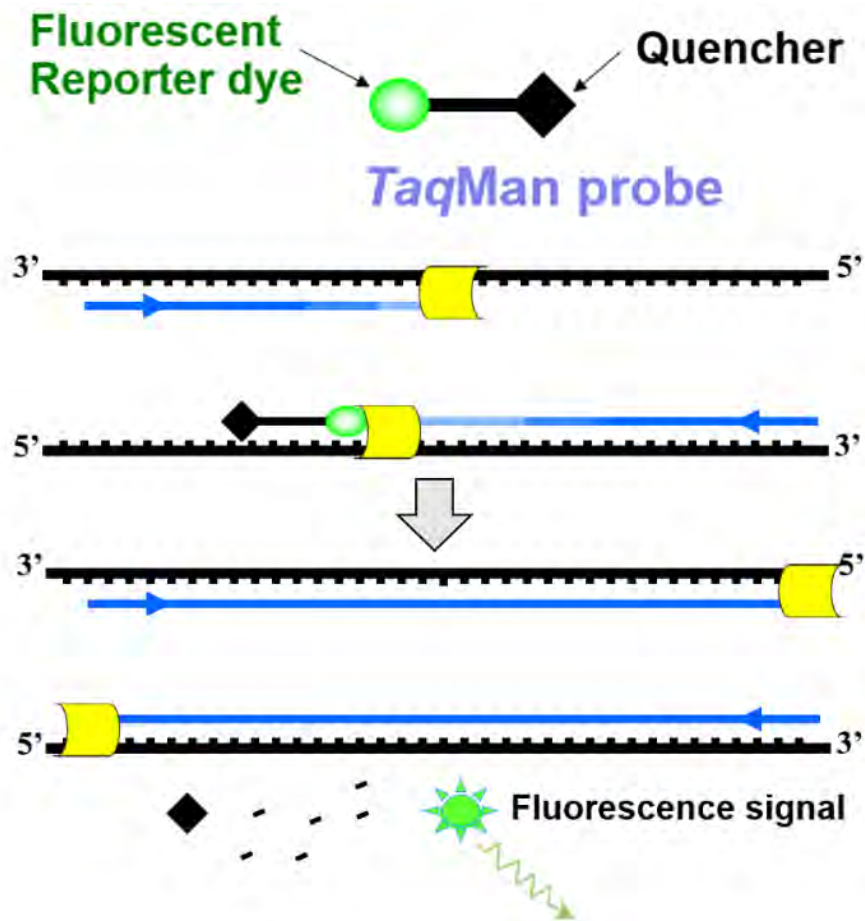
- **DNA binding fluorescent dyes** ได้แก่ SYBR Green I, LC Green และ SYTO 9 เป็นต้น โดยสารเรืองแสงที่นิยมใช้มากที่สุดคือ SYBR Green I ซึ่งสามารถจับกับบริเวณ minor groove ของ DNA สายคู่แบบไม่จำเพาะ เมื่อจับแล้วจะเกิดการเปลี่ยนแปลงโครงสร้างและเปล่งแสงที่มีความยาวคลื่นประมาณ 530 nm ดังนั้นเมื่อเกิดปฏิกิริยาการเพิ่มจำนวน DNA ด้วย real-time PCR สัญญาณของสารเรืองแสง SYBR Green I ก็จะเพิ่มขึ้นตามปริมาณ DNA ที่เพิ่มขึ้น (รูปที่ 6.13)
- **Probes based assay** ได้แก่ TaqMan probe, FRET Hybridization probes, Molecular Beacons, Scorpion probe และ Sun rise probe เป็นต้น ซึ่ง probe แต่ละชนิดจะมีหลักกลไกการทำงานแตกต่างกันไป แต่ probe ที่ได้รับความนิยมมากที่สุดคือ TaqMan probe ซึ่งเป็น oligonucleotide สายเดี่ยว (25-30 bp) ที่ติดฉลากด้วยสารเรืองแสง (fluorescent dye) ที่ปลาย 5' และตัวดับขั้ว (quencher) ที่ปลาย 3' ซึ่งในสภาวะปกติจะไม่มีการเรืองแสงเนื่องจากสารเรืองแสงอยู่ใกล้กับตัวดับขั้ว แต่ในขณะที่เกิด ปฏิกิริยาการเพิ่มจำนวน DNA โดยอาศัย เอนไซม์ Taq DNA polymerase เอนไซม์ชนิดนี้จะมี 5'→3' exonuclease activity ซึ่งสามารถย่อย TaqMan probe ได้ ทำให้โมเลกุลของสารเรืองแสงและตัวดับขั้วหลุดออกจากกัน ดังนั้นจึงเกิดการเรืองแสงขึ้น และความเข้มของสารเรืองแสงก็จะเพิ่มขึ้นตามปริมาณ DNA ที่เพิ่มขึ้น (รูปที่ 6.14)

ตารางที่ 6.3 สรุปข้อดีและข้อเสียของรูปแบบ SYBR Green I และ TaqMan probe

รูปแบบ	ข้อดี	ข้อเสีย
SYBR Green I	ง่าย และราคาถูก	ไม่มีความจำเพาะ
	ทำ melting curve analysis ได้	ตรวจแบบ multiplex ไม่ได้
TaqMan probe	มีความจำเพาะสูง	มีค่าใช้จ่ายสูง
	ตรวจแบบ multiplex ได้	ทำ melting curve analysis ไม่ได้



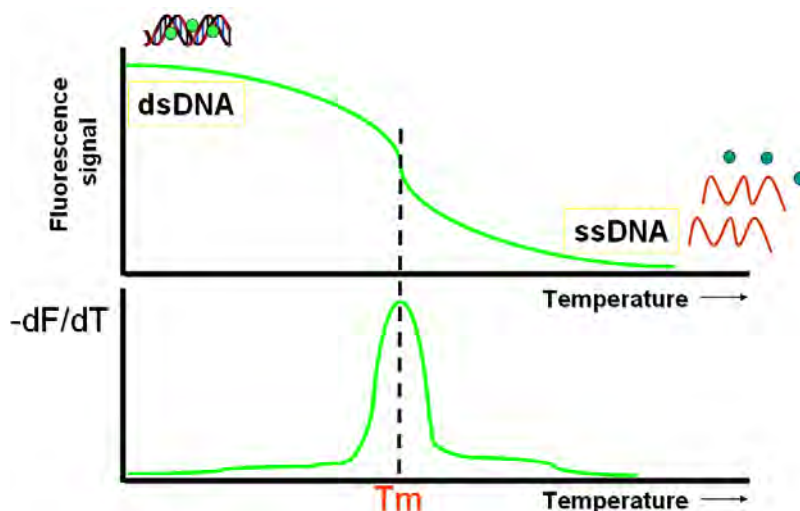
รูปที่ 6.13 หลักการทำงานของ SYBR Green I fluorescent dye



รูปที่ 6.14 ลักษณะโครงสร้าง และหลักการทำงานของ TaqMan probe

## Melting curve analysis

หลังจากการเพิ่มจำนวน DNA ด้วยเทคนิค real-time PCR ที่ใช้ SYBR Green I แล้วสามารถวิเคราะห์หาค่า melting temperature ( $T_m$ ) ซึ่งเป็นอุณหภูมิที่ทำให้ 50% ของ DNA แยกสายเป็น single stranded DNA โดยการใช้เทคนิค melting curve analysis ซึ่งทำได้โดยการค่อยๆ เพิ่มอุณหภูมิให้แก่ double stranded DNA พร้อมกับตรวจวัดปริมาณสารเรืองแสงที่ลดลงอย่างต่อเนื่อง ซึ่งที่ melting temperature จะเป็นอุณหภูมิซึ่งทำให้มีการลดลงของสารเรืองแสงอย่างมีนัยสำคัญ เนื่องจากที่อุณหภูมินี้จะทำให้ 50% ของ DNA แยกสายกลายเป็น single stranded DNA และทำให้โมเลกุล SYBR Green I หลุดออกจาก minor groove จึงมีการเรืองแสงลดลง เมื่อเขียนกราฟแสดงความสัมพันธ์ระหว่างอัตราการลดลงของสัญญาณเรืองแสงต่ออุณหภูมิ (แกน Y) และอุณหภูมิ (แกน X) จะได้เป็น melting curve (รูปที่ 15) และเมื่อลากเส้นจาก Melting peak ลงมาตัดแกน X ก็จะได้ค่า melting temperature ( $T_m$ ) ซึ่งค่า  $T_m$  ขึ้นกับปัจจัยต่าง ๆ ได้แก่ %GC content, amplicon length และ  $Mg^{2+}$  ดังนั้นเทคนิค melting curve analysis สามารถนำมาประยุกต์ใช้ในการแยกแยะ PCR product ที่มี %GC content หรือ amplicon length ต่างกันได้ เช่น การจำแนก primer dimers ออกจาก target gene การจำแนกชนิดของเชื้อไวรัสในตระกูลเดียวกัน การจำแนกสายพันธุ์ของเชื้อไวรัส และการตรวจสอบการเกิด insertion / deletion ของยีน เป็นต้น

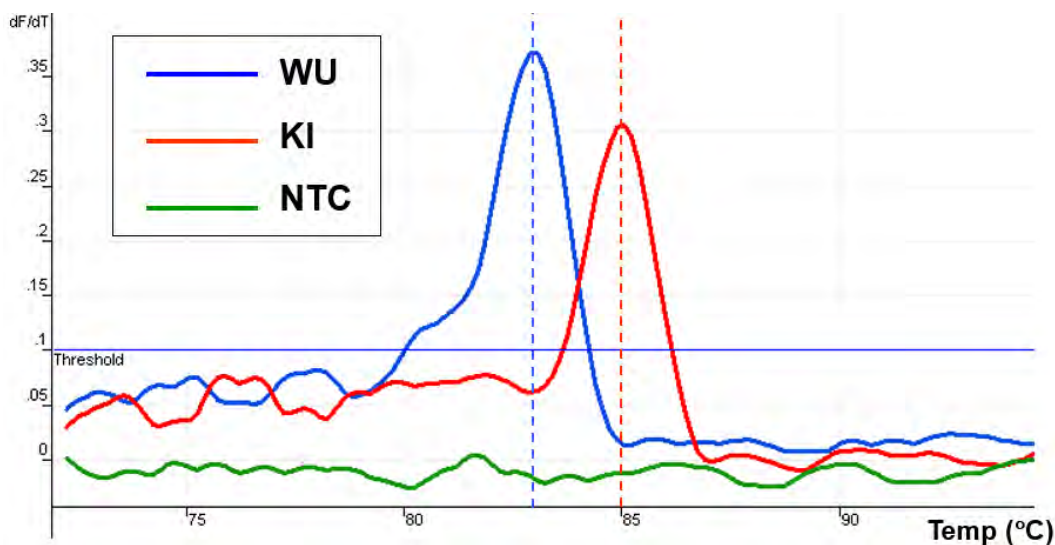


รูปที่ 6.15 หลักการหา melting temperature ( $T_m$ ) โดยการใช้ melting curve analysis

➤ ตัวอย่างการประยุกต์ใช้เทคนิค real-time PCR และ melting curve analysis

1. การตรวจวินิจฉัยเพื่อจำแนกระหว่าง WU และ KI polyomaviruses

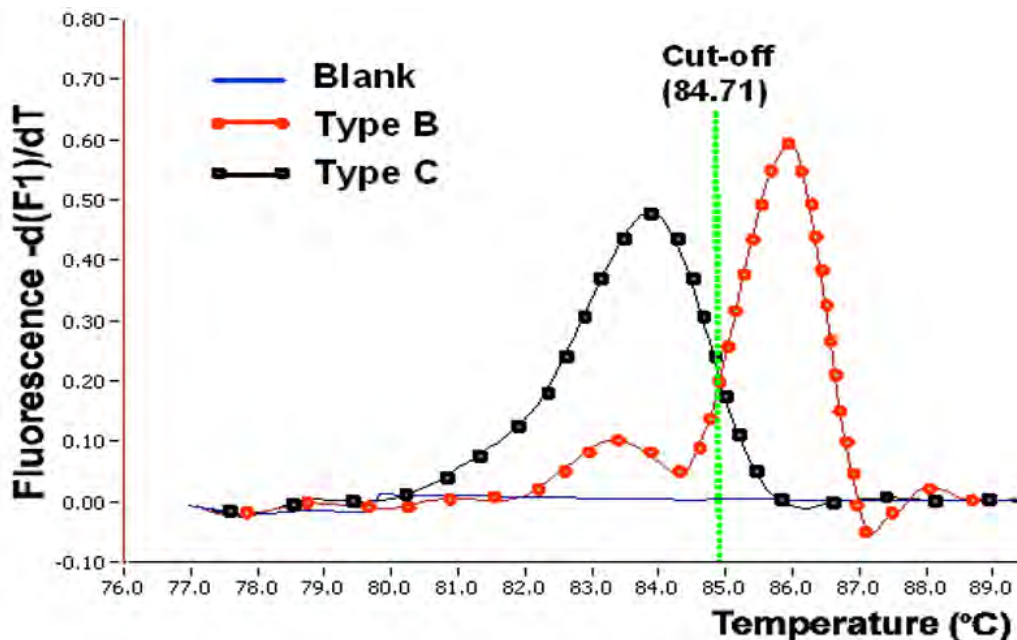
เนื่องจาก WU และ KI polyomaviruses เป็นเชื้อไวรัสอุบัติใหม่ที่อาจก่อโรคระบบทางเดินหายใจในมนุษย์ได้ ลักษณะทางไวรัสวิทยาของ WU และ KI polyomaviruses จัดอยู่ใน *Polyomaviridae* family เป็น non-enveloped virus ขนาดเส้นผ่าศูนย์กลางประมาณ 40-50 nm มีสารพันธุกรรมเป็น circular dsDNA ขนาดประมาณ 5 Kb ซึ่ง genome ของ WU และ KI polyomaviruses มีความเหมือนกันประมาณ 70% ดังนั้นจึงสามารถใช้ universal primers ที่จำเพาะในส่วน conserved region ของยีน VP1 ในการเพิ่มจำนวนยีนของทั้ง WU และ KI polyomaviruses ซึ่งจะมี %GC และ amplicon length แตกต่างกัน ทำให้สามารถใช้เทคนิค melting curve analysis ในการจำแนกชนิดของเชื้อไวรัส (WU หรือ KI) โดยอาศัยความต่างของ melting temperature (รูปที่ 6.16)



รูปที่ 6.16 การจำแนก WU และ KI polyomaviruses โดยใช้ melting curve analysis [จากผลงานวิจัยของ รศ.ดร.สัญญา พยุงกร <https://www.ncbi.nlm.nih.gov/pubmed/18639584>]

## 2. การตรวจวินิจฉัยเพื่อจำแนกระหว่าง genotype B และ C ของ hepatitis B virus

เชื้อไวรัสตับอักเสบบี (hepatitis B virus, HBV) มีลักษณะเป็นดีเอ็นเอไวรัสที่มีลักษณะเป็นวงกลมสายคู่แบบไม่สมบูรณ์ (partially double-stranded circular DNA) ขนาดประมาณ 3,200 bp ในปัจจุบันสามารถจำแนกตามความแตกต่างของลักษณะทางพันธุกรรมได้เป็น 10 genotypes (A-J) โดย genotypes ที่พบบ่อยในผู้ป่วยที่ติดเชื้อไวรัสตับอักเสบบีในประเทศไทย ได้แก่ genotype B และ C ซึ่งมีผลงานวิจัยพบว่าผู้ป่วยที่ติดเชื้อ HBV genotype C จะส่งผลทำให้มีอาการของโรคที่รุนแรงกว่า และมีความเสี่ยงสูงที่จะดำเนินโรคไปเป็นโรคตับแข็ง และมะเร็งตับ ดังนั้นในการตรวจวินิจฉัยทางคลินิกเพื่อจำแนก genotype ของเชื้อไวรัสตับอักเสบบี จึงมีความสำคัญในการพยากรณ์โรค และวางแผนการรักษาผู้ป่วย จากการศึกษาวิจัยพบว่ายีน *PreS1* มีความเหมาะสมในการจำแนก genotypes ของ HBV ดังนั้นจึงสามารถใช้ universal primers ที่จำเพาะในส่วน conserved region ในการเพิ่มจำนวนยีน *PreS1* ของทั้ง genotype B และ C ซึ่งจะมี %GC แตกต่างกัน ทำให้สามารถใช้เทคนิค melting curve analysis ในการจำแนก genotypes (B หรือ C) ของเชื้อไวรัส HBV ได้ (รูปที่ 6.17)

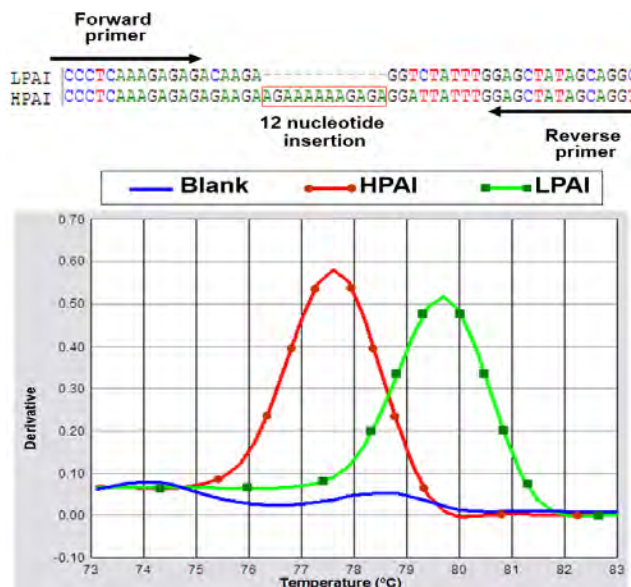


รูปที่ 6.17 การจำแนก genotypes ของเชื้อไวรัสตับอักเสบบี โดยใช้ melting curve analysis

[จากผลงานวิจัยของ รศ.ดร.สัญญาชัย พยุงกร <https://www.ncbi.nlm.nih.gov/pubmed/15288955>]



3. การตรวจวินิจฉัยเพื่อจำแนกระหว่าง HPAI และ LPAI ของ influenza A viruses จากการศึกษาเชื้อไวรัสไข้หวัดนก พบว่าสามารถแบ่งตามความรุนแรงของการก่อโรคได้เป็น 2 ชนิดคือ ชนิดก่อโรคไม่รุนแรง (low pathogenic avian influenza, LPAI) และชนิดก่อโรครุนแรง (high pathogenic avian influenza, HPAI) ซึ่งมีลักษณะที่แตกต่างกันในส่วนของ cleavage site บนโมเลกุลของ hemagglutinin (HA) โดยเชื้อไวรัสชนิดที่ก่อโรครุนแรง (HPAI) จะมี polybasic amino acid insertion (RRRK) ที่บริเวณ cleavage site ของโปรตีน HA ซึ่งส่งผลให้โปรตีน HA นี้ถูกตัดด้วยเอนไซม์ furin (ubiquitous proteases) ซึ่งจะพบได้ในทุกอวัยวะของร่างกาย เป็นเหตุให้มีการติดเชื้อแบบรุนแรงและมีการแพร่กระจายของเชื้อไวรัสไปทั่วทุกอวัยวะของร่างกาย (systemic infection) ดังนั้นเมื่อเปรียบเทียบลำดับนิวคลีโอไทด์ในส่วนของยีน HA ระหว่าง LPAI และ HPAI จะพบ nucleotide insertion (12 bp) บนยีน HA ของ HPAI ดังนั้นจึงสามารถใช้ universal primers ที่จำเพาะในส่วน conserved region ของยีน HA ในการเพิ่มจำนวนยีน ซึ่งจะมี %GC และ amplicon length แตกต่างกัน ทำให้สามารถใช้เทคนิค melting curve analysis ในการจำแนกระหว่าง LPAI และ HPAI ได้ (รูปที่ 6.18)



รูปที่ 6.18 การจำแนกระหว่าง LPAI และ HPAI โดยใช้ melting curve analysis

[จากผลงานวิจัยของ รศ.ดร.สัณชัย พยุงกร <https://www.ncbi.nlm.nih.gov/pubmed/16715581>]



## Multiplex real-time PCR

เทคนิค real-time PCR สามารถนำมาประยุกต์ใช้เป็นระบบ multiplex สำหรับการตรวจสอบหลายๆ ยีนพร้อมๆ กัน โดยทั่วไปเครื่อง real-time PCR จะมีตัวตรวจวัดสัญญาณเรืองแสง (fluorescent detectors) ประมาณ 4-6 ช่วงความยาวคลื่น ทำให้สามารถใช้ *TaqMan* probes ที่ติดฉลากด้วยสารเรืองแสงต่างชนิดกันในการทำ multiplex real-time PCR ได้ ซึ่งมีข้อควรระวังในการพิจารณาเลือกชนิดของสารเรืองแสง (fluorescent reporter dye) ดังนี้

- ควรเลือกใช้สารเรืองแสงชนิดที่เครื่อง real-time PCR ทั่วไปสามารถตรวจวัดได้ เนื่องจากเครื่อง real-time PCR แต่ละยี่ห้อ แต่ละรุ่น จะมีความสามารถในการตรวจวัดสัญญาณเรืองแสงจำนวนสีแตกต่างกัน เครื่องบางรุ่นตรวจวัดได้เพียง 4 สีเท่านั้น แต่เครื่องบางรุ่นสามารถตรวจวัดได้ 6 สี ดังนั้นจึงนิยมเลือกใช้สารเรืองแสงชนิดที่เป็นพื้นฐาน สามารถตรวจวัดได้ทุกเครื่อง เช่น FAM และ VIC เป็นต้น
- ควรเลือกใช้ชนิดของ quenchers ให้เหมาะสมกับชนิดของ reporter dyes หากเลือกใช้ quencher ที่ไม่เหมาะสมจะไม่สามารถบดบังสัญญาณเรืองแสงของ reporter dye ได้
- เนื่องจากในการปิเปตสารละลาย PCR master mix ใส่ในแต่ละหลอด อาจเกิดความผิดพลาด (pipetting errors) ทำให้การเปรียบเทียบสัญญาณเรืองแสงที่เกิดขึ้นในแต่ละหลอด (samples) มีความผิดพลาดได้ ดังนั้นนักวิจัยจึงนิยมใช้ ROX fluorescent dye เป็น passive reference dye ผสมใน PCR reagent master mix เพื่อใช้ในการ normalize fluorescent signal ทำให้สามารถเปรียบเทียบสัญญาณเรืองแสงระหว่างหลอด ได้แม่นยำมากขึ้น ด้วยเหตุนี้จึงไม่ควรเลือกใช้ ROX เป็น reporter dye ของ *TaqMan* probe ในกรณีที่ใช้ ROX เป็น passive reference dye แล้ว
- MGB (minor groove binder) จัดเป็น non-fluorescent quencher ที่มีคุณสมบัติพิเศษคือ สามารถใช้ได้กับ reporter dyes ทุกชนิด และเพิ่มเสถียรภาพของ probe ทำให้จับกับ DNA template ได้ดีขึ้นด้วย ดังนั้น *TaqMan*-MGB probe (15-20 bp) จะมีความยาวสั้นกว่า *TaqMan* probe ทั่วไป จึงนิยมใช้ใน allelic discrimination assay ซึ่งจะทำให้จำแนก alleles ได้ชัดเจนกว่า *TaqMan* probe ทั่วไป

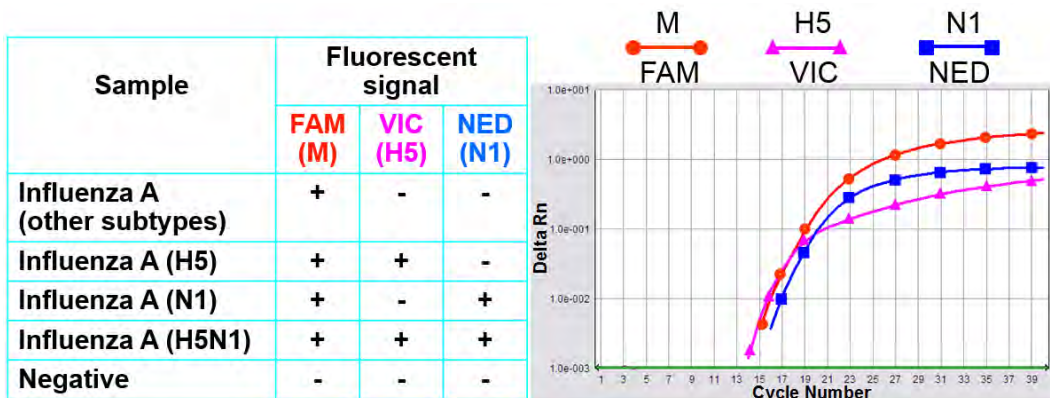
➤ ตัวอย่างการประยุกต์ใช้เทคนิค multiplex real-time PCR

1. การตรวจวินิจฉัยเชื้อไวรัสไข้หวัดนก สายพันธุ์ H5N1

เชื้อไวรัสไข้หวัดนก (H5N1 avian influenza A virus) จัดอยู่ใน *Orthomyxoviridae* family มีสารพันธุกรรมเป็น negative single strand RNA ที่มีลักษณะเป็นท่อน ๆ จำนวนทั้งสิ้น 8 ยีน ซึ่งเชื้อชนิดนี้สามารถแบ่งออกเป็น subtype ตามชนิดของแอนติเจน hemagglutinin (HA) และ neuraminidase (NA) ดังนั้นในการตรวจวินิจฉัยเชื้อไวรัสไข้หวัดนกสายพันธุ์ H5N1 จึงต้องมีการตรวจสอบหลายยีนดังนี้

- Matrix (M) เพื่อใช้ในการตรวจสอบว่าเป็น influenza A virus
- Hemagglutinin (HA) เพื่อใช้ในการระบุสายพันธุ์ H5
- Neuraminidase (NA) เพื่อใช้ในการระบุสายพันธุ์ N1

ดังนั้นในการประยุกต์ใช้เทคนิค real-time PCR ในการตรวจวินิจฉัยเชื้อไวรัสไข้หวัดนกสายพันธุ์ H5N1 จึงใช้ primers จำนวน 3 คู่ และ TaqMan probes จำนวน 3 ตัว ที่ติดฉลากด้วยสารเรืองแสงที่มีสีแตกต่างกันสำหรับตรวจสอบยีน M, H5 และ N1 พร้อมกันใน tube เดียว ทำให้สามารถตรวจวินิจฉัยและแปลผลการติดเชื้อไวรัสไข้หวัดนกสายพันธุ์ H5N1 ได้อย่างรวดเร็ว (รูปที่ 6.19)



รูปที่ 6.19 การตรวจเชื้อไวรัสไข้หวัดนก ด้วยเทคนิค multiplex real-time PCR

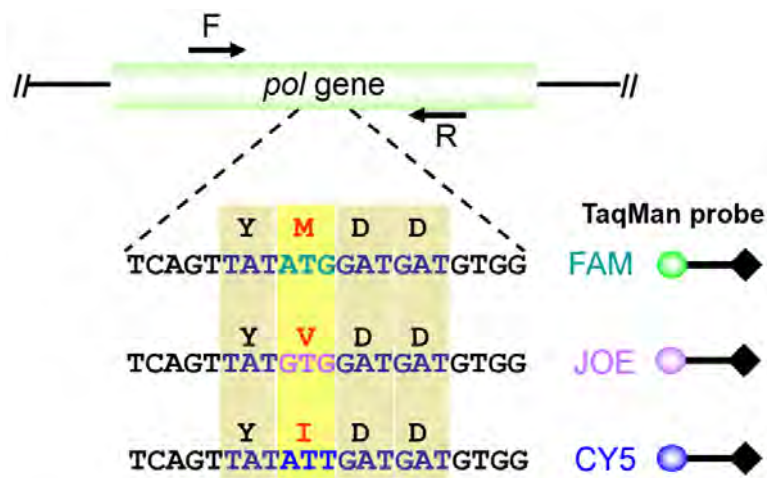
[จากผลงานวิจัยของ รศ.ดร.สัญญา พยุงกร <https://www.ncbi.nlm.nih.gov/pubmed/16183140>]

2. การตรวจวินิจฉัยเชื้อไวรัสตับอักเสบบี ที่ดื้อต่อยาต้านไวรัส lamivudine

Lamivudine เป็นยาต้านไวรัส (anti-viral drug) จำพวก nucleoside analogue ที่สามารถยับยั้ง reverse-transcriptase activity ของเอนไซม์ polymerase (Pol) ของเชื้อไวรัสตับอักเสบบีได้ อย่างไรก็ตามเมื่อใช้ยาดังกล่าวในการรักษาผู้ป่วยติดเชื้อไวรัสตับอักเสบบีไประยะหนึ่ง ตัวเชื้อไวรัสจะมีการปรับตัวในการกลายพันธุ์เพื่อดื้อต่อยา (drug resistant mutants) จากการศึกษาวิจัยพบว่าสามารถใช้ลำดับนิวคลีโอไทด์บนยีน polymerase ของเชื้อไวรัสตับอักเสบบีในการพิจารณาว่าเชื้อไวรัสที่พบอยู่ในร่างกายของผู้ป่วยจะตอบสนอง (sensitive) หรือดื้อ (resistant) ต่อยา lamivudine ได้ ดังนี้

- YMDD เป็น motif ที่พบใน lamivudine sensitive HBV
- YVDD เป็น motif ที่พบใน lamivudine resistant mutant HBV
- YIDD เป็น motif ที่พบใน lamivudine resistant mutant HBV

ดังนั้นในการประยุกต์ใช้เทคนิค real-time PCR ในการตรวจวินิจฉัยลักษณะการดื้อยา lamivudine ของเชื้อไวรัสตับอักเสบบี จึงอาศัย primers จำนวน 1 คู่ และ TaqMan probes จำนวน 3 ตัว ที่ติดฉลากด้วยสารเรืองแสงที่มีสีแตกต่างกันสำหรับตรวจสอบ YMDD, YVDD และ YIDD พร้อมกันในหลอดเดียว ทำให้สามารถตรวจวินิจฉัยและแปลผลการดื้อยา lamivudine ของเชื้อไวรัสตับอักเสบบีในผู้ป่วย ได้อย่างรวดเร็ว (รูปที่ 6.20)

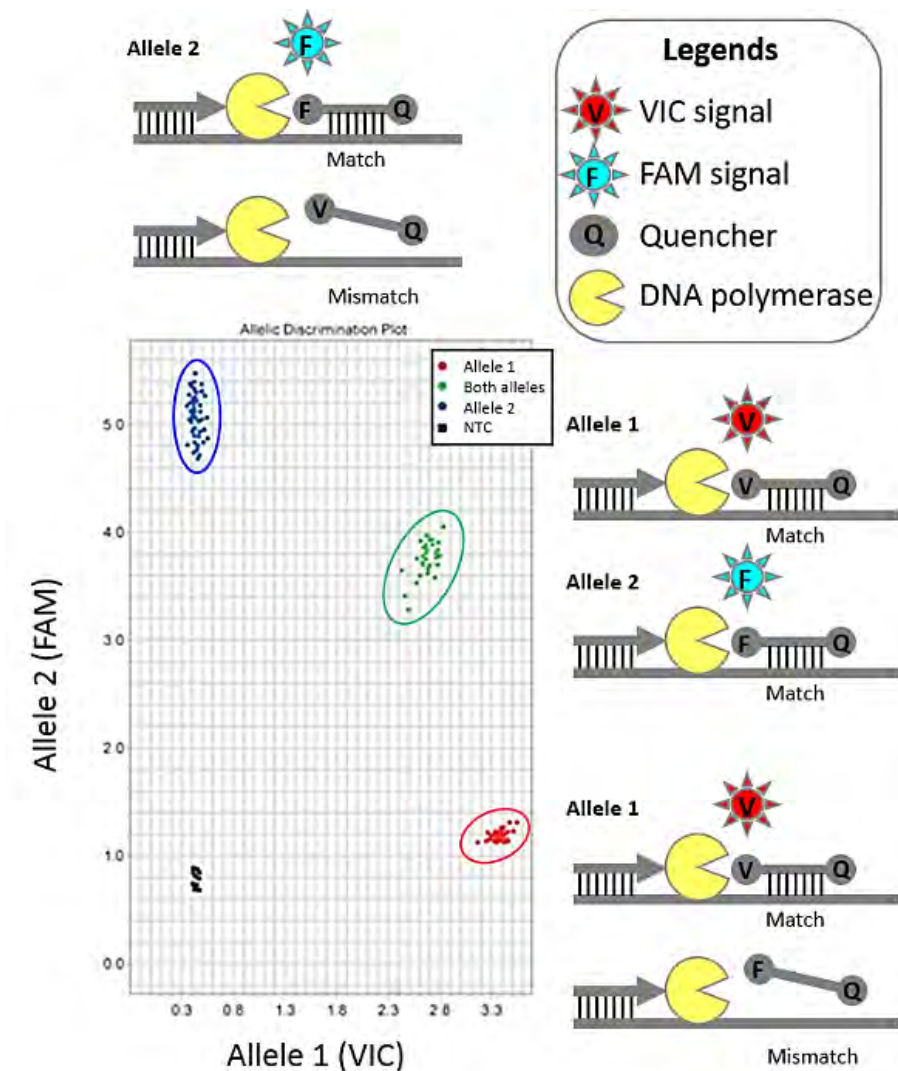


รูปที่ 6.20 การตรวจจลัษณะการดื้อยา lamivudine ด้วยเทคนิค multiplex real-time PCR

### Allelic discrimination assay

เป็น multiplex real-time PCR เพื่อใช้ในการตรวจสอบ alleles (SNPs หรือ point mutation) ของยีนได้ โดยอาศัย TaqMan probes 2 ตัวที่ติดฉลากด้วยสารเรืองแสงต่างชนิดกัน (รูปที่ 6.19) ดังนี้

- Allele 1 (wild type) specific probe ติดฉลากด้วย VIC
- Allele 2 (mutant) specific probe ติดฉลากด้วย FAM



รูปที่ 6.19 หลักการจำแนก genotypes ด้วย allelic discrimination plot

เมื่อทำการเพิ่มจำนวนยีนที่สนใจด้วย real-time PCR ที่ใส่ TaqMan probes 2 ตัวในปฏิกิริยาเสร็จแล้ว โปรแกรมจะทำการสร้างกราฟ allelic discrimination plot โดยให้แกน X เป็นสัญญาณเรืองแสงชนิด VIC และแกน Y เป็นสัญญาณเรืองแสงชนิด FAM ตามลำดับ ซึ่งสามารถแปลผลได้ (ตารางที่ 6.4) ดังนี้

**ตารางที่ 6.4** ตัวอย่างการแปลผล genotypes ด้วยเทคนิค allelic discrimination plot

Sample	Fluorescent signal		Interpretation of genotypes
	VIC	FAM	
A	+++	-	Homozygous allele 1 (wild type)
B	-	+++	Homozygous allele 2 (mutant)
C	++	++	Heterozygous alleles

ในปัจจุบัน real-time PCR เป็นเทคนิคที่ได้รับความนิยมเป็นอย่างสูง ในการประยุกต์ใช้สำหรับการตรวจวินิจฉัยโรคติดเชื้อ เป็นเนื่องจากเป็นเทคนิคที่มีลักษณะเด่นดังนี้

- Rapid (less time consuming) ได้ผลการตรวจสอบที่รวดเร็ว สามารถวิเคราะห์ผลได้ภายใน 3-5 ชั่วโมง
- No post-PCR processing step ไม่จำเป็นต้องนำ PCR product มาแยกด้วย agarose gel electrophoresis จึงเป็นการลดความเสี่ยงในการปนเปื้อน เนื่องจากเป็นวิธีที่มีกระบวนการขั้นตอนเดียว
- High specificity มีความจำเพาะสูง เนื่องจากอาศัย primers และ probes ที่มีความจำเพาะต่อยีนที่ต้องการตรวจสอบ
- High sensitivity มีความไวในการตรวจสอบสูงเนื่องจากอาศัยหลักการตรวจสอบสารเรืองแสง (fluorescent signal) ขณะที่มีการเพิ่มจำนวนของสารพันธุกรรมดังนั้นจึงสามารถตรวจสอบได้แม้สิ่งส่งตรวจจะมีปริมาณ DNA อยู่เล็กน้อยก็ตาม
- High accuracy and reproducibility ให้ผลการตรวจวินิจฉัยที่แม่นยำ และสามารถตรวจสอบซ้ำแล้วให้ผลเหมือนเดิม

- Multi-color detection สามารถตรวจวัดหลายๆ ยีนพร้อมกันได้ (multiplex) ในกรณีที่ใช้ TaqMan probe ที่ติดฉลากด้วยสารเรืองแสงต่างสีกัน
- Melting curve analysis สามารถใช้ในการแยกแยะยีนที่มี Tm แตกต่างกันในกรณีที่ใช้ SYBR Green I
- Quantitative data สามารถวัดปริมาณเชิงเปรียบเทียบ หรือวัดปริมาณสัมบูรณ์ได้

#### เอกสารอ้างอิง และแนะนำอ่านเพิ่มเติม

1. Arya M, Shergill IS, Williamson M, Gommersall L, Arya N, Patel HR. Basic principles of real-time quantitative PCR. *Expert Rev Mol Diagn.* 2005 Mar;5(2):209-19. Review.
2. Chanthra N, Payungporn S, Chuaypen N, Piratanantatavorn K, Pinjaroen N, Poovorawan Y, Tangkijvanich P. Single Nucleotide Polymorphisms in STAT3 and STAT4 and Risk of Hepatocellular Carcinoma in Thai Patients with Chronic Hepatitis B. *Asian Pac J Cancer Prev.* 2015;16(18):8405-10.
3. Chieochansin T, Chutinimitkul S, Payungporn S, Theamboonlers A, Tangkijvanich P, Komolmit P, Poovorawan Y. Rapid detection of lamivudine-resistant hepatitis B virus mutations by PCR-based methods. *Tohoku J Exp Med.* 2006 Sep;210(1):67-78.
4. Payungporn S, Tangkijvanich P, Jantaradsamee P, Theamboonlers A, Poovorawan Y. Simultaneous quantitation and genotyping of hepatitis B virus by real-time PCR and melting curve analysis. *J Virol Methods.* 2004 Sep 15;120(2):131-40.

5. Payungporn S, Phakdeewirot P, Chutinimitkul S, Theamboonlers A, Keawcharoen J, Oraveerakul K, Amonsin A, Poovorawan Y. Single-step multiplex reverse transcription-polymerase chain reaction (RT-PCR) for influenza A virus subtype H5N1 detection. *Viral Immunol.* 2004;17(4):588-93.
6. Payungporn S, Chutinimitkul S, Chaisingh A, Damrongwattanapokin S, Buranathai C, Amonsin A, Theamboonlers A, Poovorawan Y. Single step multiplex real-time RT-PCR for H5N1 influenza A virus detection. *J Virol Methods.* 2006 Feb;131(2):143-7.
7. Payungporn S, Chutinimitkul S, Chaisingh A, Damrongwattanapokin S, Nuansrichay B, Pinyochon W, Amonsin A, Donis RO, Theamboonlers A, Poovorawan Y. Discrimination between highly pathogenic and low pathogenic H5 avian influenza A viruses. *Emerg Infect Dis.* 2006 Apr;12(4):700-1.
8. Payungporn S, Chieochansin T, Thongmee C, Panjaworayan N, Samransamruajkit R, Theamboonlers A, Poovorawan Y. Detection and discrimination of WU/KI polyomaviruses by real-time PCR with melting curve analysis. *J Virol Methods.* 2008 Oct;153(1):70-3.
9. Payungporn S, Poomipak W, Makkoch J, Rianthavorn P, Theamboonlers A, Poovorawan Y. Detection of oseltamivir sensitive/resistant strains of pandemic influenza A virus (H1N1) from patients admitted to hospitals in Thailand. *J Virol Methods.* 2011 Nov;177(2):133-9.
10. Suwannakarn K, Chutinimitkul S, Payungporn S, Chieochansin T, Theamboonlers A, Amonsin A, Damrongwattanapokin S, Mai le Q, Hanh NH, Poovorawan Y. Assay to detect H5N1 oseltamivir resistance. *Emerg Infect Dis.* 2006 Dec;12(12):1995-6.





# 7

การออกแบบไพรเมอร์ด้วยคอมพิวเตอร์

Computational based primers design

## การออกแบบไพรเมอร์ด้วยคอมพิวเตอร์

### Computational based primers design

จากบทที่แล้วจะเห็นได้ว่าเทคนิคที่เกี่ยวข้องกับพีซีอาร์ (PCR-related techniques) ได้รับความนิยมเป็นอย่างมากทั้งในด้านงานวิจัย และการประยุกต์ใช้ในการตรวจวินิจฉัยโรคต่างๆ ดังนั้นในบทนี้จึงมุ่งเน้นเกี่ยวกับการออกแบบต่างๆ ที่เกี่ยวข้องกับเทคนิค PCR ซึ่งจะช่วยทำให้สามารถพัฒนาเทคนิค PCR ในการเพิ่มจำนวนยีนที่สนใจได้อย่างมีประสิทธิภาพ

#### การออกแบบ thermal profiles สำหรับ PCR

Thermal profiles คือรูปแบบของค่าอุณหภูมิ และเวลา สำหรับปฏิกิริยา PCR ที่เรากำหนดในเครื่องเพิ่มจำนวนสารพันธุกรรม (thermocycler) ซึ่ง conventional PCR โดยทั่วไป จะกำหนด thermal profiles พื้นฐาน ดังตารางที่ 7.1

ตารางที่ 7.1 Thermal profiles พื้นฐานที่ใช้สำหรับ PCR

Step	Temp (°C)	Time	Cycle
Initial denaturation	94	3 - 10 min	1
Denature	94	15 - 30 sec	35-40
Annealing	See Tm of primer	15 - 30 sec	
Extension	72	See product size	
Final extension	72	3 - 10 min	1

จากตารางที่ 7.1 จะเห็นได้ว่า thermal profiles จะมีตัวแปร 2 ปัจจัย ได้แก่

- Annealing temperature อุณหภูมิที่ใช้จะขึ้นอยู่กับค่า melting temperature (Tm) ของ primer คำนวณได้จาก สูตร  $T_m = 2 \times (\text{no.of [A+T]}) + 4 \times (\text{no.of [G+C]})$  โดยที่ Annealing temperature =  $T_m - 5$
- Extension time เวลาที่ใช้จะขึ้นอยู่กับความยาวของ PCR product โดยประมาณได้จาก PCR product ขนาด 1,000 bp ใช้เวลาประมาณ 1 นาที (ขึ้นกับเอนไซม์ที่ใช้)

ขนาดความยาวของ PCR product จะขึ้นอยู่กับระยะห่างระหว่าง forward primer และ reverse primer ควรคำนึงถึงขนาดความยาวของ PCR product ให้เหมาะสมด้วย ดังนี้

- ถ้า PCR product ขนาดใหญ่ (>1kb) จะมีความไว (sensitivity) ในการตรวจสอบลดลง เพราะจะได้ band ของ PCR product บาง หรืออาจไม่เห็น band เนื่องจากสาเหตุ ดังนี้
  - DNA template คุณภาพไม่ดี มีการขาดเป็นท่อนสั้นๆ
  - dNTPs ไม่เพียงพอในการต่อสาย DNA ขนาดใหญ่
  - Inhibitor ปนเปื้อนอยู่ใน DNA ทำให้ปฏิกิริยาการเพิ่มจำนวนเกิดได้ไม่ดี
  - Activity ของเอนไซม์เสื่อมสภาพ ทำให้การเพิ่มจำนวน DNA เกิดได้ไม่ดี
- ถ้า PCR product ขนาดเล็ก (<100 bp) จะตรวจสอบด้วย agarose gel electrophoresis ได้ยาก เพราะ band ของ PCR product จะไปซ้อนทับกับ primer-dimer หรือสีย้อม loading dye ทำให้มองเห็น band ไม่ชัด

### หลักการออกแบบ primers

หัวใจสำคัญของเทคนิค PCR ก็คือ oligonucleotide primers ที่มีความจำเพาะต่อยีนที่ต้องการเพิ่มจำนวน ซึ่ง primers สำหรับ conventional PCR ควรมีลักษณะพื้นฐานต่างๆ ดังนี้

1. ความยาวของ oligonucleotide primer ควรอยู่ระหว่าง 18-30 bp
2. %GC content ของ oligonucleotide primer ควรอยู่ในช่วง 40-60%
3. Melting temperature (T<sub>m</sub>) ของ primer ควรอยู่ในช่วง 55-66°C ซึ่งคำนวณได้จากสูตร  $T_m = 2 \times (\text{no. of [A+T]}) + 4 \times (\text{no. of [G+C]})$
4. T<sub>m</sub> ของ forward และ reverse primers ควรมีค่าเท่ากัน หรือต่างกันไม่เกิน 2°C
5. ในกรณีที่ T<sub>m</sub> ของ forward และ reverse primers มีค่าต่างกัน ควรใช้ T<sub>m</sub> ที่มีค่าน้อยในการคำนวณค่า annealing temperature
6. Primer ควร perfect complementary และจำเพาะกับยีนที่สนใจเท่านั้น ควรนำไปวิเคราะห์ด้วยการ BLAST เพื่อค้นหาว่า primer คล้ายคลึงกับยีนอื่นๆ หรือไม่

หากเป็นไปได้ ควรมีลักษณะเพิ่มเติม (optional) ดังนี้

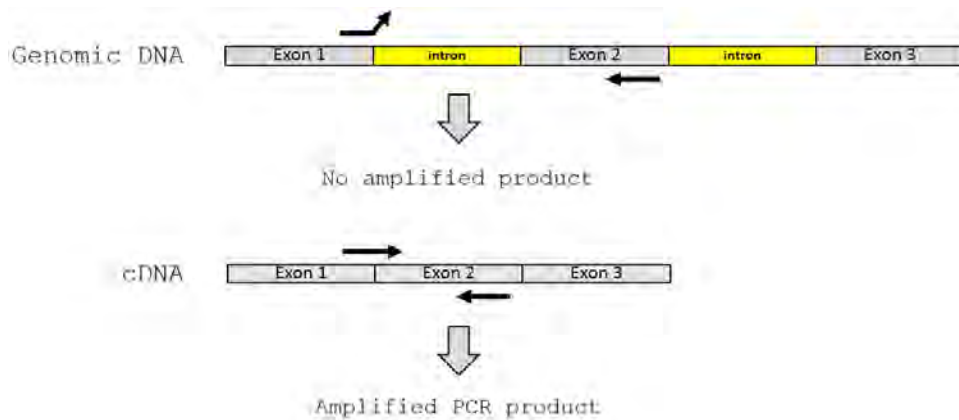
1. หากมีเบส G หรือ C ที่ปลายด้าน 3' ของ primer จะช่วยเพิ่ม priming efficiency
2. หลีกเลี่ยง GC clamp (G หรือ C มากกว่า 3 ตัว) ในบริเวณ 5 เบสสุดท้ายที่ปลายด้าน 3'
3. หลีกเลี่ยงลำดับเบสเดิมซ้ำๆ มากกว่า 3 ตัว เช่น AAA, TTTT, CCCCC เป็นต้น
4. หากจำเป็นต้องมี mismatch ระหว่าง primer และ DNA template ควรเลี่ยง mismatch บริเวณ 5 bp สุดท้ายที่ปลายด้าน 3' ของ primer
5. หลีกเลี่ยงการเกิด secondary structure ของ primers (รูปที่ 7.1)



รูปที่ 7.1 Secondary structure ของ primers

นอกจากนี้ยังมีเงื่อนไขที่ควรพิจารณาประกอบในการออกแบบ primers ในกรณีต่างๆ ดังนี้

- Conventional PCR ควรออกแบบ primers ในการเพิ่มจำนวนยีนเป้าหมาย ให้ได้ขนาดของ PCR product ประมาณ 200-800 bp ที่สามารถตรวจสอบได้ชัดเจนด้วย agarose gel electrophoresis
- RT-PCR ควรออกแบบเป็น primer spans intron/exon boundary คือออกแบบ primer ให้จับบริเวณรอยต่อระหว่าง exons เพื่อให้สามารถเพิ่มจำนวนยีนที่มาจาก cDNA template เท่านั้น แต่จะไม่เพิ่มจำนวนยีนที่มาจาก genomic DNA (รูปที่ 7.2) เป็นการป้องกันการเพิ่มจำนวน genomic DNA contamination ที่ปะปนอยู่กับ cDNA
- Multiplex PCR ควรออกแบบ primers ให้มีลักษณะ ดังนี้
  - Primers ทุกตัวมีค่า  $T_m$  เท่ากัน หรือใกล้เคียงกัน
  - ไม่เกิด primer-dimer ระหว่าง primers ทุกตัว
  - PCR product ของ primer แต่ละคู่ ควรมีความต่างกันอย่างน้อย 50 bp เพื่อให้สามารถแยกและแปลผลได้ชัดเจนด้วย agarose gel electrophoresis



รูปที่ 7.2 Primer spans intron/exon boundary สำหรับ RT-PCR

- PCR-RFLP ควรออกแบบ primers ให้ได้ขนาดของ PCR product ประมาณ 300-500 bp และแปลผลความแตกต่างได้ชัดเจน ระหว่าง PCR product ที่ถูกตัด และไม่ถูกตัด ด้วยเอนไซม์ตัดจำเพาะ
- Allele specific PCR ควรออกแบบ allele specific primer โดยให้ตำแหน่งที่จะใช้ในการเลือกจับระหว่าง wild type และ mutant alleles อยู่ที่ปลายด้าน 3' ของ primer
- Real-time PCR ควรออกแบบ primers ให้ได้ PCR product ขนาดเล็ก (100-250 bp) จะทำให้มีประสิทธิภาพในการเพิ่มจำนวนสูง และจะได้กราฟ amplification plot เป็น sigmoid curve เป็นการเพิ่มความไว (sensitivity) ของการตรวจสอบ
- Real-time PCR ที่ใช้ TaqMan probe ในการตรวจวัด มีหลักการออกแบบ probe ดังนี้
  - จำเพาะต่อยีนเป้าหมาย ตรวจสอบได้โดยการ BLAST
  - Tm ของ probe ควรมากกว่า Tm ของ primers ประมาณ 8-10°C
  - หลีกเลี่ยงเบส G ที่ตำแหน่งปลาย 5' เพื่อไม่ให้เกิด quenching effect
  - หลีกเลี่ยงลำดับเบสเดิมซ้ำๆ มากกว่า 3 ตัว เช่น AAA, TTTT, CCCCC เป็นต้น
- Multiplex real-time PCR ควรเลือกใช้ fluorescent reporter dye และ quencher ที่เหมาะสมกัน (ตารางที่ 7.1) และควรเลือกใช้ fluorescent reporter dye ชนิดที่เครื่อง real-time PCR สามารถตรวจวัดได้

ตารางที่ 7.1 การเลือกใช้ reporter dyes และ quenchers สำหรับ multiplex real-time PCR

Reporter dye	Quenchers					Absorption (nm)	Emission (nm)	Color
	TAMRA	BHQ1	BHQ2	BHQ3	MGB			
FAM	✓	✓	⊗	⊗	✓	495	520	Green
TET	✓	✓	⊗	⊗	✓	521	536	Green
JOE / VIC	✓	✓	✓	⊗	✓	520	548	Yellow
MAX	✓	✓	✓	⊗	✓	525	555	Yellow
ROX	⊗	⊗	✓	✓	✓	576	601	Orange
CY5	⊗	⊗	✓	✓	✓	643	667	Red

- Multiplex real-time PCR สำหรับ allelic discrimination plot นิยมใช้เป็น *TaqMan* MGB probes ซึ่ง MGB จัดเป็น non-fluorescent quencher ที่มีคุณสมบัติพิเศษคือสามารถใช้ได้กับ reporter dyes ทุกชนิด และเป็น minor groove binder ที่สามารถเพิ่มเสถียรภาพของ probe ทำให้จับกับ DNA template ได้ดีขึ้นด้วย ดังนั้น *TaqMan* MGB probe (15-20 bp) มักจะมีความยาวสั้นกว่า *TaqMan* probe ทั่วไป จึงนิยมใช้ใน allelic discrimination assay ซึ่งตำแหน่งนิวคลีโอไทด์ที่แตกต่างกันระหว่าง allele 1 (wild type) และ allele 2 (mutant) ควรจะอยู่ที่ตำแหน่งกึ่งกลางของ probe เช่น หาก *TaqMan* MGB probe มีความยาว 17 bp ตำแหน่งที่จะใช้แยกระหว่าง allele 1 และ 2 ควรจะอยู่ในตำแหน่งที่ 9 ของ probe เป็นต้น



## โปรแกรมคอมพิวเตอร์สำหรับการออกแบบ primers

ในปัจจุบันมีโปรแกรมคอมพิวเตอร์สำหรับการออกแบบ primers โดยผู้ใช้งานสามารถใส่ลำดับนิวคลีโอไทด์ของยีนที่สนใจเข้าสู่โปรแกรม จากนั้นตั้งค่าต่างๆ ตามที่ต้องการ แล้วสั่งให้โปรแกรมทำการคัดเลือกและออกแบบ primers ที่มีลักษณะเหมาะสมได้ ทั้งนี้แต่ละโปรแกรมหรือแต่ละ website ก็จะมีหลักการเบื้องต้นในการออกแบบคล้ายๆ กัน แต่จะมีความแตกต่างกันไปในส่วนของ default parameters ต่างๆ ดังนั้นผู้ใช้งานจึงควรตระหนักถึงประเด็นที่มีความแตกต่างกันด้วย ตัวอย่างของโปรแกรมที่นิยมใช้ในการออกแบบ primers มีดังนี้

### ➤ โปรแกรม Primer3web [<http://primer3.ut.ee>]

เป็นโปรแกรม free online web-based สำหรับการออกแบบ primers และ hybridization probe โดยผู้ใช้งานสามารถใส่ข้อมูลลำดับนิวคลีโอไทด์ของยีนที่ต้องการใช้เป็น template ในการออกแบบ primer (รูปที่ 7.3) จากนั้นตั้งค่า parameters ต่างๆ ที่สำคัญ ได้แก่

- การกำหนดลำดับนิวคลีโอไทด์บริเวณที่สนใจเพื่อใช้ในการออกแบบ primers ได้แก่
  - Targets กำหนดเป้าหมายบริเวณที่สนใจ (เช่น ตำแหน่งที่เกิดการกลายพันธุ์) เพื่อให้โปรแกรมออกแบบ primers คร่อมบริเวณดังกล่าว
  - Overlap Junction List กำหนดช่วงที่จะออกแบบ primer ให้จับบริเวณรอยต่อระหว่าง exons เพื่อใช้เป็น primers สำหรับ cDNA template
  - Excluded Regions กำหนดช่วงที่ไม่สนใจ ในการออกแบบ primer
  - Included Region กำหนดช่วงที่สนใจ ในการออกแบบ primer
  - Force Left Primer Start กำหนดช่วงเริ่มต้นของการค้นหา forward primer
  - Force Left Primer End กำหนดช่วงสิ้นสุดของการค้นหา forward primer
  - Force Right Primer Start กำหนดช่วงเริ่มต้นของการค้นหา reverse primer
  - Force Right Primer End กำหนดช่วงสิ้นสุดของการค้นหา reverse primer


- การกำหนดคุณสมบัติของ primers (General Primer Picking Conditions)
  - Primer Size กำหนดช่วงความยาวของ primers ที่ต้องการ
  - Primer Tm กำหนดช่วง Tm ของ primers ที่ต้องการ
  - Max Tm Difference กำหนดค่าความต่างสูงสุดของ Tm ระหว่างคู่ primers
  - Primer GC% กำหนดช่วง GC% ของ primers ที่ต้องการ
  - Product Size Ranges กำหนดช่วงความยาวของ PCR product ที่ต้องการ
  - Number To Return กำหนดจำนวน candidate primers ที่ต้องการให้แสดง

โดยทั่วไปมักนิยมใช้ค่า default ของโปรแกรม แล้วกำหนดเพียงปัจจัยที่เราต้องการเป็นพิเศษ เช่น Primer Tm, Max Tm Differences และ Product Size Ranges เมื่อตั้งค่าต่างๆ เสร็จแล้ว คลิกที่คำสั่ง Pick Primers โปรแกรมก็จะประมวลผล และแสดงผลการออกแบบ primers (รูปที่ 7.4)

**Primer3web** version 4.1.0 - Pick primers from a DNA sequence.

[disclaimer](#)   [code](#)  
[cautions](#)

Select the [Task](#) for primer selection generic

[Template masking](#) before primer design ([available species](#))


[Select species](#) Example: Mus musculus

[Nucleotides to mask in 5' direction](#) 1

[Primer failure rate cutoff](#) < 0.1

[Nucleotides to mask in 3' direction](#) 0

Paste source sequence below (5'→3', string of ACGTNacgtn -- other letters treated as N -- numbers and blanks ignored). FASTA format ok. Please N-out undesirable sequence (vector, ALUs, LINEs, etc.) or use a [Mispriming Library \(repeat library\)](#). NONE

ATGGAACCATTAAAGACAGGAGCCCATATCGAACCCATGAGCTGTCCTATTGGTGAAGTTCCTCTCCATAC  
 AACTCAAGATTTGAGTCAGTCGCTTGGTCAGCAAGTGCTTGCATGATGGCATCAATTGGCTAACAAATTGGAAT  
 TTCTGGCCAGACAAATGGGGCAGTGGCTGTGTTAAAGTACAACGGCATAAATACAGACACTATCAAGAGTTGG  
 AGAAACAATATATTGAGAACAACAAGAGTCTGAATGTGCATGTGTAATGGTCTTCTTACTGTAATGACCGA  
 TGGACCAAGTGTGGACAGCCCTACACAAGATCTTCAGAATAGAAAAGGGAAGATAATTAATCAGTCGAAA  
 TGAATGCCCTAATTACTATGAGGAATGCTCCTGTTATCCTGATTCTAGTGAATCACATGTGTGTCAGG

Pick left primer, or use left primer below

Pick hybridization probe (internal oligo), or use oligo below

Pick right primer, or use right primer below (5' to 3' on opposite strand)

Pick Primers
Download Settings
Reset Form

รูปที่ 7.3 โปรแกรม Primer3web [http://primer3.ut.ee]



➤ โปรแกรม Primer3Plus [https://primer3plus.com/cgi-bin/dev/primer3plus.cgi]

เป็นโปรแกรม free online web-based สำหรับการออกแบบ primers และ hybridization probe ซึ่งมีการพัฒนาต่อยอดมาจากโปรแกรม Primer3Web เพื่อให้ผู้ใช้สามารถใช้งานได้ง่ายขึ้น (user friendly) ทั้งนี้โปรแกรม Primer3Plus มีการแบ่งค่า parameters ออกเป็นหมวดหมู่ ได้แก่ Main (รูปที่ 7.5), General Setting (รูปที่ 7.6), Advanced Setting (รูปที่ 7.7), Internal Oligo (รูปที่ 7.8), Penalty Weights (รูปที่ 7.9) และ Advanced Sequence (รูปที่ 7.10)

Primer3Plus		<a href="#">More...</a>	<a href="#">Source Code</a>
pick primers from a DNA sequence		<a href="#">Help</a>	<a href="#">About</a>
Load server settings: <input type="text" value="Default"/>	<input type="button" value="Activate Settings"/>	Select primer pairs to detect the given template sequence. Optionally targets and included/excluded regions can be specified.	
Task: <input type="text" value="generic"/>	<input type="button" value="Pick Primers"/>	<input type="button" value="Reset Form"/>	
<b>Main</b>	<b>General Settings</b>	<b>Advanced Settings</b>	<b>Internal Oligo</b>
<b>Penalty Weights</b>	<b>Advanced Sequence</b>		
Sequence Id: <input type="text"/>			
Paste template sequence below		Or upload sequence file: <input type="text"/> <input type="button" value="Browse..."/> <input type="button" value="Upload File"/>	
<pre> ATGGAACCATTAAGACAGGAGCCCATATCGAACCCCTAATGAGCTGTCCTATTGGTGAAGTCCCTCTCCATACAACTCAAGATTGAGTCAGTCGCCTGGTC AGCAAGTGCCTTGTTCATGATGGCATCAATTGGCTAACCAATTGGAATTTCTGGCCAGACAATGGGCGAGTGGCTGTGTTAAAGTACAACGGCATAATAACAGAC ACTATCAAGAGTTGGAGAACAATATATTGAGAACAAGAGTCTGAATGTGCATGTGTAATGGTCTTGTCTTACTGTAATGACCGATGGACCAAGTGTG GACAGGCTCATAAAGATCTTCAGAATAGAAAAGGAAAGATAATTAATCAGTCGAAATGAATGCCCTAATTATTAATGAGGAATGCTCCTGTTATCC TGATCTAGTGAATCACATGTGTGTCAGGGATAACTGGCATGGCTCGAATCGACCGTGGGTGTCCTTCAACCAGAACTGGAAATCAGATAGGATACATA TGCAAGTGGGATTTTCGGAGACAAATCCAGCCCTAATGATAAGACAGGCAGTTGTGGTCCAGTATCGTCTAATGGAGCAAATGGAGTAAAGGA                     </pre>			
Mark selected region: <input type="button" value="&lt;&gt;"/> <input type="button" value="[]"/> <input type="button" value="{}"/> <input type="button" value="Clear"/>			<input type="button" value="Save Sequence"/>
Excluded Regions:	< <input type="text"/> >		
Targets:	[ <input type="text"/> ]		
Included Region:	{ <input type="text"/> }		
Primer overlap positions:	- <input type="text"/>		
Pair OK Region List:	<input type="text"/>		
<input checked="" type="checkbox"/> Pick left primer	<input type="checkbox"/> Pick hybridization probe	<input checked="" type="checkbox"/> Pick right primer	
or use <a href="#">left primer</a> below.	(internal oligo) or use <a href="#">oligo</a> below.	or use <a href="#">right primer</a> below (5'→3' on opposite strand).	
<input type="text"/>	<input type="text"/>	<input type="text"/>	

รูปที่ 7.5 โปรแกรม Primer3Plus [https://primer3plus.com/cgi-bin/dev/primer3plus.cgi]

Main	General Settings	Advanced Settings	Internal Oligo	Penalty Weights	Advanced Sequence
<b>Product Size Ranges</b> 501-600 601-700 401-500 701-850 851-1000 1001-1500 1501-3000 3001-5000 401-					
Primer Size	Min: 18	Opt: 20	Max: 27		
Primer Tm	Min: 57.0	Opt: 60.0	Max: 63.0	Max Tm Difference: 100.0	
Primer GC%	Min: 20.0	Opt: 50.0	Max: 80.0		
Concentration of monovalent cations:	50.0	Annealing Oligo Concentration:	50.0		
Concentration of divalent cations:	1.5	Concentration of dNTPs:	0.6		
Mispriming/Repeat Library: NONE					

รูปที่ 7.6 การตั้งค่า General Settings ของโปรแกรม Primer3Plus

Main	General Settings	Advanced Settings	Internal Oligo	Penalty Weights	Advanced Sequence
Max Poly-X:	5	Table of thermodynamic parameters:	SantaLucia 1998		
Max #N's:	0	Salt correction formula:	SantaLucia 1998		
CG Clamp:	0	Use Thermodynamic Primer Alignment:	<input checked="" type="checkbox"/> Activates Settings Starting with TH:		
Max End GC:	5	Use Thermodynamic Template Alignment:	<input type="checkbox"/> Activates TH: Settings-VERY SLOW		
Number To Return:	5	Max End Stability:	9.0		
5 Prime Junction Overlap:	7	3 Prime Junction Overlap:	4		
Min Left Primer End Distance:	3	Min Right Primer End Distance:	3		
Max Self Complementarity:	8.00	Max Pair Complementarity:	8.00		
TH: Max Self Complementarity:	47.00	TH: Max Pair Complementarity:	47.00		
Max End Self Complementarity:	3.00	Max Pair End Complementarity:	3.00		
TH: Max End Self Compl.:	47.00	TH: Max Pair End Complementarity:	47.00		
TH: Max Hairpin:	47.00				
Max Template Mispriming:	12.00	Pair Max Template Mispriming:	24.00		
TH: Max Template Mispriming:	47.00	TH: Pair Max Template Mispriming:	47.00		
Max Library Mispriming:	12.00	Pair Max Library Mispriming:	24.00		
Primer Must Match 5 Prime:		Internal Must Match 5 Prime:			
Primer Must Match 3 Prime:		Internal Must Match 3 Prime:			
Left Primer Acronym:	F	Internal Oligo Acronym:	IN		
Right Primer Acronym:	R	Primer Name Spacer:	_		
Product Tm	Min: <input type="text"/>	Opt: <input type="text"/>	Max: <input type="text"/>		
Product Size	Min: <input type="text"/>	Opt: <input type="text"/>	Max: <input type="text"/>	<input type="checkbox"/> Debug Information	
<input checked="" type="checkbox"/> Pick Anyway <input checked="" type="checkbox"/> Liberal Base <input type="checkbox"/> Do not treat ambiguity codes in libraries as consensus <input type="checkbox"/> Use Lowercase Masking					
<b>Sequencing</b>					
Lead	50	Spacing	500		
Accuracy	20	Interval	250		

รูปที่ 7.7 การตั้งค่า Advanced Settings ของโปรแกรม Primer3Plus

การออกแบบไพรเมอร์ด้วยคอมพิวเตอร์

Main	General Settings	Advanced Settings	Internal Oligo	Penalty Weights	Advanced Sequence
<u>Excluded Region:</u> <input type="text"/>					
Hyb Oligo Size:	Min: <input type="text" value="18"/>	Opt: <input type="text" value="20"/>	Max: <input type="text" value="27"/>		
Hyb Oligo Tm:	Min: <input type="text" value="57.0"/>	Opt: <input type="text" value="60.0"/>	Max: <input type="text" value="63.0"/>		
Hyb Oligo GC%	Min: <input type="text" value="20.0"/>	Opt: <input type="text" value="50.0"/>	Max: <input type="text" value="80.0"/>		
Hyb Oligo Monovalent Cations Concentration:	<input type="text" value="50.0"/>		Hyb Oligo DNA Concentration:	<input type="text" value="50.0"/>	
Hyb Oligo Divalent Cations Concentration:	<input type="text" value="0.0"/>		Hyb Oligo dNTP Concentration:	<input type="text" value="0.0"/>	
Max #Ns:	<input type="text" value="0"/>		Hyb Oligo Max Poly-X:	<input type="text" value="5"/>	
Hyb Oligo Self Complementarity:	<input type="text" value="12.00"/>		Hyb Oligo Max End Self Complementarity:	<input type="text" value="12.00"/>	
TH: Hyb Oligo Self Complementarity:	<input type="text" value="47.00"/>		TH: Hyb Oligo Max End Self Complementarity:	<input type="text" value="47.00"/>	
TH: Hyb Oligo Max Hairpin:	<input type="text" value="47.00"/>		Max Library Mishyb:	<input type="text" value="12.00"/>	
Hyb Oligo Mishyb Library:	<input type="text" value="NONE"/>				
Hyb Oligo Min Sequence Quality:	<input type="text" value="0"/>				

รูปที่ 7.8 การตั้งค่า Internal Oligo ของโปรแกรม Primer3Plus

Main	General Settings	Advanced Settings	Internal Oligo	Penalty Weights	Advanced Sequence
<b>For Primers</b>					
Size	Lt: <input type="text" value="1.0"/>	Gt: <input type="text" value="1.0"/>			
Tm	Lt: <input type="text" value="1.0"/>	Gt: <input type="text" value="1.0"/>			
GC%	Lt: <input type="text" value="0.0"/>	Gt: <input type="text" value="0.0"/>			
Template Mispriming	<input type="text" value="0.0"/>				
TH: Template Mispriming	<input type="text" value="0.0"/>				
Library Mispriming	<input type="text" value="0.0"/>				
Self Complementarity	<input type="text" value="0.0"/>				
TH: Self Complementarity	<input type="text" value="0.0"/>				
End Self Complementarity	<input type="text" value="0.0"/>				
TH: End Self Complementarity	<input type="text" value="0.0"/>				
TH: Hairpin	<input type="text" value="0.0"/>				
#N's	<input type="text" value="0.0"/>				
Sequence Quality	<input type="text" value="0.0"/>				
End Sequence Quality	<input type="text" value="0.0"/>				
Position Penalty	<input type="text" value="0.0"/>				
End Stability	<input type="text" value="0.0"/>				
Inside Target Penalty:	<input type="text" value="-1.0"/>				
Outside Target Penalty:	<input type="text" value="0.0"/>				
Set Inside Target Penalty to allow primers inside a target.					
<b>For Internal Oligos</b>					
Size	Lt: <input type="text" value="1.0"/>	Gt: <input type="text" value="1.0"/>			
Tm	Lt: <input type="text" value="1.0"/>	Gt: <input type="text" value="1.0"/>			
GC%	Lt: <input type="text" value="0.0"/>	Gt: <input type="text" value="0.0"/>			
Library Mishyb	<input type="text" value="0.0"/>				
Self Complementarity	<input type="text" value="0.0"/>				
TH: Self Complementarity	<input type="text" value="0.0"/>				
End Self Complementarity	<input type="text" value="0.0"/>				
TH: End Self Complementarity	<input type="text" value="0.0"/>				
TH: Hairpin	<input type="text" value="0.0"/>				
Hyb Oligo #N's	<input type="text" value="0.0"/>				
Sequence Quality	<input type="text" value="0.0"/>				
End Sequence Quality	<input type="text" value="0.0"/>				
<b>For Primer Pairs</b>					
Product Size	Lt: <input type="text" value="0.0"/>	Gt: <input type="text" value="0.0"/>			
Tm Difference	<input type="text" value="0.0"/>				
Product Tm	Lt: <input type="text" value="0.0"/>	Gt: <input type="text" value="0.0"/>			
Template Mispriming	<input type="text" value="0.0"/>				
TH: Template Mispriming	<input type="text" value="0.0"/>				
Library Mispriming	<input type="text" value="0.0"/>				
Pair Complementarity	<input type="text" value="0.0"/>				
TH: Pair Complementarity	<input type="text" value="0.0"/>				
Pair End Complementarity	<input type="text" value="0.0"/>				
TH: Pair End Complementarity	<input type="text" value="0.0"/>				
Primer Penalty Weight	<input type="text" value="1.0"/>				
Hyb Oligo Penalty Weight	<input type="text" value="0.0"/>				

รูปที่ 7.9 การตั้งค่า Penalty Weights ของโปรแกรม Primer3Plus

Main	General Settings	Advanced Settings	Internal Oligo	Penalty Weights	Advanced Sequence
Force Left Primer Start:	<input type="text"/>	Force Right Primer Start:	<input type="text"/>		
Force Left Primer End:	<input type="text"/>	Force Right Primer End:	<input type="text"/>		
First Base Index:	<input type="text" value="1"/>	Sequence Quality:	<div style="border: 1px solid gray; height: 100px; width: 100%;"></div>		
Start Codon Position:	<input type="text"/>				
Min Sequence Quality:	<input type="text" value="0"/>				
Min End Sequence Quality:	<input type="text" value="0"/>				
Sequence Quality Range Min:	<input type="text" value="0"/>				
Sequence Quality Range Max:	<input type="text" value="100"/>				

รูปที่ 7.10 การตั้งค่า Advanced Sequence ของโปรแกรม Primer3Plus

โดยทั่วไปมักนิยมตั้งค่าเฉพาะในหมวด General Settings เพื่อกำหนดปัจจัยที่สำคัญต่างๆ เช่น Primer Tm, Max Tm Differences และ Product Size Ranges เมื่อตั้งค่าต่างๆ เสร็จแล้ว คลิกที่คำสั่ง Pick Primers โปรแกรมก็จะประมวลผล และแสดงผลการออกแบบ primers (รูปที่ 7.11)

## Primer3Plus

pick primers from a DNA sequence

[More...](#)
[Source Code](#)

[Help](#)
[About](#)

---

[< Back](#)

Pair 1:

Left Primer 1:

Start: 17    Length: 20 bp    Tm: 60.0 C    GC: 60.0 %    Any: 0.0    End: 0.0    TB: 5.0    HP: 0.0    3' Stab: 4.1    Penalty: 0.035

Right Primer 1:

Start: 480    Length: 20 bp    Tm: 60.0 C    GC: 55.0 %    Any: 0.0    End: 0.0    TB: 6.0    HP: 40.2    3' Stab: 2.5    Penalty: 0.038

Pair:    Product Size: 464 bp    Any: 0.0    End: 0.0    TB: 11.0    Penalty: 0.073

Send to Primer3Manager
Reset Form

1	ATGGAACCAT	TAAAGACAGG	AGCCCATATC	GAACCCTAAT	GAGCTGTCCT
51	ATTGGTGAAG	TCCCCTCTCC	ATACAACCTCA	AGATTGAGT	CAGTCGCTTG
101	GTCAGCAAGT	GCTTGTCATG	ATGGCATCAA	TTGGTAACA	ATTGGAATTT
151	CTGGCCAGA	CAATGGGGCA	GTGGCTGTGT	TAAAGTACAA	CGGCATAATA
201	ACAGACACTA	TCAAGAGTTG	GAGAAACAAT	ATATTGAGAA	CACAAGAGTC
251	TGAATGTGCA	TGTGTAATG	GTTCTTGCTT	TACTGTAATG	ACCGATGGAC
301	CAAGTGATGG	ACAGGCCTCA	TACAAGATCT	TCAGAATAGA	AAAGGGAAAG
351	ATAATTAAAT	CAGTCGAAAT	GAATGCCCTT	AATTATTACT	ATGAGGAATG
401	GTCCTGTTAT	CCTGATTCTA	GTGAAATCAC	ATGTGTGTGC	AGGGATAACT
451	CGCATGGCTC	GAATCGACCG	TGGGTGTCTT	TCAACCAGAA	TCTGGAATAT
501	CAGATAGGAT	ACATATGCAG	TGGGATTTTC	GGAGACAATC	CACGCCCTAA
551	TGATAAGACA	GGCAGTTGTG	GTCCAGTATC	GTCTAATGGA	GCAATGGAG
601	TAAAGGA				

รูปที่ 7.11 ตัวอย่างผลการออกแบบ primers ด้วยโปรแกรม Primer3Plus



➤ โปรแกรม Primer-BLAST [<https://www.ncbi.nlm.nih.gov/tools/primer-blast>]

เป็นโปรแกรม free online web-based ที่พัฒนาโดย NCBI โดยใช้กระบวนการออกแบบ primer ตามโปรแกรม Primer3 แล้วเพิ่มขั้นตอนการนำลำดับนิวคลีโอไทด์ของ primer ไปค้นหาเปรียบเทียบกับลำดับนิวคลีโอไทด์ในฐานข้อมูล (BLAST) เพื่อตรวจสอบความจำเพาะของ primer ว่ามีความคล้ายคลึงกับยีนอื่นๆ ที่ไม่ใช่เป็นยีนเป้าหมายหรือไม่ จึงทำให้โปรแกรมนี้สามารถออกแบบ primers ที่มีความจำเพาะสูง และเป็นที่ยอมรับใช้อย่างมาก โปรแกรม Primer-BLAST ประกอบด้วย 4 ส่วนสำหรับการตั้งค่าต่างๆ (รูปที่ 7.12) ดังนี้

- PCR Template เป็นส่วนที่ให้ผู้ผู้ใช้ข้อมูลลำดับนิวคลีโอไทด์ของยีนเป้าหมาย โดยอาจใส่เป็น accession no. หรือ ลำดับนิวคลีโอไทด์ที่เป็น FASTA format หรือจะ upload FASTA file ก็ได้ จากนั้นตั้งค่า Range สำหรับช่วงเริ่มต้น (From) และสิ้นสุด (To) ที่จะให้โปรแกรมวิเคราะห์และออกแบบ forward และ reverse primers หากไม่ได้กำหนดช่วง (Range) โปรแกรมก็จะใช้ลำดับนิวคลีโอไทด์ทั้งสายในการออกแบบ primers
- Primer Parameters เป็นการกำหนดค่าต่างๆ ของ primers ได้แก่
  - Use my own forward primer ใช้ forward primer ที่เรามีอยู่แล้ว
  - Use my own reverse primer ใช้ reverse primer ที่เรามีอยู่แล้ว
  - PCR Product size กำหนดช่วงขนาดสั้นที่สุด (min) และยาวที่สุด (max) ของ PCR product
  - # of primers to return จำนวนคู่ของ primers ที่จะให้โปรแกรมแสดง
  - Primer melting temperature กำหนดค่าต่ำสุด (min), เหมาะสม (Opt) และสูงสุด (max) ของ Tm
  - Max Tm difference กำหนดค่าความต่างสูงสุดของ Tm ระหว่างคู่ primers
- Exon/intron selection กำหนดให้โปรแกรมออกแบบ primer ที่จับบริเวณรอยต่อระหว่าง exons เพื่อใช้เป็น primers สำหรับเพิ่มจำนวน cDNA template แต่ไม่สามารถเพิ่มจำนวน genomic DNA ในเทคนิค RT-PCR โดยสามารถตั้งค่าต่างๆ ได้ดังนี้

- Exon junction span สามารถเลือกได้ 3 แบบ ได้แก่ No preference, Must span และ May not span
  - Exon junction match กำหนดจำนวนเบสขั้นต่ำของ primer บริเวณด้าน 5' และ 3' ในการจับกับ exon
  - Intron inclusion กำหนดว่า forward และ reverse primers จะต้องครอบคลุมส่วนที่เป็น intron ใน genomic DNA
  - Intron length range กำหนดช่วงความยาวของ intron ที่สั้นที่สุด (min) และยาวที่สุด (max)
- Primer Pair Specificity Checking Parameters
  - Specificity check ตรวจสอบความจำเพาะของ primer
  - Search mode เลือกได้ 3 แบบ ได้แก่ Automatic, User guided และ No user guidance
  - Database เลือกฐานข้อมูลสำหรับการ BLAST ได้แก่ Refseq mRNA, Refseq representative genomes, Genomes for selected organisms, nr, Refseq RNA และ Custom เป็นต้น
  - Organism เลือกชนิดของสิ่งมีชีวิต
  - Primer specificity stringency กำหนดค่าในการคำนวณความจำเพาะของ primers เช่น หากนำลำดับนิวคลีโอไทด์ของ primer ไปเปรียบเทียบกับยีนอื่นๆ (unintended targets) จะต้อง mismatch อย่างน้อย 2 bp ในช่วง 5 เบสสุดท้ายทางด้าน 3' ของ primer เป็นต้น
  - Max target size ค่าความยาวสูงสุดของ non-specific PCR product
  - Allow splice variants หากยีนเป้าหมายมีการ splicing ได้หลายแบบ จะกำหนดให้ primers ดังกล่าวสามารถเพิ่มจำนวนยีนที่เป็น splice variants ได้

**Primer-BLAST** *A tool for finding specific primers*

NCBI/ **Primer-BLAST: Finding primers specific to your PCR template (using Primer3 and BLAST).**

[Reset page](#) [Save search parameters](#) [Retrieve recent results](#) [Publication](#) [Tips for finding specific primers](#)

**PCR Template**

Enter accession, gi, or FASTA sequence (A refseq record is preferred) [Clear](#)

Or, upload FASTA file [Browse...](#)

**Range**

Forward primer From To [Clear](#)

Reverse primer From To

**Primer Parameters**

Use my own forward primer (5'->3' on plus strand)  [Clear](#)

Use my own reverse primer (5'->3' on minus strand)  [Clear](#)

PCR product size Min Max  
70 1000

# of primers to return  10

Primer melting temperatures (T<sub>m</sub>) Min Opt Max Max T<sub>m</sub> difference  
57.0 60.0 63.0 3 [Clear](#)

**Exon/intron selection**

A refseq mRNA sequence as PCR template input is required for options in the section [Clear](#)

Exon junction span  No preference [Clear](#)

Exon junction match Exon at 5' side Exon at 3' side  
 7  4 [Clear](#)

Minimal number of bases that must anneal to exons at the 5' or 3' side of the junction [Clear](#)

Intron inclusion  Primer pair must be separated by at least one intron on the corresponding genomic DNA [Clear](#)

Intron length range Min Max  
 1000  1000000 [Clear](#)

**Primer Pair Specificity Checking Parameters**

Specificity check  Enable search for primer pairs specific to the intended PCR template [Clear](#)

Search mode  Automatic [Clear](#)

Database  Refseq mRNA [Clear](#)

Exclusion  Exclude predicted Refseq transcripts (accession with XM, XR prefix)  Exclude uncultured/environmental sample sequences [Clear](#)

Organism  Homo sapiens  
Enter an organism name (or organism group name such as enterobacteriaceae, rodents), taxonomy id or select from the suggestion list as you type. [Clear](#)  
[Add more organisms](#)

Entrez query (optional)  [Clear](#)

Primer specificity stringency Primer must have at least  2 total mismatches to unintended targets, including at least  2 mismatches within the last  5 bps at the 3' end. [Clear](#)  
Ignore targets that have  6 or more mismatches to the primer. [Clear](#)

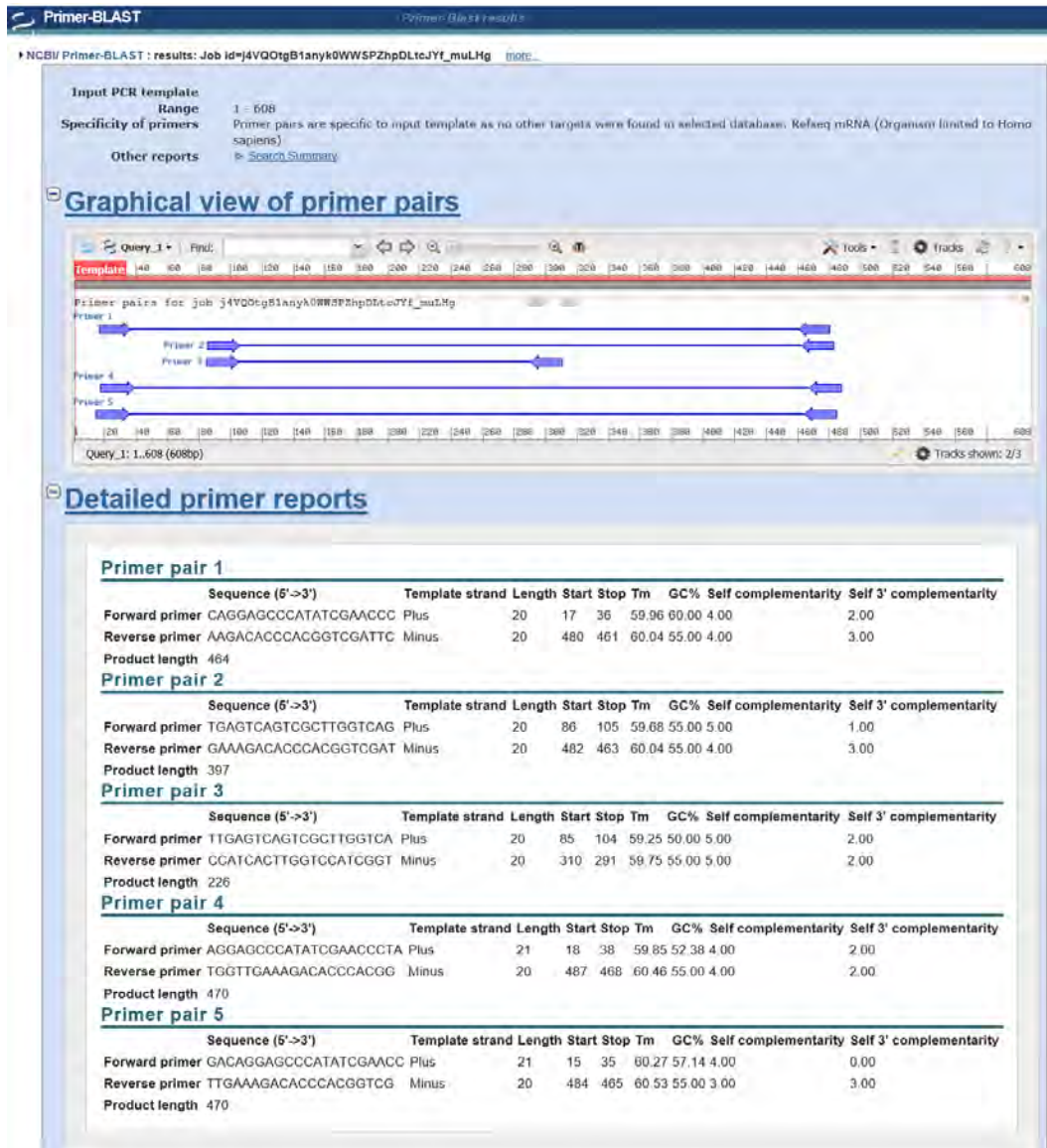
Max target size  4000 [Clear](#)

Allow splice variants  Allow primer to amplify mRNA splice variants (requires refseq mRNA sequence as PCR template input) [Clear](#)

[Get Primers](#)  Show results in a new window  Use new graphic view [Clear](#)

รูปที่ 7.12 โปรแกรม Primer-BLAST [https://www.ncbi.nlm.nih.gov/tools/primer-blast]

เมื่อตั้งค่าต่างๆ เสร็จแล้ว คลิกที่คำสั่ง Get Primers โปรแกรมก็จะประมวลผล และแสดงผลการออกแบบ primers เป็นภาพกราฟิก และ รายละเอียดของ primers ได้แก่ sequence, strand, length, start, stop, Tm, GC%, self-complementarity และ product length ของ primers แต่ละคู่ (รูปที่ 7.13)



รูปที่ 7.13 ตัวอย่างผลการออกแบบ primers ด้วยโปรแกรม Primer-BLAST

➤ โปรแกรม PrimerQuest Tool [https://sg.idtdna.com/Primerquest/Home/Index]

เป็นโปรแกรม free online web-based สำหรับการออกแบบ primers และ probes ซึ่งพัฒนาโดยบริษัท Integrated DNA Technologies โดยแบ่งเป็น 2 ส่วน (รูปที่ 7.14) ดังนี้

- Sequence Entry เป็นส่วนที่ให้ผู้ใส่ข้อมูลลำดับนิวคลีโอไทด์ของยีนเป้าหมาย โดยสามารถใส่ได้หลายรูปแบบ เช่น paste FASTA sequence, download sequence from GenBank, accession no. และ upload excel file
- Choose Your Design เป็นการเลือก assay ที่เราต้องการให้โปรแกรมออกแบบ ได้แก่
  - Custom Design Parameters ปรับเปลี่ยนค่าในการออกแบบ (รูปที่ 7.15)
  - PCR ออกแบบ primers สำหรับ conventional PCR (รูปที่ 7.16)
  - qPCR (intercalating dyes) ออกแบบ primers ที่ให้ PCR product ของสั้นๆ เหมาะสำหรับเทคนิค real-time PCR ที่ใช้ SYBR green I (รูปที่ 7.17)
  - qPCR (probe) ออกแบบ primers และ TaqMan probe ที่เหมาะสมสำหรับเทคนิค TaqMan real-time PCR (รูปที่ 7.18)



รูปที่ 7.14 PrimerQuest Tool [https://sg.idtdna.com/Primerquest/Home/Index]



## Custom Design Parameters

### Task Settings

#### Results

Results to Return  (1-50)

Vary Forward Primer Location  Off Low Med High

Vary Reverse Primer Location  Off Low Med High

#### Reaction Conditions

Monovalent Salt (Na<sup>+</sup>) (mM)

Primer DNA Conc (nM)

Divalent Salt (Mg<sup>2+</sup>) (mM)

Probe DNA Conc (nM)

dNTP Conc (mM)

### Primer Criteria

#### Primer Details

	Min	Opt	Max			
Primer Tm (°C)	<input type="text" value="59"/>	<input type="text" value="62"/>	<input type="text" value="65"/>	45	<input type="range" value="62"/>	75
Primer GC (%)	<input type="text" value="35"/>	<input type="text" value="50"/>	<input type="text" value="65"/>	20	<input type="range" value="50"/>	80
Primer Size (nt)	<input type="text" value="17"/>	<input type="text" value="22"/>	<input type="text" value="30"/>	12	<input type="range" value="22"/>	45

#### Misc Settings

3' GC Clamp (nt)

### Amplicon Criteria

#### Amplicon Details

	Min	Opt	Max			
Amplicon Size (bp)	<input type="text" value="75"/>	<input type="text" value="100"/>	<input type="text" value="150"/>		<input type="range" value="100"/>	

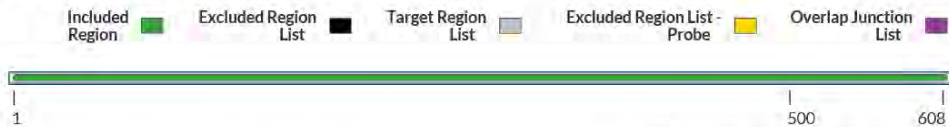
### Probe Criteria

#### Probe Details

	Min	Opt	Max			
Probe Tm (°C)	<input type="text" value="64"/>	<input type="text" value="68"/>	<input type="text" value="72"/>	45	<input type="range" value="68"/>	85
Probe GC (%)	<input type="text" value="40"/>	<input type="text" value="50"/>	<input type="text" value="60"/>	20	<input type="range" value="50"/>	80
Probe Size (nt)	<input type="text" value="20"/>	<input type="text" value="24"/>	<input type="text" value="30"/>	12	<input type="range" value="24"/>	45

### Custom Target Region

#### Regions Considered



Included Region

Excluded Region List

Excluded Region List - Probe

#### Primer Focus Region

Target Region List

#### Partial Design Input

Forward Primer Input

Probe Input

Reverse Primer Input

#### Design Across Junctions

Overlap Junction List

Min Overlap on 3' end of Primer

Min Overlap on 5' end of Primer

#### Primer Force Location Settings

Force Forward Start Position (5' End)

Force Forward Stop Position (3' End)

Force Reverse Start Position (5' End)

Force Reverse Stop Position (3' End)

GET ASSAYS

รูปที่ 7.15 การตั้งค่า Custom Design Parameters ของโปรแกรม PrimerQuest Tool

การออกแบบไพรเมอร์ด้วยคอมพิวเตอร์

Parameter Set: General PCR (Primers only)  
 Sequence Name: Target gene  
 Amplicon Length: 505

		Start	Stop	Length	Tm	GC%
Forward	<b>CCTATTGGTGAAGTTCCTCTC (Sense)</b>	48	70	22	62	50
Reverse	<b>CATTAGGGCGTGGATTGTCT (AntiSense)</b>	533	553	20	62	50

Base	Sequence
1	ATGGAACCATTAAGACAGGAGCCCATATCGAACCCCTAATGAGCTGT <b>CCTATTGGTGAAGTTCCTCTC</b> CATACAACCTCAAGATTTGAGTCAGTCGCTTG
101	GTCAGCAAGTGCTTGTGCATGATGGCATCAATTGGCTAACAAATTTGGAATTTCTGGCCAGACAATGGGGCAGTGGCTGTGTTAAAGTACAACGGCATAATA
201	ACAGACACTATCAAGAGTTGGAGAAACAATATATTGAGAACACAAGAGTCTGAATGTGCATGTGTAATGGTTCTTGCTTTACTGTAATGACCGATGGAC
301	CAAGTGATGGACAGGCCTCATACAAGATCTTCAGAATAGAAAAGGAAAGATAATTAATCAGTCGAAATGAATGCCCTAATTATTACTATGAGGAATG
401	CTCCTGTTATCCTGATTCTAGTGAATCACAATGTGTGTCAGGGATAACTGGCATGGCTCGAATCGACCGTGGGTGTCTTTCAACAGAACTCTGGAATAT
501	CAGATAGGATACATATGCAGTGGGATTTTCGG <b>AGACAAATCCACGCCCTAATG</b> ATAAGACAGGCAGTTGTGGTCCAGTATCGTCTAATGGAGCAAATGGAG
601	TAAAAGGA

รูปที่ 7.16 ตัวอย่างผลการออกแบบ primers สำหรับ conventional PCR

Parameter Set: qPCR Intercalating Dyes (Primers only)  
 Sequence Name: Target gene  
 Amplicon Length: 107

		Start	Stop	Length	Tm	GC%
Forward	<b>TAACTGGCATGGCTCGAATC (Sense)</b>	446	466	20	62	50
Reverse	<b>CATTAGGGCGTGGATTGTCT (AntiSense)</b>	533	553	20	62	50

Base	Sequence
1	ATGGAACCATTAAGACAGGAGCCCATATCGAACCCCTAATGAGCTGTCTTATTGGTGAAGTTCCTCTCCATACAACCTCAAGATTTGAGTCAGTCGCTTG
101	GTCAGCAAGTGCTTGTGCATGATGGCATCAATTGGCTAACAAATTTGGAATTTCTGGCCAGACAATGGGGCAGTGGCTGTGTTAAAGTACAACGGCATAATA
201	ACAGACACTATCAAGAGTTGGAGAAACAATATATTGAGAACACAAGAGTCTGAATGTGCATGTGTAATGGTTCTTGCTTTACTGTAATGACCGATGGAC
301	CAAGTGATGGACAGGCCTCATACAAGATCTTCAGAATAGAAAAGGAAAGATAATTAATCAGTCGAAATGAATGCCCTAATTATTACTATGAGGAATG
401	CTCCTGTTATCCTGATTCTAGTGAATCACAATGTGTGTCAGGGATA <b>TAACTGGCATGGCTCGAATC</b> GACCGTGGGTGTCTTTCAACAGAACTCTGGAATAT
501	CAGATAGGATACATATGCAGTGGGATTTTCGG <b>AGACAAATCCACGCCCTAATG</b> ATAAGACAGGCAGTTGTGGTCCAGTATCGTCTAATGGAGCAAATGGAG
601	TAAAAGGA

รูปที่ 7.17 ตัวอย่างผลการออกแบบ primers สำหรับ qPCR (intercalating dyes)

Parameter Set: RT-qPCR (Primers with Probe)  
 Sequence Name: Target gene  
 Amplicon Length: 107

		Start	Stop	Length	Tm	GC%
Forward	<b>TAACTGGCATGGCTCGAATC (Sense)</b>	446	466	20	62	50
Probe	<b>ACCGTGGGTGTCTTTCAACAGAA (Sense)</b>	467	491	24	68	50
Reverse	<b>CATTAGGGCGTGGATTGTCT (AntiSense)</b>	533	553	20	62	50

Base	Sequence
1	ATGGAACCATTAAGACAGGAGCCCATATCGAACCCCTAATGAGCTGTCTTATTGGTGAAGTTCCTCTCCATACAACCTCAAGATTTGAGTCAGTCGCTTG
101	GTCAGCAAGTGCTTGTGCATGATGGCATCAATTGGCTAACAAATTTGGAATTTCTGGCCAGACAATGGGGCAGTGGCTGTGTTAAAGTACAACGGCATAATA
201	ACAGACACTATCAAGAGTTGGAGAAACAATATATTGAGAACACAAGAGTCTGAATGTGCATGTGTAATGGTTCTTGCTTTACTGTAATGACCGATGGAC
301	CAAGTGATGGACAGGCCTCATACAAGATCTTCAGAATAGAAAAGGAAAGATAATTAATCAGTCGAAATGAATGCCCTAATTATTACTATGAGGAATG
401	CTCCTGTTATCCTGATTCTAGTGAATCACAATGTGTGTCAGGGATA <b>TAACTGGCATGGCTCGAATC</b> <b>ACCGTGGGTGTCTTTCAACAGAA</b> TCTGGAATAT
501	CAGATAGGATACATATGCAGTGGGATTTTCGG <b>AGACAAATCCACGCCCTAATG</b> ATAAGACAGGCAGTTGTGGTCCAGTATCGTCTAATGGAGCAAATGGAG
601	TAAAAGGA

รูปที่ 7.18 ตัวอย่างผลการออกแบบ primers และ TaqMan probe สำหรับ qPCR (probe)



จากที่นำเสนอไปแล้วข้างต้น จะเห็นได้ว่าโปรแกรมคอมพิวเตอร์สำหรับการออกแบบ primers ไม่ว่าจะเป็น Primer3web, Primer3Plus, Primer-BLAST และ PrimerQuest Tool จะเริ่มต้นการออกแบบ primers จากลำดับนิวคลีโอไทด์เพียงสายเดียว แล้วคำนวณลักษณะ thermodynamics ของ primers ได้แก่ GC%, Tm, secondary structure และ PCR product size เพื่อให้ได้ primers ที่เหมาะสม โดยมีได้ค่านิ่งว่าลำดับนิวคลีโอไทด์ส่วนที่เลือกมาเป็น primers นั้นอยู่ที่บริเวณใดของยีน บริเวณดังกล่าวมีการกลายพันธุ์เกิดขึ้นหรือไม่ ทั้งนี้การออกแบบ primers สำหรับการตรวจสอบหาเชื้อไวรัส หรือการตรวจวินิจฉัยการติดเชื้อไวรัสควรมีการคำนึงถึงตำแหน่งของ primers ด้วย เนื่องจากเชื้อไวรัสมีอัตราการกลายพันธุ์สูงมาก ดังนั้นจึงมีความเป็นไปได้ว่า primers ที่ใช้จะเกิด mismatch กับสารพันธุกรรมของเชื้อไวรัสเนื่องจากเกิดการกลายพันธุ์ ซึ่งจะส่งผลให้ primers ดังกล่าวไม่สามารถเพิ่มจำนวนสารพันธุกรรมของเชื้อไวรัสได้ และจะไม่สามารถตรวจสอบหาการติดเชื้อไวรัสได้ หรืออาจเกิดผลเป็น false negative ในการตรวจวินิจฉัยโรคติดเชื้อไวรัส

ดังนั้นในการออกแบบ primers และ probes สำหรับการตรวจสอบสารพันธุกรรมของเชื้อไวรัส จึงนิยมออกแบบตามวัตถุประสงค์ 2 รูปแบบ ดังนี้

- Universal primers ออกแบบจากส่วนที่เป็นบริเวณอนุรักษ์ (conserved regions) ที่ได้จาก nucleotide sequences alignment สำหรับตรวจ screening (+/-) เบื้องต้น
  - ตรวจหา common gene ที่พบในไวรัสชนิดต่างๆ ที่จัดอยู่ในตระกูลเดียวกัน เช่น Dengue virus, Yellow fever virus, Japanese encephalitis virus, West Nile virus และ Zika virus เป็นต้น
  - ตรวจหา common gene ที่พบในสายพันธุ์ต่างๆ (subtypes, genotypes, serotypes) ของเชื้อไวรัสชนิดเดียวกัน
- Genotype / subtype specific primers & probe ออกแบบจากส่วนที่เป็นบริเวณแปรเปลี่ยน (variable region) ที่ได้จาก nucleotide sequences alignment ใช้สำหรับ allele specific PCR หรือ allelic discrimination plot ใน real-time PCR เพื่อการจำแนกสายพันธุ์ ตรวจการกลายพันธุ์ และตรวจการดื้อยาของไวรัส เป็นต้น

## การออกแบบ primers และ probes จาก nucleotide sequences alignment

1. เริ่มต้นจากการดาวน์โหลดลำดับนิวคลีโอไทด์ของเชื้อไวรัสที่สนใจ สายพันธุ์ต่างๆ จากฐานข้อมูลทางชีวภาพ (biological database) หรือฐานข้อมูลทางด้านไวรัสวิทยา (virological database) แล้วจัดเตรียมข้อมูลลำดับนิวคลีโอไทด์ทั้งหมดไว้ใน text file (FASTA format) ของโปรแกรม Notepad ดังที่กล่าวไปแล้วในบทก่อนหน้า

2. เปิดใช้งานโปรแกรม BioEdit แล้วเปิด text file ที่เตรียมไว้ จากนั้นคลิกที่คำสั่ง Accessory Application > ClustalW Multiple alignment จะมีหน้าต่างปรากฏขึ้นเพื่อให้ตั้งค่าที่จะใช้ในการจัดเรียงลำดับนิวคลีโอไทด์ เมื่อตั้งค่าเสร็จแล้ว (นิยมใช้ค่า default) ให้คลิกที่คำว่า Run ClustalW จากนั้นกด OK จะปรากฏหน้าต่างการทำงานของโปรแกรม แล้วก็จะแสดงผลขึ้นมาอัตโนมัติ ดังที่กล่าวไปแล้วในบทก่อนหน้า

3. หากต้องการจำแนกส่วนที่เหมือนกันและแตกต่างกันอย่างชัดเจน ให้คลิกที่ปุ่ม “View conservation by plotting identities to a standard as a dot” โปรแกรมจะใช้ลำดับนิวคลีโอไทด์สายแรก เป็นตัวตั้งต้นในการเปรียบเทียบกับลำดับนิวคลีโอไทด์สายอื่นๆ โดยแสดงผล ดังนี้

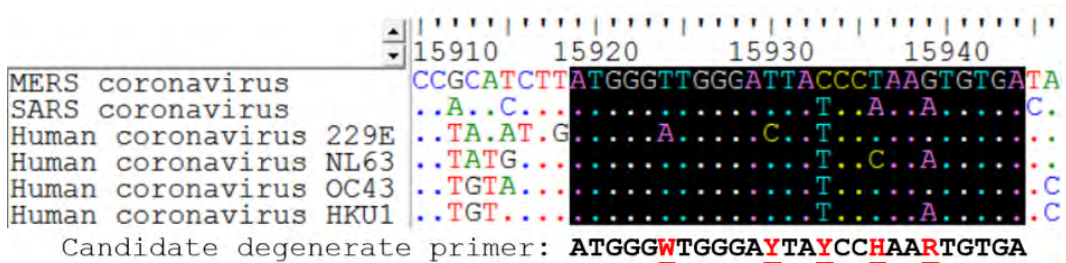
- บริเวณอนุรักษ์ (conserved region) จะแสดงเป็นเครื่องหมาย (.)
- บริเวณแปรเปลี่ยน (variable region) จะแสดงเป็นเบสที่แตกต่าง (A/T/C/G)

4. สามารถสั่งให้โปรแกรมค้นหาบริเวณอนุรักษ์ (conserved regions) แบบอัตโนมัติ โดยคลิกที่คำสั่ง Alignment > Find Conserved Regions จะมีหน้าต่างปรากฏขึ้นเพื่อให้ตั้ง จากนั้นคลิกที่คำว่า Start โปรแกรมก็จะแสดงผลการค้นหาบริเวณอนุรักษ์

5. บริเวณอนุรักษ์ที่มีความยาวประมาณ 20-25 bp สามารถนำมาใช้ในการออกแบบ universal primers ได้ อย่างไรก็ตามเนื่องจากเชื้อไวรัสบางชนิดมีอัตราการกลายพันธุ์สูงมาก ซึ่งจะทำให้หาบริเวณอนุรักษ์ (conserved regions) ที่มีความยาว 20-25 bp เพื่อใช้เป็น universal primer ไม่ได้ จึงจำเป็นต้องมีการปรับเปลี่ยนลำดับเบสให้เป็น degenerate primer (ตารางที่ 7.2) เพื่อให้รองรับไวรัสสายพันธุ์ต่างๆ ได้ (รูปที่ 7.19)

ตารางที่ 7.2 Degeneracy codes สำหรับ degenerate primers

Degeneracy code	Bases	Complement code
M	A / C	K
R	A / G	Y
W	A / T	W
S	G / C	S
Y	C / T	R
K	G / T	M
V	A / G / C	B
H	A / C / T	D
D	A / G / T	H
B	C / G / T	V
N	A / G / C / T	N

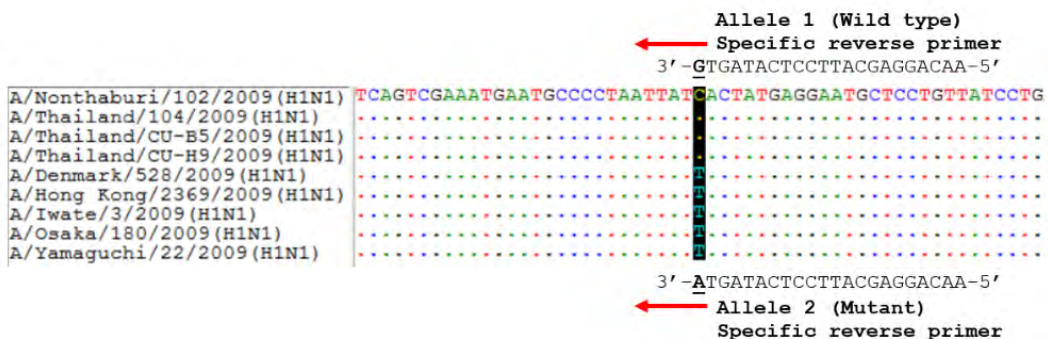


รูปที่ 7.19 ตัวอย่าง degenerate primers สำหรับ coronavirus สายพันธุ์ต่างๆ ที่พบในมนุษย์

6. ควรคำนึงถึงความจำเพาะของ degenerate primers โดยพิจารณาปัจจัย ดังนี้

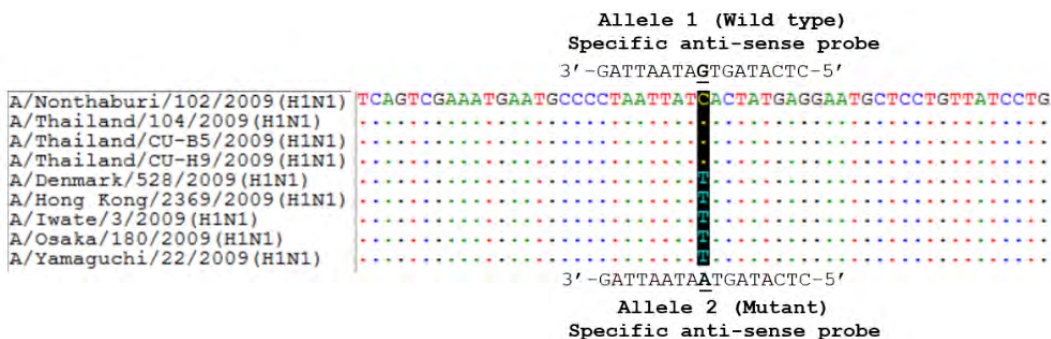
- ไม่ควรใช้ degeneracy bases บริเวณ 5 bp ที่ปลายด้าน 3' ของ primer
- ไม่ควรใช้ degeneracy bases ในตำแหน่งที่อยู่ติดๆ กัน
- ไม่ควรมี degeneracy bases เกิน 5 ตำแหน่งใน primer แต่ละสาย

7. Genotype / subtype specific primers สำหรับ allele specific PCR ออกแบบจากส่วนที่เป็นบริเวณแปรเปลี่ยน (variable region) ที่ได้จาก nucleotide sequences alignment โดยให้ตำแหน่งที่จะใช้ในการเลือกจับระหว่าง wild type และ mutant alleles อยู่ที่ปลายด้าน 3' ของ primer (รูปที่ 7.20)



รูปที่ 7.20 ตัวอย่าง allele specific primers สำหรับตรวจการดื้อยาของเชื้อไวรัสไข้หวัดใหญ่

8. Genotype / subtype specific probes สำหรับ allelic discrimination plot ใน multiplex real-time PCR ออกแบบจากส่วนที่เป็นบริเวณแปรเปลี่ยน (variable region) ที่ได้จาก nucleotide sequences alignment โดยให้ตำแหน่งที่จะใช้ในการเลือกจับระหว่าง allele 1 (wild type) และ allele 2 (mutant) ควรจะอยู่ที่ตำแหน่งกึ่งกลางของ probe เช่น หาก TaqMan MGB probe มีความยาว 17 bp ตำแหน่งที่จะใช้แยกระหว่าง allele 1 และ 2 ควรจะอยู่ในตำแหน่งที่ 9 ของ probe เป็นต้น (รูปที่ 7.21)



รูปที่ 7.21 ตัวอย่าง allele specific probes สำหรับตรวจการดื้อยาของเชื้อไวรัสไข้หวัดใหญ่

9. เนื่องจาก nucleotide sequences alignment จะแสดงลำดับนิวคลีโอไทด์เพียงสายเดียวเท่านั้น แต่การเพิ่มจำนวนด้วยเทคนิค PCR จะใช้ forward และ reverse primers ซึ่งมีทิศสวนทางกัน และเข้าจับกับ DNA template คนละสาย ดังนั้นจึงควรคำนึงถึง strand ของ candidate primers และ probes ดังนี้

- Forward primer และ sense strand probe สามารถ copy & paste ลำดับนิวคลีโอไทด์จาก nucleotide sequences alignment มาใช้เป็น candidate primer/probe
- Reverse primer และ anti-sense strand probe เมื่อ copy & paste ลำดับนิวคลีโอไทด์จาก nucleotide sequences alignment มาแล้ว จะต้องทำการกลับทิศทางและเปลี่ยนเป็นเบสคู่สม (reverse and complement) ก่อนจึงจะใช้เป็น candidate primer/probe ได้

10. นำลำดับนิวคลีโอไทด์ของ candidate primers / probes ไปวิเคราะห์ด้วยการ BLAST เพื่อค้นหาว่า candidate primers / probes มีความคล้ายคลึงกับยีนอื่นๆ ที่ไม่ใช่ยีนเป้าหมายหรือไม่

11. วิเคราะห์คุณสมบัติพื้นฐานของ candidate primers / probes เช่น length, %GC และ Tm จากนั้นปรับให้มีความเหมาะสมตามหลักการออกแบบ primers / probes ที่กล่าวไปแล้วข้างต้น เช่น Tm ของ forward และ reverse primers ควรมีค่าเท่ากัน หรือต่างกันไม่เกิน 2°C และ Tm ของ probe ควรสูงกว่า primers ประมาณ 8-10°C เป็นต้น

12. วิเคราะห์โอกาสในการเกิด secondary structures (hairpin, self-complementary และ primer dimer) ของ candidate primers / probes

ซึ่งการวิเคราะห์คุณสมบัติพื้นฐานและการทำนายการเกิด secondary structures ของ candidate primers / probes สามารถใช้โปรแกรมต่างๆ ช่วยในการวิเคราะห์ได้ ดังนี้

➤ โปรแกรม Oligo Calc [<http://biotools.nubic.northwestern.edu/OligoCalc.html>]

Oligo Calc (Oligonucleotide Properties Calculator) เป็นโปรแกรมสำหรับการคำนวณคุณสมบัติต่างๆ ของ oligonucleotides ที่สามารถใช้ในการออกแบบ primers และ probes ได้ โดยผู้ใช้สามารถใส่ข้อมูลลำดับนิวคลีโอไทด์ของ candidate primers / probes ที่เลือกเอาไว้ เข้าไปในช่อง Nucleotide base code จากนั้นกำหนด 5' / 3' modification (การติดฉลากด้วยหมู่เคมี หรือสารเรืองแสง) ความเข้มข้นของ primer และ ความเข้มข้นของ salt แล้วคลิกที่ Calculate เพื่อให้โปรแกรมคำนวณคุณสมบัติต่างๆ (รูปที่ 7.22) ดังนี้

**Oligo Calc: Oligonucleotide Properties Calculator**

Enter Oligonucleotide Sequence Below  
*OD calculations are for single-stranded DNA or RNA*

**Nucleotide base codes**

GGT TGG GAY TAY CCH AAR TGT GA

Reverse Complement Strand(5' to 3') is:  
TCA CAY TTD GGR TAR TCC CAA CC

5' modification (if any)      3' modification (if any)      Select molecule

50 nM Primer      1 Measured Absorbance at 260 nanometers

50 mM Salt (Na<sup>+</sup>)

**Calculate**    **Swap Strands**    **BLAST**    **mfold**

**Physical Constants**      **Melting Temperature (T<sub>m</sub>) Calculations**

Length: 23    Molecular Weight: 7088.7 to 7158.4    GC content: 39 to 57%    1 51.7 to 58.8 °C (Basic)

1 ml of a sol'n with an Absorbance of 1 at 260 nm    2 59.2 to 66.6 °C (Salt Adjusted)

is 3.797 to 4.01 microMolar and contains 26.9 to 28.7 micrograms.    3 48.35 to 51.25 °C (Nearest Neighbor)

**Thermodynamic Constants Conditions: 1 M NaCl at 25°C at pH 7.**

RlnK 33.404 cal/(°K\*mol)      deltaH 160.6 to 178.8 Kcal/mol

deltaG 23.8 to 27.2 Kcal/mol      deltaS 425.1 to 473.9 cal/(°K\*mol)

**Deprecated Hairpin/self dimerization calculations**

5 (Minimum base pairs required for single primer self-dimerization)

4 (Minimum base pairs required for a hairpin)

**Check Self-Complementarity**

รูปที่ 7.22 Oligo Calc [<http://biotools.nubic.northwestern.edu/OligoCalc.html>]

- Reverse Complement Strand โปรแกรมจะแสดงผลการกลับทิศทาง และเปลี่ยนเป็นเบสคู่สม สำหรับ reverse primer และ anti-sense strand probe
- Physical constants ได้แก่ length, molecular weight, %GC content และ Tm
- Thermodynamic constants ได้แก่  $\Delta G$ ,  $\Delta H$ ,  $\Delta S$  และ RlnK
- Deprecated Hairpin/self-dimerization calculations สามารถตั้งค่าเพื่อใช้ในการคำนวณหาโอกาสในการเกิด secondary structure หากคลิกที่ Swap Strands โปรแกรมจะทำการสลับลำดับนิวคลีโอไทด์ระหว่างช่อง Nucleotide base code และ Reverse Complement Strand
- หากคลิกที่ BLAST โปรแกรมจะนำลำดับนิวคลีโอไทด์ในช่อง Nucleotide base code ไป BLAST ในฐานข้อมูล NCBI
- หากคลิกที่ mfold โปรแกรมจะวิเคราะห์การ folding ของ ลำดับนิวคลีโอไทด์
- หากคลิกที่ Check Self-Complementarity โปรแกรมจะวิเคราะห์และทำนายลักษณะการเกิด hairpin, 3' complementarity และ self-annealing (รูปที่ 7.23)

### Potential hairpin formation :

None !

### 3' Complementarity:

None !

All potential self-annealing sites are marked in red (allowing 1 mis-match) :

```

5'      GGWTGGGAYTAYCCHAARTGTGA      3'
3' AGTGTRAAHCCYATYAGGGTWGG          5'

5'      GGWTGGGAYTAYCCHAARTGTGA      3'
3' AGTGTRAAHCCYATYAGGGTWGG          5'

5'      GGWTGGGAYTAYCCHAARTGTGA      3'
3' AGTGTRAAHCCYATYAGGGTWGG          5'

5'      GGWTGGGAYTAYCCHAARTGTGA      3'
3' AGTGTRAAHCCYATYAGGGTWGG          5'
    
```

รูปที่ 7.23 ตัวอย่างการทำนายลักษณะ secondary structure ด้วยโปรแกรม Oligo Calc



➤ โปรแกรม PCR Primer Stats [[http://www.bioinformatics.org/sms2/pcr\\_primer\\_stats.html](http://www.bioinformatics.org/sms2/pcr_primer_stats.html)]

PCR Primer Stats เป็น application ใน Sequence Manipulation Suite (SMS) ใช้ในการคำนวณหาคุณสมบัติต่างๆ ของ candidate primers / probes ว่ามีความเหมาะสมต่อ PCR หรือไม่ โดยผู้ใช้สามารถใส่ข้อมูลลำดับนิวคลีโอไทด์ของ candidate primers / probes ที่เลือกเอาไว้ แล้วคลิกที่ Submit เพื่อให้โปรแกรมคำนวณคุณสมบัติต่างๆ (รูปที่ 7.23)

PCR Primer Stats results

-----  
General properties:  
-----

```

Primer name: Candidate primer
Primer sequence: GGWTGGGAYTAYCCHAARTGTGA
Sequence length: 23
Base counts: G=7; A=5; T=4; C=2; Other=5;
GC content (%): 39.13 to 56.52
Molecular weight (Daltons): 7088.67 to 7167.75
nmol/A260: 4.18 to 4.45
micrograms/A260: 29.90 to 31.60
Basic Tm (degrees C): 52 to 59
Salt adjusted Tm (degrees C): 47 to 54
Nearest neighbor Tm (degrees C): 61.65 to 68.72
    
```

PCR suitability tests (Pass / Warning):  
-----

```

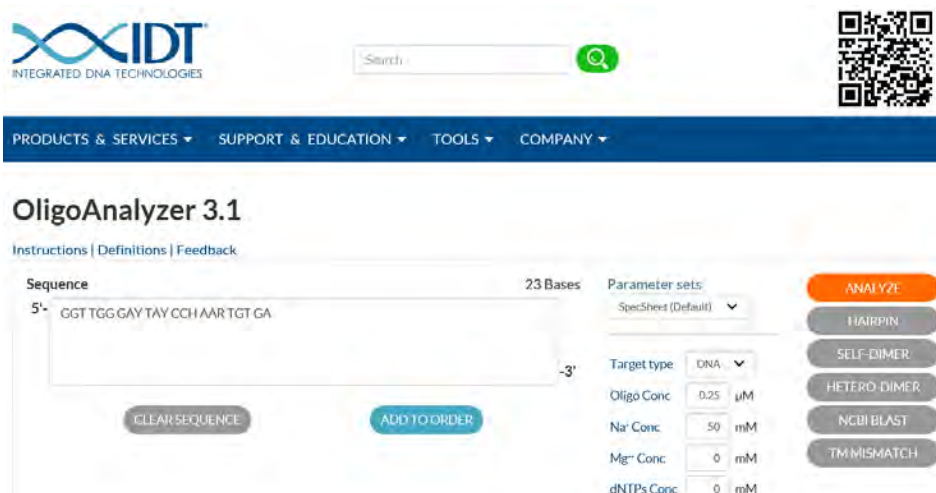
Single base runs: Pass
Dinucleotide base runs: Pass
Length: Pass
Percent GC: Warning: %GC is less than 40;
Tm (Nearest neighbor): Warning: Tm is greater than 58;
GC clamp: Pass
Self-annealing: Pass
Hairpin formation: Pass
    
```

รูปที่ 7.23 ตัวอย่างการคำนวณคุณสมบัติต่างๆ ของ primer ด้วยโปรแกรม PCR Primer Stats

➤ โปรแกรม OligoAnalyzer 3.1 [<https://sg.idtdna.com/calc/analyzer>]

เป็นโปรแกรมสำหรับการคำนวณคุณสมบัติต่างๆ ของ oligonucleotides ที่สามารถช่วยในการออกแบบ primers และ probes ได้ (รูปที่ 7.24) โดยผู้ใช้สามารถใส่ข้อมูลลำดับนิวคลีโอไทด์ของ candidate primers / probes เพื่อให้โปรแกรมคำนวณคุณสมบัติต่างๆ ได้ ดังนี้

- Analyze เป็นการวิเคราะห์คุณสมบัติพื้นฐานของ oligonucleotide เช่น sequence, complementary strand, length, %GC, Tm, molecular weight, extinction coefficient, nmole/OD และ  $\mu\text{g}/\text{OD}$  เป็นต้น (รูปที่ 7.25)
- Hairpin เป็นการวิเคราะห์ค่า thermodynamic ต่างๆ ได้แก่  $\Delta G$ ,  $\Delta H$ ,  $\Delta S$  และจำลองลักษณะการเกิด hairpin (รูปที่ 7.26)
- Self-dimer วิเคราะห์ค่า  $\Delta G$  และจำลองการเกิด self-complementarity (รูปที่ 7.27)
- Hetero-dimer วิเคราะห์ค่า  $\Delta G$  และจำลองการเกิด primer dimer (รูปที่ 7.28)
- NCBI-BLAST นำลำดับนิวคลีโอไทด์ไป BLAST ในฐานข้อมูล NCBI
- Tm mismatch วิเคราะห์ค่า Tm ที่แตกต่างกันระหว่าง exact match และ single base mismatch สำหรับใช้ในการออกแบบ allele specific primer และ allele specific probe (รูปที่ 7.29)



รูปที่ 7.24 โปรแกรม OligoAnalyzer 3.1 [<https://sg.idtdna.com/calc/analyzer>]

## RESULTS





DILUTION

RESUSPENSION

SEQUENCE	5'- GGT TGG GAY TAY CCH AAR TGT GA -3'		
COMPLEMENT	5'- TCA CAY TTD GGR TAR TCC CAA CC -3'		
LENGTH	23		
GC CONTENT	47.1%		
MELT TEMP RANGE	MIN	MEAN	MAX
	52.5 °C	56.5 °C	60.8 °C
MOLECULAR WEIGHT	7124.7 g/mole		
EXTINCTION COEFFICIENT	229446.95 L/(mole·cm)		
nmole/OD <sub>260</sub> :	4.36		
µg/OD <sub>260</sub> :	31.05		

รูปที่ 7.25 ตัวอย่างการวิเคราะห์คุณสมบัติของ primer ด้วยโปรแกรม OligoAnalyzer 3.1

## Structures

Structure Name	Image	$\Delta G(\text{kcal.mole}^{-1})$	$T_m$ (°C)	$\Delta H(\text{kcal.mole}^{-1})$	$\Delta S$ (cal.K <sup>-1</sup> .mole <sup>-1</sup> )	Output	
1		0.62	12.2	-13.7	-48.02	Ct	Det
2		1.48	-5.7	-12.9	-48.24	Ct	Det
3		1.53	-13.5	-10.3	-39.66	Ct	Det
4		1.61	-25.1	-8	-32.25	Ct	Det

รูปที่ 7.26 ตัวอย่างการวิเคราะห์ hairpin ของ primer ด้วยโปรแกรม OligoAnalyzer 3.1

## Homo-Dimer Analysis

**Dimer Sequence:**

5'- GGTTGGGAYTAYCCHAARTGTGA -3'

**Maximum Delta G: -42.74 kcal/mole**

**Delta G: -13.02 kcal/mole Base Pairs: 7**

5' GGTTGGGAYTAYCCHAARTGTGA

| | | | | | | : : : : : : :

3' AGTGTRAAHCCYATYAGGGTTGG

รูปที่ 7.27 ตัวอย่างการวิเคราะห์ homo-dimer ด้วยโปรแกรม OligoAnalyzer 3.1

## Hetero-Dimer Analysis

**Primary Sequence: 5'- GGTTGGGAYTAYCCHAARTGTGA -3'**

**Secondary Sequence: 5'- TAYTATCARAAYAATGTCTTTATGTC -3'**

**Maximum Delta G: -42.74 kcal/mole**

**Delta G: -5.55 kcal/mole Base Pairs5**

5' GGTTGGGAYTAYCCHAARTGTGA

: | | | | : : : : :

3' CTGTATTTCTGTAAYAARACTATYAT

รูปที่ 7.28 ตัวอย่างการวิเคราะห์ hetero-dimer ด้วยโปรแกรม OligoAnalyzer 3.1



➤ **ฐานข้อมูล PrimerBank** [<https://pga.mgh.harvard.edu/primerbank>]

เป็นฐานข้อมูลที่รวบรวม primers สำหรับ PCR และ real-time PCR ของยีนต่างๆ ในมนุษย์และหนูทดลอง ผู้ใช้สามารถค้นหา primers โดยใช้ข้อมูล species, accession no., หรือชื่อยีนที่สนใจ (รูปที่ 7.30) จากนั้นคลิก Submit โปรแกรมก็จะแสดงผลรายละเอียดของ primers ได้แก่ amplicon size, primer sequence, length, Tm และ location (รูปที่ 7.31)

**PrimerBank**  
PCR Primers for Gene Expression Detection and Quantification

Home/Search PCR Protocol Primer Statistics Comments **Primer Submission** Links Citation Policy Help/FAQ

**Primer Search**

**Search for PCR Primers**

Search by

Species

For text

You can blast your sequence against the primerbank sequence DB [here](#).

**Order Oligos**

You can have primers synthesized and PCR reaction products sequenced at:

**DNA Core Facility**  
Center for Computational and Integrative Biology

PrimerBank is a public resource for PCR primers. These primers are designed for gene expression detection or quantification (real-time PCR). PrimerBank contains over 306,800 primers covering most known human and mouse genes. There are several ways to search for primers: GenBank Accession, NCBI protein accession, NCBI Gene ID, Gene Symbol<sup>New!</sup>, PrimerBank ID or Keyword (gene description) or you can blast your gene sequence against the primerbank Sequence DB<sup>New!</sup>.

The primer design algorithm has been extensively tested by real-time PCR experiments for PCR specificity and efficiency. We have tested 26,855 primer pairs that correspond to 27,681 mouse genes by Real Time PCR followed by agarose gel electrophoresis and sequencing of the PCR products. The design success rate is 82.6% (22,187 successful primer pairs) based on agarose gel electrophoresis.

All experimental validation data for mouse primers are available from PrimerBank. In order to view, please follow the appropriate links seen on the primer information page.

รูปที่ 7.30 ฐานข้อมูล PrimerBank [<https://pga.mgh.harvard.edu/primerbank>]

## PrimerBank

The following matches are found for GenBank Accession (Human): "NM\_001256799"

### Gene Descriptions:

NCBI GeneID	2597
GenBank Accession	NM_001256799
NCBI Protein Accession	NP_001243728
Species	Human
Coding DNA Length	882
Gene Description	Homo sapiens glyceraldehyde-3-phosphate dehydrogenase (GAPDH), transcript variant 2, mRNA.

### Primer Pair 1 (Click here for cDNA and amplicon sequence):

PrimerBank ID 378404907c1

Amplicon Size 197

	Sequence (5' → 3')	Length	Tm	Location
Forward Primer	GGAGCGAGATCCCTCCAAAAT	21	61.6	108-128
Reverse Primer	GGCTGTTGTCATACTTCTCATGG	23	60.9	304-282

รูปที่ 7.31 ตัวอย่างผลการค้นหา primers จากฐานข้อมูล PrimerBank

### ➤ ฐานข้อมูล VirOligo [<http://virologo.okstate.edu/main.php>]

เป็นฐานข้อมูลที่รวบรวม primers สำหรับการตรวจหาชิ้นของไวรัสชนิดต่างๆ ด้วยเทคนิค PCR และ real-time PCR ผู้ใช้สามารถค้นหา primers โดยเลือกชื่อไวรัสตามลำดับตัวอักษร (รูปที่ 7.32) จากนั้นคลิกที่ชื่อของเชื้อไวรัสที่สนใจ โปรแกรมก็จะแสดงผลรายละเอียด ได้แก่ Virus, PMID, Type (PCR / real-time PCR), primer name, Tm และ Temp (รูปที่ 7.33)





A Simple Solution for Oligonucleotide Design

[About Us](#)

[FAQ](#)

[Update](#)

[Links](#)

[Database Search](#)

## Database Search

[Click here for a list of viruses in the database](#)

Please enter the search query or click the following.  
Articles without primers listed will not appear unless the PMID is entered directly.

Two search options:  
1. Search by virus name and/or primer data  
2. Search by PMID, VirOligoID, or taxonomy ID:

Virus name:

Oligo Sequence:

Tm:  °C ±

Oligo length:  bp ±

**Search**

### Viruses listed in the VirOligo database

Select Letter Index For Virus Name:  ▼

- |  |   |
|--|---|
| <a href="#">Adeno-associated virus 2</a>                 | <a href="#">Bovine rotavirus</a>              |
| <a href="#">Adult diarrheal rotavirus</a>                | <a href="#">Bovine torovirus</a>              |
| <a href="#">African swine fever virus (strain BA71V)</a> | <a href="#">Bovine viral diarrhea virus</a>   |
| <a href="#">Alcelaphine herpesvirus 1</a>                | <a href="#">Bunyamwera virus</a>              |
| <a href="#">Alphavirus</a>                               | <a href="#">Caliciviridae</a>                 |
| <a href="#">Amsacta moorei Entomopoxvirus</a>            | <a href="#">California encephalitis virus</a> |
| <a href="#">Avian infectious bronchitis virus</a>        | <a href="#">Camelpox virus</a>                |
| <a href="#">Avian influenza virus</a>                    | <a href="#">Canine adenovirus</a>             |
| <a href="#">Avian pneumovirus</a>                        | <a href="#">Canine distemper virus</a>        |
| <a href="#">B19 virus</a>                                | <a href="#">Canine herpesvirus</a>            |
| <a href="#">Banana bunchy top virus</a>                  | <a href="#">Canine parvovirus</a>             |
| <a href="#">Banzi virus</a>                              | <a href="#">Cantagalo orthopoxvirus</a>       |
| <a href="#">Barley mild mosaic virus</a>                 | <a href="#">Caprine herpesvirus 1</a>         |
| <a href="#">Barley yellow mosaic virus</a>               | <a href="#">Cauliflower mosaic virus</a>      |
| <a href="#">Bean common mosaic virus</a>                 | <a href="#">Cervid herpesvirus 1</a>          |
| <a href="#">BK polyomavirus</a>                          |   |



รูปที่ 7.32 ฐานข้อมูล VirOligo [http://virologo.okstate.edu/main.php]

## Database Search

Total 797 Oligos

<b>Common Data ID:</b> <a href="#">MP0808</a>	<b>Virus:</b> Human papillomavirus type 16, HPV-16	<b>Type:</b> PCR
<b>PMID:</b> 12067999	<b>Publication date:</b> 06/02	
<b>Temp:</b> (94C 1min>57C 1min>72C 2min)*40	<b>Type of Usage:</b>	<b>Tm:</b> 70 C at salt 1000.0mM
<b>Primer Name:</b> <a href="#">PM00418</a>	<b>Type of Usage:</b>	<b>Tm:</b> 72 C at salt 1000.0mM
<b>Primer Name:</b> <a href="#">PM00419</a>		
<b>Common Data ID:</b> <a href="#">SG0107</a>	<b>Virus:</b> Human papillomavirus type 16	<b>Type:</b> Real-time PCR
<b>PMID:</b> 14522932	<b>Publication date:</b> 09/03	
<b>Temp:</b> 50C 120sec>95C 600sec>(95C 15sec>60C 60sec)*40	<b>Type of Usage:</b> Forward Primer	<b>Tm:</b> 74.4C at salt 1000.0 mM
<b>Primer Name:</b> <a href="#">HPV 16 L1-278F</a>	<b>Type of Usage:</b> Reverse Primer	<b>Tm:</b> 71C at salt 1000.0 mM
<b>Primer Name:</b> <a href="#">HPV 16 L1-417R</a>	<b>Type of Usage:</b> PCR Probe	<b>Tm:</b> 84.6C at salt 1000.0 mM
<b>Primer Name:</b> <a href="#">HPV 16 L1-302pr</a>		

รูปที่ 7.33 ตัวอย่างผลการค้นหา primers/probes จากฐานข้อมูล VirOligo

### เอกสารอ้างอิง และแนะนำอ่านเพิ่มเติม

1. Duitama J, Kumar DM, Hemphill E, Khan M, Mandoiu II, Nelson CE. PrimerHunter: a primer design tool for PCR-based virus subtype identification. *Nucleic Acids Res.* 2009 May;37(8):2483-92.
2. Kibbe WA. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* 2007 Jul;35(Web Server issue):W43-6.
3. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. *Bioinformatics.* 2007 May 15;23(10):1289-91.
4. Onodera K, Melcher U. VirOligo: a database of virus-specific oligonucleotides. *Nucleic Acids Res.* 2002 Jan 1;30(1):203-4.
5. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol.* 2000;132:365-86.
6. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3--new capabilities and interfaces. *Nucleic Acids Res.* 2012 Aug;40(15):e115.
7. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 2007 Jul;35(Web Server issue):W71-4.
8. Wang X, Spandidos A, Wang H, Seed B. PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res.* 2012 Jan;40(Database issue):D1144-9.
9. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics.* 2012 Jun 18;13:134.

# 8

เทคนิคอาร์เอ็นเอไอในด้านอณูไวรัสวิทยา

RNAi technique in molecular virology

## เทคนิคอาร์เอ็นเอในด้านอณูไวรัสวิทยา

### RNAi technique in molecular virology

การแสดงออกของยีนเกี่ยวข้องกับกระบวนการสำคัญที่เรียกว่า Central Dogma ซึ่งประกอบด้วยกระบวนการถอดรหัส (transcription) จาก DNA ไปสู่ RNA และกระบวนการแปลรหัส (translation) จาก codon บนสาย mRNA ไปเป็นโปรตีน ในสิ่งมีชีวิตที่เป็น Eukaryotes จะมีกระบวนการ RNA interference (RNAi) ที่เกิดขึ้นโดย double stranded RNA (dsRNA) สามารถยับยั้งการแสดงออกของยีนได้อย่างจำเพาะ โดยไปมีผลยับยั้งหลังจากที่ยีนเกิดกระบวนการถอดรหัสแล้ว (post-transcriptional gene silencing) ซึ่งอาจทำให้ mRNA ของยีนนั้นถูกทำลาย (mRNA degradation) หรือยับยั้งการสร้างโปรตีน (translational repression) ของยีน

#### ประวัติการค้นพบกระบวนการ RNA interference (RNAi)

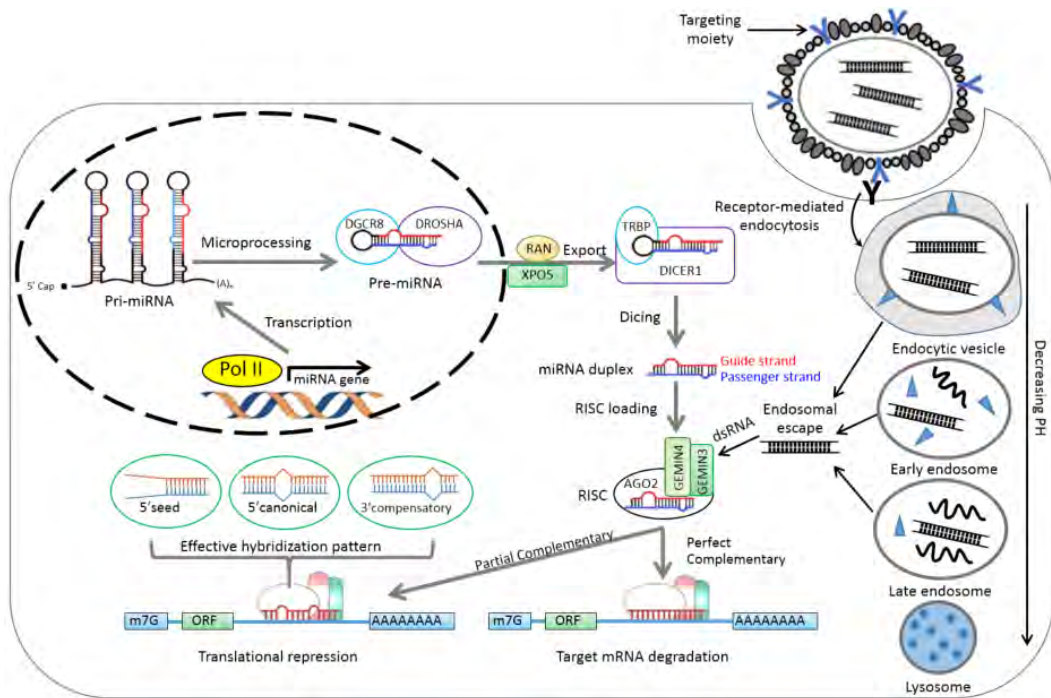
กระบวนการ RNAi ถูกค้นพบและพิสูจน์ครั้งแรก โดย Andrew Fire และ Craig Mello ซึ่งพบว่า double strand RNA (dsRNA) สามารถเหนี่ยวนำให้เกิดกระบวนการ RNAi ยับยั้งการแสดงออกของยีนได้อย่างจำเพาะ ในตัวหนอน *C. elegans* จากนั้น Andrew Fire และ Craig Mello ก็มีการทดลองและพิสูจน์กลไก RNAi ในเชิงลึกต่าง ๆ มีผลงานวิจัยตีพิมพ์เป็นที่ประจักษ์เป็นจำนวนมาก จนกระทั่งทั้งคู่ได้รับรางวัลโนเบล สาขาการแพทย์ (The Nobel prize in Physiology or Medicine) ในปี ค.ศ. 2006

#### กลไกของกระบวนการ RNA interference (RNAi)

หลังจากที่ได้มีการค้นพบหลักการของกระบวนการ RNAi ก็ได้มีการศึกษาเพิ่มเติมในเชิงลึกทำให้ทราบข้อมูลเกี่ยวกับกลไก RNAi อย่างชัดเจน โดยสามารถแบ่งออกเป็น 2 ประเภทตามชนิดของ RNA ที่มากระตุ้น ได้แก่ endogenous microRNA และ exogenous dsRNA ดังนี้

## 1. กระบวนการ RNAi ที่เกิดขึ้นโดย endogenous microRNA (miRNA)

เป็นกระบวนการธรรมชาติที่เกิดขึ้นเพื่อควบคุมการแสดงออกของยีนภายในเซลล์ ซึ่ง miRNA เป็น RNA ขนาดเล็กที่ถูกสร้างมาจาก non-coding gene (ส่วนที่ไม่ได้แปลรหัสในการสร้างโปรตีน) ปัจจุบันพบว่าในมนุษย์มียีนที่เกี่ยวข้องในการสร้าง miRNA มากกว่าหนึ่งพันยีน กระบวนการสร้าง miRNA เริ่มจากกระบวนการ transcription ภายใน nucleus โดยอาศัย RNA polymerase II ในการถอดรหัสจาก DNA ให้ได้เป็น primary microRNA (Pri-miRNA) ซึ่งเป็น RNA สายยาวมากกว่า 100 nt ประกอบด้วย 5'-cap, 3' poly A tail และโครงสร้าง secondary structure ที่เป็น hairpin-loop จากนั้นจะเข้าสู่กระบวนการ cropping โดยอาศัย microprocessing complex ซึ่งประกอบด้วย Drosha (RNase III endonuclease) และโปรตีน Digeorge Syndrome Critical Region 8 (DGCR8) โดยโปรตีน DGCR8 จะมี RNA recognizing domain ที่ใช้ในการจับกับ hairpin-loop ของ Pri-miRNA แล้วเอนไซม์ Drosha จะทำหน้าที่ตัด Pri-miRNA ให้มีขนาดสั้นลง ได้เป็น precursor microRNA (Pre-miRNA) ขนาดประมาณ 60-70 nt ที่ยังมีลักษณะเป็น hairpin-loop จากนั้น Pre-miRNA จะถูกขนส่ง (export) ออกจากนิวเคลียสไปสู่ cytoplasm โดยอาศัยโปรตีนขนส่ง Exportin-5 และพลังงานจาก Ran-GTP จากนั้นใน cytoplasm จะมีเอนไซม์ Dicer (RNase III endonuclease) ตัด Pre-miRNA ที่บริเวณ hairpin-loop ได้เป็น mature miRNA duplex ซึ่งมีลักษณะเป็น dsRNA ขนาดประมาณ 21-23 nt แล้วเข้าจับกับ RNA Induced Silencing Complex (RISC) ซึ่งประกอบด้วยโปรตีนหลายชนิด ได้แก่ argonaute-2 (AGO2), GEMIN3 และ GEMIN4 จากนั้น GEMIN3 ซึ่งมี helicase activity จะทำการแยกสาย miRNA duplex ให้กลายเป็น miRNA สายเดี่ยว (guide strand) เพื่อนำ RISC ไปจับกับยีน mRNA เป้าหมายที่มีลำดับเบสคู่สมกับ guide strand miRNA ซึ่งโดยส่วนใหญ่ miRNA จะจับกับ mRNA แบบ partial complementary ส่งผลให้เกิดการยับยั้งการสร้างโปรตีน (translational repression) ดังรูปที่ 8.1



รูปที่ 8.1 กลไกของกระบวนการ RNA interference (RNAi)

## 2. กระบวนการ RNAi ที่เกิดขึ้นโดย exogenous double stranded RNA (dsRNA)

เป็นกระบวนการที่เกิดจากการเหนี่ยวนำให้ dsRNA จากภายนอก (exogenous dsRNA) เข้าสู่ cytoplasm ของเซลล์ โดยอาศัยพาหะ (carriers) หาก dsRNA ที่ใส่เข้าไปในเซลล์มีขนาดใหญ่ก็就会被ตัดโดยเอนไซม์ Dicer (RNase III endonuclease) ให้กลายเป็น short interfering RNA (siRNA) ซึ่งมีลักษณะเป็น dsRNA ขนาดประมาณ 21-23 nt แล้วเข้าจับกับ RISC จากนั้น GEMIN3 ซึ่งมี helicase activity จะทำการแยกสาย siRNA duplex ให้กลายเป็น siRNA สายเดี่ยว (guide strand) เพื่อนำ RISC ไปจับกับยีน mRNA เป้าหมายที่มีลำดับเบสคู่สมกับ guide strand siRNA ซึ่งโดยส่วนใหญ่ siRNA จะจับกับ mRNA แบบ perfect complementary เนื่องจากเป็นการออกแบบให้มีลำดับเบสเหมือนกับยีนที่ต้องการจะยับยั้งการแสดงออก จึงส่งผลให้เกิดกระบวนการย่อยสลาย mRNA (mRNA cleavage) โดยอาศัยการทำงานของเอนไซม์ AGO2 ดังรูปที่ 8.1

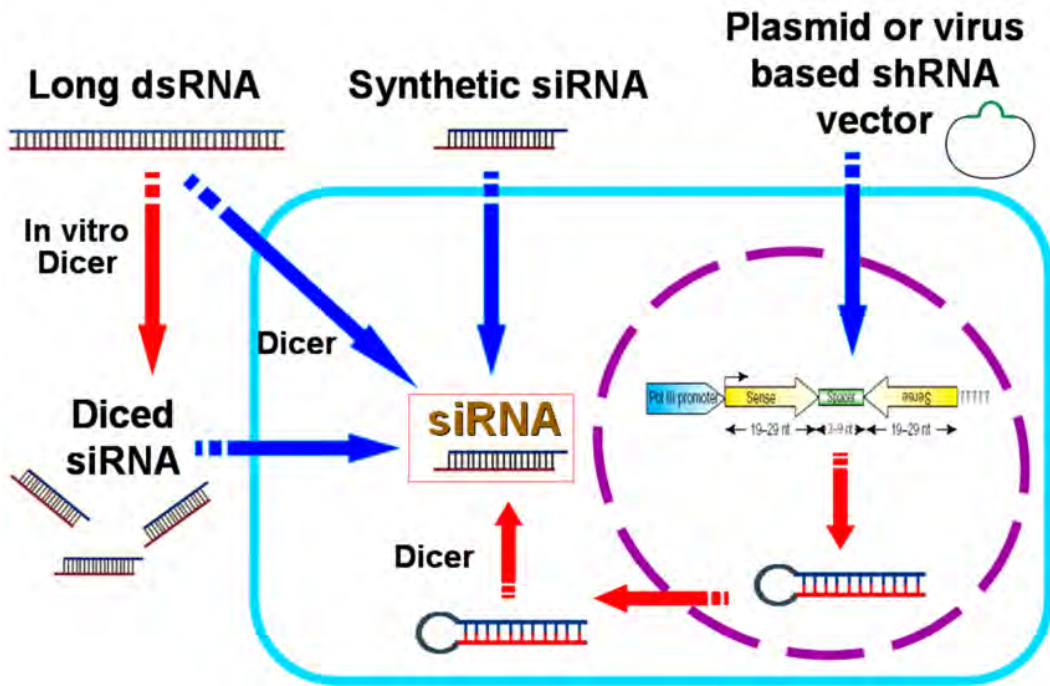
## วิวัฒนาการร่วมของกระบวนการ RNAi ใน eukaryotes

จากการศึกษาที่ผ่านมาพบว่า กระบวนการ RNAi สามารถพบได้ในสิ่งมีชีวิต eukaryotes ประเภทต่าง ๆ ได้แก่ ยีสต์ (*Saccharomyces pombe*), พืช (*Arabidopsis thaliana*), แมลง (*Drosophila melanogaster*), หนอน (*Caenorhabditis elegans*) และมนุษย์ (*Homo sapiens*) เป็นต้น ซึ่งมีกลไกหลักของกระบวนการ RNAi คล้ายคลึงกันมาก ดังนั้นนักวิทยาศาสตร์จึงมีความเชื่อมั่นว่า RNAi เป็นกระบวนการที่สำคัญของเซลล์ eukaryotes สำหรับการควบคุมการแสดงออกของยีนผ่าน endogenous miRNA ซึ่งในปัจจุบันพบว่า miRNA มีบทบาทสำคัญในกระบวนการต่างๆ ของเซลล์ ได้แก่ cell cycle, cell proliferation, apoptosis, cell differentiation, metabolism และ development เป็นต้น นอกจากนี้ RNAi ที่ตอบสนองต่อ exogenous dsRNA ยังทำหน้าที่เป็น defense mechanism เพื่อป้องกันการติดเชื้อไวรัสด้วย เนื่องจากไวรัสส่วนใหญ่จะมีสารพันธุกรรมเป็น RNA

## เทคโนโลยีในการสร้าง siRNA

จากการศึกษากระบวนการ RNAi ที่เกิดขึ้นโดย exogenous dsRNA พบว่า โมเลกุลหลักที่มีความสำคัญก็คือ short interfering RNA (siRNA) ซึ่งเป็นตัวกำหนดความจำเพาะว่าจะไปยับยั้งการแสดงออกของยีนใด การทำให้เซลล์ได้รับ siRNA ดังกล่าวนี้อาศัยวิธีการ (รูปที่ 8.2) ดังนี้





รูปที่ 8.2 เทคโนโลยีในการสร้าง siRNA ประเภทต่างๆ

- **Long dsRNA**: เป็นการใส่ dsRNA สายยาวเข้าสู่เซลล์ ซึ่งหลังจากเข้าสู่เซลล์แล้ว dsRNA จะถูกเอนไซม์ Dicer ใน cytoplasm ตัดให้กลายเป็น siRNAs หลาย ๆ โมเลกุล เพื่อเหนี่ยวนำให้เกิดการยับยั้งการแสดงออกของยีนต่อไป ข้อดีของวิธีนี้คือนักวิจัยสามารถสังเคราะห์ long dsRNA ขึ้นเองได้ในห้องปฏิบัติการ โดยใช้เทคนิค *in vitro* transcription ซึ่งจะได้ปริมาณ dsRNA ที่เพียงพอและประหยัดค่าใช้จ่าย แต่มีข้อจำกัดคือในสัตว์มีกระดูกสันหลัง (vertebrate) หากใส่ dsRNA ที่มีขนาดความยาวมากกว่า 50 bp จะเหนี่ยวนำให้เกิด interferon response ซึ่งจะมีการส่งสัญญาณกระตุ้นการทำงานของเอนไซม์ RNaseL ทำให้เกิดการทำลาย RNA ทั้งหมดภายในเซลล์ (global RNA degradation) นอกจากนี้ยังมีสัญญาณยับยั้งการทำงานของ eIF-2 $\alpha$  ทำให้เกิดการยับยั้งการสร้างโปรตีนทั้งหมดภายในเซลล์ (global translation repression) ผลที่ตามมาคือเซลล์ที่ได้รับ long dsRNA (>50 bp) จะตาย ดังนั้นวิธีนี้จึงไม่นิยม

- ***In vitro* diced siRNA:** เป็นการปรับปรุงข้อเสียของ long dsRNA โดยหลังจากการสังเคราะห์ long dsRNA แล้ว จะนำมาตัดด้วยเอนไซม์ dicer (ที่ซื้อมาจากบริษัท) ในหลอดทดลอง (*in vitro*) ให้ได้เป็น pool siRNAs ที่มีขนาดประมาณ 21-23 nt แล้วจึงนำไปใส่เข้าสู่เซลล์ ซึ่งจะไม่เกิดการเหนี่ยวนำ interferon response ใน vertebrate อย่างไรก็ตามจากผลการศึกษาพบว่าเกิดการยับยั้งการแสดงออกของยีนอย่างไม่จำเพาะ (non-specific genes silencing หรือ off-target) เนื่องจาก long dsRNA ที่ถูกตัดด้วย dicer จะมีการตัดแบบสุ่ม ทำให้ได้ pool siRNAs หลายโมเลกุลที่มีลำดับเบสแตกต่างกันไป เมื่อนำใส่เข้าไปในเซลล์ siRNAs แต่ละตัวอาจไปจับกับ mRNA ได้หลายยีน (โดยเฉพาะกลุ่ม gene family) เป็นผลให้เกิดการยับยั้งการแสดงออกของยีนอย่างไม่จำเพาะ
- **Synthetic siRNA:** เป็นการสังเคราะห์ siRNA โดยใช้ปฏิกิริยาเคมี ที่ใช้สารตั้งต้นเป็น ribonucleoside phosphoamidite เกิดปฏิกิริยา coupling, capping, oxidation และ deprotection วนเวียนไปจนกระทั่งได้ siRNA ขนาด 21-23 nt ที่มีลำดับเบสตามต้องการ ข้อดีของเทคนิคนี้คือสามารถสร้าง siRNA ที่มีลำดับเบสตามต้องการได้อย่างถูกต้อง ดังนั้นจึงให้ผลการยับยั้งการแสดงออกของยีนอย่างจำเพาะ (specific silencing) และสามารถสังเคราะห์เป็น modified siRNA ได้ ซึ่งมีประโยชน์ เช่น เพิ่มความเสถียรของ siRNA เพิ่มประสิทธิภาพในการยับยั้งการแสดงออกของยีน และเพิ่มความจำเพาะต่อ target mRNA เป็นต้น นอกจากนี้สามารถเลือกสังเคราะห์ siRNAs หลาย ๆ ตำแหน่งเพื่อใช้เป็น pool siRNAs ยับยั้งยีนเดียวกันได้อย่างมีประสิทธิภาพ แต่มีข้อเสียคือค่าใช้จ่ายสูง และมีระยะเวลาในการยับยั้งการแสดงออกของยีนค่อนข้างสั้น (transient effect)

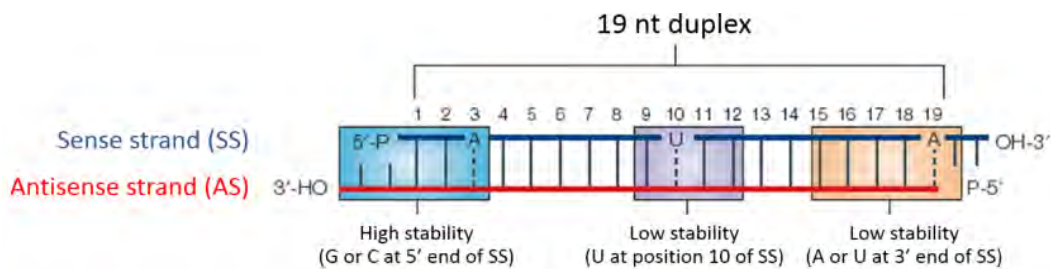
- **Short hairpin RNA (shRNA) expression vector:** เป็นการใช้ DNA vector ที่มีลักษณะเป็น DNA plasmid เมื่อใส่เข้าไปในเซลล์แล้ว จะเคลื่อนที่เข้าสู่นิวเคลียส แล้วเกิดกระบวนการถอดรหัส (transcription) โดยอาศัยการทำงานของ RNA polymerase ได้เป็น RNA ที่มีลักษณะเป็น hairpin ขนาดสั้น ๆ (ไม่เกิน 50 bp) จึงเรียกว่า short hairpin RNA (shRNA) จากนั้นเมื่อออกมาสู่ cytoplasm ก็จะถูกเอนไซม์ dicer ตัดส่วนที่เป็น hairpin loop ออกไป กลายเป็น siRNA เพื่อไปยับยั้งการแสดงออกของยีนต่อไป วิธีการนี้มีข้อดีหลายด้าน เช่น สามารถทำได้เองในห้องปฏิบัติการ เพิ่มปริมาณ DNA plasmid ได้มาก ประหยัดค่าใช้จ่าย ควบคุมการแสดงออกของ shRNA ได้ (inducible vector) ระยะเวลาในการยับยั้งการแสดงออกของยีนค่อนข้างนาน (stable silencing) ข้อเสียคือต้องทำการ cloning vector และการออกแบบ shRNA ที่จำเพาะต่อยีนมีความซับซ้อน

### การเปรียบเทียบเทคโนโลยีการสร้าง siRNA

จากการเปรียบเทียบพบว่า long dsRNA และ *in vitro* diced siRNA ไม่ค่อยมีความโดดเด่น และยังมีข้อจำกัดค่อนข้างมากดังที่กล่าวไปแล้วข้างต้น นอกจากนี้ long dsRNA และ *in vitro* diced siRNA ไม่สามารถใช้สำหรับการทดลองในสิ่งมีชีวิต (*in vivo*) ดังนั้นจะเห็นได้ว่าเทคโนโลยีการสร้าง siRNA ที่ได้รับความนิยมในปัจจุบัน คือ synthetic siRNA และ shRNA expression vector ซึ่งทั้ง 2 เทคโนโลยีมีจุดเด่นในด้าน ต่าง ๆ ได้แก่ scalable (สามารถเพิ่มปริมาณได้) และ *in vivo* studies (สามารถใช้สำหรับการทดลองในสิ่งมีชีวิตได้) อย่างไรก็ตาม synthetic siRNA มีข้อได้เปรียบในด้าน modification สามารถปรับเปลี่ยนโมเลกุลของ siRNA ได้ตามต้องการ และ fidelity (มีความถูกต้องของลำดับเบสถูกต้อง 100%) ในขณะที่ shRNA expression vector มีข้อได้เปรียบในด้าน stable silencing (ยับยั้งการแสดงออกของยีนได้ยาวนานหลายวัน) และ inducible expression (สามารถกำหนดและควบคุมการแสดงออกของ shRNA ได้)

## การออกแบบ siRNA

ตามมาตรฐาน siRNA ที่สามารถยับยั้งการแสดงออกของยีนได้อย่างมีประสิทธิภาพ (effective gene silencing) คือใช้ siRNA ที่ความเข้มข้น 100 nM สามารถยับยั้งการแสดงออกของยีน (gene knockdown) โดยทำให้ระดับของ target mRNA มีปริมาณลดลงได้ไม่ต่ำกว่า 70% จะถือว่าเป็น functional siRNA จากการศึกษาวิจัยเพื่อทดสอบหาคุณสมบัติพื้นฐานของ functional siRNA พบว่าการออกแบบ siRNA ควรจะมีลักษณะ (รูปที่ 8.3) ดังต่อไปนี้



รูปที่ 8.3 ลักษณะของ siRNA ตามหลักการของ rational siRNA design guidelines

### ➤ ลักษณะพื้นฐาน (basic rules)

- siRNA ประกอบด้วย RNA 2 สาย แต่ละสายมีความยาว 21 bp เมื่อมาเข้าคู่กันจะได้ RNA duplex (double stranded RNA) ขนาด 19 bp มี 2 bp overhang ที่ปลาย 3'
- %GC contents ควรอยู่ในช่วงระหว่าง 30-52% เพื่อให้การจับมีประสิทธิภาพที่ดี
- หลีกเลี่ยงยีนบริเวณ introns เนื่องจาก introns ไม่ได้อยู่บน mature mRNA
- หลีกเลี่ยงยีนบริเวณ untranslated regions (UTR) เนื่องจากบริเวณ UTR จะมีโปรตีนมาจับอยู่ ทำให้ siRNA ไม่สามารถเข้าไปจับบริเวณ UTR ได้
- หลีกเลี่ยงยีนบริเวณหลังจาก start codon (ATG) อย่างน้อย 75 nt เนื่องจากบริเวณนี้อาจมี ribosome มาเกาะอยู่ ทำให้ siRNA ไม่สามารถเข้าไปจับได้
- siRNA จะต้องมีความจำเพาะต่อ mRNA ของยีนที่ต้องการยับยั้งเท่านั้น ไม่ได้รวมจับกับยีนอื่น มิฉะนั้นจะทำให้เกิด non-specific gene silencing

➤ ลักษณะพิเศษเพิ่มเติม (additional guidelines)

- ไม่มี internal repeat เช่น TTTT, AAAA, CCCC หรือ GGGG เป็นต้น เพราะจะทำให้ความจำเพาะของ siRNA ลดลง
- มีเสถียรภาพสูงทางด้าน 5' ของ sense strand (ควรมีเบส G หรือ C)
- มีเสถียรภาพต่ำทางด้าน 3' ของ sense strand (ควรมีเบส A หรือ U)
- ตำแหน่ง 3, 10 และ 19 ของ sense strand ควรเป็น "A", "U" และ "A" ตามลำดับ

ซึ่งหลักเกณฑ์ต่างๆ ข้างต้นมีเหตุผลอธิบายประกอบ ดังสรุปในตารางที่ 8.1

ตารางที่ 8.1 อธิบายเหตุผลชี้แจงหลักเกณฑ์ในการออกแบบ siRNA

Criteria	Probable reason
30-50% G+C content	Facilitates interaction with RISC and unwinding
Low stability at the 5' antisense strand	Promotes antisense-strand selection
High stability at the 5' sense strand	Blocks sense-strand selection
Absence of internal repeats or palindromes	Increases the concentration of functional siRNA
"A" at position 3 and 19 of sense strand	Promotes antisense-strand selection
"U" at position 10 of sense strand	Promotes RISC mediated cleavage of mRNA and dissociation of the RISC-siRNA complex
Perform stringent homology searches	Minimizes potential nonspecific gene silencing

## การสร้าง shRNA expression vector

การสร้าง shRNA expression vector ประกอบด้วยขั้นตอน 6 ขั้นตอน ดังนี้

1. Identify target site: ใช้หลักการของ rational siRNA design guideline หรือ online siRNA design tool ในการออกแบบลำดับของ siRNA สาย sense target sequence

2. Design an appropriate shRNA: จากข้อที่ 1 สิ่งที่ได้คือส่วนของ sense target sequence (19-21 nt) จากนั้นนำมาทำ reverse & complement จะได้ส่วนของ antisense target sequence (19-21 nt) นำ loop sequence (4-10 nt) มาเชื่อมต่อระหว่างส่วนของ sense และ antisense target sequences (รูปที่ 8.4)



รูปที่ 8.4 การออกแบบและสังเคราะห์ oligonucleotides สำหรับการสร้าง shRNA vector

3. Synthesize a dsDNA fragment that codes for the shRNA: จากข้อที่ 2 สิ่งที่ได้คือ oligonucleotide สายยาวที่เรียกว่า Top strand oligo จากนั้นนำมาทำเป็นสายที่มีลำดับเบสคู่สม (complementary strand) จะได้ Bottom strand oligo แล้วต่อปลาย 5' ด้วยลำดับเบสที่เหมาะสมต่อการ ligation เข้าสู่ vector (รูปที่ 8.4) เมื่อได้ลำดับเบสของ top และ bottom strand oligos แล้ว สามารถสั่งให้บริษัทที่ให้บริการทำการสังเคราะห์ oligonucleotides

4. Clone the fragment into an RNA expression vector: จากข้อ 3 เมื่อได้ oligonucleotides มาแล้วนำ Top และ Bottom strands มาทำการ anneal กันใน annealing buffer จะได้เป็น dsDNA จากนั้นนำไปเชื่อม (ligation) ต่อกับ shRNA expression vector (รูปที่ 8.5) โดยอาศัยเอนไซม์ ligase

5. Transform into bacterial culture: จากข้อ 4 นำ recombinant vector มา transform เข้าสู่ competent cells (*E.coli*) ด้วยวิธี heat shock หรือ electroporation จากนั้นนำไป spread บน agar plate ที่ผสมด้วย antibiotic ที่เหมาะสมกับ vector ที่ใช้ (เช่น ถ้าใน vector มียีนที่ต้านยา Kanamycin ให้ใส่ยา Kanamycin ลงใน agar เป็นต้น)

6. Select colony and purify recombinant plasmids: คัดเลือก colony ที่สามารถทนต่อยาปฏิชีวนะได้ จากนั้นนำมาเลี้ยงเพิ่มจำนวนใน LB broth แล้วทำการสกัด plasmid จากนั้นอาจทำการยืนยันผลโดยการทำ nucleotide sequencing เพื่อให้แน่ใจว่า plasmid ดังกล่าวมีลำดับของ shRNA ถูกต้อง

### โปรแกรมคอมพิวเตอร์สำหรับการออกแบบ siRNA และ shRNA

ในปัจจุบันมีนัก bioinformatics ได้เขียนโปรแกรมสำหรับช่วยอำนวยความสะดวกในการออกแบบ siRNA และ shRNA โดยผู้ใช้สามารถใส่ลำดับนิวคลีโอไทด์ของยีนที่สนใจเข้าสู่โปรแกรม แล้วสั่งให้โปรแกรมทำการคัดเลือกและออกแบบ siRNA และ shRNA ที่มีลักษณะเหมาะสมได้ ทั้งนี้แต่ละโปรแกรมหรือแต่ละ website ก็จะมีหลักการเบื้องต้นในการออกแบบคล้ายๆ กัน แต่จะมีความแตกต่างกันไปใน criteria ของ additional guidelines ดังนั้นผู้ใช้งานจึงควรตระหนักถึงประเด็นที่มีความแตกต่างกันด้วย ตัวอย่างของโปรแกรมที่นิยมใช้ในการออกแบบ siRNA หรือ shRNA มีดังนี้



➤ **DSIR** [<http://biodev.extra.cea.fr/DSIR/DSIR.html>]

DSIR (Designer of Small Interfering RNA) เป็นโปรแกรม web-based สำหรับใช้ในการออกแบบ siRNA และ shRNA โดยผู้ใช้งานสามารถใส่ลำดับนิวคลีโอไทด์ในรูปแบบ FASTA format ของยีนที่สนใจเข้าสู่โปรแกรม จากนั้นตั้งค่าต่างๆ ในการออกแบบ (รูปที่ 8.5) ได้แก่

- RNA design เลือกออกแบบ siRNA (19 nt หรือ 21 nt) หรือ shRNA
- Score threshold ตั้งค่าคะแนนที่จะใช้ในการคัดกรองการออกแบบ
- Filter 4 or more nucleotide runs: ตั้งค่าเพื่อหลีกเลี่ยงลำดับเบสที่ซ้ำกัน (>4 nt)
- Avoid immunostimulatory motif: ตั้งค่าเพื่อหลีกเลี่ยงส่วนที่สามารถกระตุ้นระบบภูมิคุ้มกัน

เมื่อคลิกที่คำว่า Run analysis โปรแกรมจะทำการออกแบบ siRNA หรือ shRNA ออกมาให้เลือกจำนวนหลายโมเลกุล (รูปที่ 8.6) โดยเรียงตามลำดับคะแนนของการออกแบบจากมากไปหาน้อย ซึ่งแสดงข้อมูลต่างๆ ดังนี้

- siRNA\_id เป็นการระบุ ID ของ siRNA
- Position ระบุตำแหน่งของ siRNA บน target gene
- SS sequence ลำดับนิวคลีโอไทด์ของ sense strand
- AS sequence ลำดับนิวคลีโอไทด์ของ antisense strand
- Score คะแนนของการออกแบบ ถ้าคะแนนมาก จะมีความน่าเชื่อถือมาก

นอกจากนี้เมื่อสนใจ siRNA ตัวใด ผู้ใช้สามารถคลิกเลือกที่ช่องด้านหน้า siRNA\_id เพื่อสั่งให้โปรแกรมทำ similarity search เปรียบเทียบกับลำดับนิวคลีโอไทด์ในฐานข้อมูล และคำนวณค่า Seed Complement Frequencies เพื่อตรวจสอบว่าลำดับนิวคลีโอไทด์ของ siRNA ดังกล่าวมีความคล้ายคลึงกับยีนอื่นๆ หรือไม่ เพื่อให้แน่ใจว่า siRNA ที่ออกแบบนั้นมีความจำเพาะต่อยีนเป้าหมายเท่านั้น



## Welcome to DSIR

DSIR is a tool for siRNA (19 or 21 nt) and shRNA target design. It implements an algorithm based on the following publication: Vert JP, Foveau N, Lajaunie C, Vandenbrouck Y. An accurate and interpretable model for siRNA efficacy prediction. BMC Bioinformatics. 2006 Nov 30;7(1):520. Recently a comparison study based on four independent siRNA datasets ranked DSIR as one of the best predictors among several programs evaluated [Matveeva et al., 2007]

### Sequence to target

First, name your design session and then enter your target sequence. You can either use the input field by simply copy-paste your sequence or upload a sequence fasta file. In both cases, your sequence has to respect the FASTA format (or DSIR will throw an error).

Design name :  \* required

Sequence :

(FASTA format only)

```
>Target gene
ATGGAACCATTAAGACAGGAGCCCATATCGAACCCCTAATGAGCTGT
CCTATTGGTGAAGTTCCTCTCCATACAACCTCAAGATTTGAGTCAGT
CGCTTGGTCAGCAAGTGCTTGTCATGATGGCATCAATTGGCTAACA
ATTGGAATTTCTGGCCCAGACAATGGGGCAGTGGCTGTGTTAAAGT
ACAACGGCATAATAACAGACACTATCAAGAGTTGGAGAAAACAATATA
TTGAGAACACAAGAGTCTGAATGTGCATGTGTAATGGTTCTTGCTT
TACTGTAATGACCGATGGACCAAGTGATGGACAGGCCTCATACAAG
```

or upload file :

### Prediction settings

The first analysis performed by DSIR is the prediction of the efficacy of all possible siRNAs targeting your input sequence. The efficacy (or score) computed by DSIR is based on a linear model that combines siRNA and target sequence features.

RNA design :  siRNA 19nt  siRNA 21nt  shRNA

Score threshold :

Filter 4 or more nucleotide runs :  Yes  No

Avoid immunostimulatory motifs :  Yes  No

รูปที่ 8.5 โปรแกรม DSIR [<http://biodev.extra.cea.fr/DSIR/DSIR.html>]

<input type="checkbox"/>	siRNA_id	Position	SS Sequence	AS Sequence	Score	Corrected score
<input type="checkbox"/>	1	268	GGUUCUUGCUUUACUGUAAUG	UUACAGUAAAGCAAGAACCAU	103.0	101.0
<input type="checkbox"/>	2	213	GAGUUGGAGAAACAUAUAUU	UAUAUUGUUUCUCCAACUCUU	98.7	96.7
<input type="checkbox"/>	3	486	GAAUCUGGAAUAUCAGAUAGG	UAUCUGAUAUUCCAGAUUCUG	96.2	92.2
<input type="checkbox"/>	4	39	GAGCUGUCCUAUUGGUGAAGU	UUCACCAAUAGGACAGCUCAU	96.2	96.2
<input type="checkbox"/>	5	61	CCCUCUCCAUAACAUCUAGA	UUGAGUUGUAUGGAGAGGGAA	95.5	95.5
<input type="checkbox"/>	6	202	GACACUAUCAAGAGUUGGAGA	UCCAACUCUUGAUAGUGUCUG	94.9	92.9
<input type="checkbox"/>	7	559	GGCAGUUGUGGUCCAGUAUCG	AUACUGGACCACAACUGCCUG	94.8	89.8
<input type="checkbox"/>	8	291	CGAUGGACCAAGUGAUGGACA	UCCAUCACUUGGUCCAUCGGU	94.1	92.1
<input type="checkbox"/>	9	191	GCAUAAUAACAGACACUAUCA	AUAGUGUCUGUUUAUUGCCG	93.8	92.8
<input type="checkbox"/>	10	558	AGGCAGUUGUGGUCCAGUAUC	UACUGGACCACAACUGCCUGU	93.8	88.8

### Similarity search

Check how siRNA sequence are specific to your target sequence.

Mismatch tolerance :

Data bank :

Compute SCF (Seed Complement Frequencies).

using seed length:

Export results :

### รูปที่ 8.6 ตัวอย่างผลการออกแบบ siRNA ด้วยโปรแกรม DSIR

➤ **BiOSETTIA: The shRNA Designer** [<http://biosettia.com/support/shrna-designer/>]

เป็นโปรแกรม web-based สำหรับใช้ในการออกแบบ shRNA โดยผู้ใช้สามารถใส่ accession no. ของ RefSeq หรือลำดับนิวคลีโอไทด์ในรูปแบบ FASTA format ของยีนที่สนใจ เข้าสู่โปรแกรม จากนั้นตั้งค่า parameters ต่างๆ ในการออกแบบ (รูปที่ 8.7) แล้วคลิกที่คำว่า Search/Convert Targets โปรแกรมจะทำการออกแบบ shRNA ออกมาให้เลือกจำนวนหลายโมเลกุล (รูปที่ 8.8)



Use this program to design shRNA oligos that are compatible with our SORT-A/B/C vectors. The design tool provides targets with the greatest chance of knocking down your gene. Please note, only one oligo is designed as it is palindromic. For more information, please download our [protocol](#) or visit <http://biosettia.com/shrna/>.

RefSeq Accession  Min GC%  Max GC%

Example: NM\_002529 Recommend GC Content: 35-55% Results displayed below

Method 1  Method 2  CDS  (Strict parameters)  Target

Nucleotide sequence to search for efficient shRNA  
-OR-  
Convert your own targets into Biosettia compatible oligos (one per line) e.g.:

```
GCCTGAATCTCTCCTTCAA
GCTTCATCTTCACTGAGTT
...
```

### รูปที่ 8.7 โปรแกรม BiOSETTIA: The shRNA Designer

[<http://biosettia.com/support/shrna-designer/>]

## Custom Sequence Query

Gene name: Custom (Custom)

Found on chromosome 0

CDS region: 0 - 0

GC Content range: 35-55

Pos	Location	GC%	shRNA Oligos: <span style="color: red;">overhang</span> -target- <span style="color: red;">loop</span> -antisense	Select
123	-	42.11	AAAAGGCATCAATTGGCTAACAATTGGATCCAAATTGTTAGCCAATTGATGCC	<input type="checkbox"/>
124	-	36.84	AAAAGCATCAATTGGCTAACAATTGGATCCAAATTGTTAGCCAATTGATGC	<input type="checkbox"/>
270	-	36.84	AAAAGGTTCTTGCTTTACTGTAATTGGATCCAAATTACAGTAAAGCAAGAACC	<input type="checkbox"/>

### รูปที่ 8.8 ตัวอย่างผลการออกแบบ shRNA ด้วยโปรแกรม BiOSETTIA: The shRNA Designer

➤ InvivoGen: siRNA Wizard v3.1 [<http://www.invivogen.com/sirnazizard/>]

เป็นโปรแกรม web-based สำหรับใช้ในการออกแบบ siRNA หรือ shRNA (รูปที่ 8.9) ซึ่งประกอบด้วย 3 ส่วน ได้แก่



### Selection of siRNA/shRNA targets

InvivoGen's siRNA Wizard™ is a software designed to help you select siRNA/shRNA sequences targeting your gene(s) of interest. This program selects siRNA/shRNA sequences that match criteria suggested by studies of RNA interference and which will have the best expression rate in [psiRNA vectors](#).

siRNA Wizard™ is composed of 3 parts:

- **Find siRNA/shRNA sequence**

Two types of searches can be performed to find siRNA/shRNA sequences:

- "Standard search" utilizes a default set of criteria to analyze your gene of interest and provide the best sequences to silence gene expression.
- "Advanced search" lets you manually set the criteria for selecting the sequences against your target gene.

- **Design hairpin insert**

Using your selected siRNA/shRNA sequence, this tool will design two complementary oligonucleotides forming the insert needed to complete InvivoGen's psiRNA vector.

- **Scramble siRNA/shRNA**

Generate a negative control for siRNA/shRNA. This tool will return a scrambled sequence with the same nucleotide composition as your siRNA/shRNA input sequence.

### รูปที่ 8.9 InvivoGen: siRNA Wizard v3.1 [<http://www.invivogen.com/sirnazizard/>]

- Find siRNA/shRNA sequence ผู้ใช้สามารถใส่ลำดับนิวคลีโอไทด์ในรูปแบบ FASTA format ของยีนที่สนใจเข้าสู่โปรแกรม จากนั้นตั้งค่า parameters ต่างๆ ในการออกแบบ (รูปที่ 8.10) แล้วคลิกที่คำว่า Search โปรแกรมจะทำการออกแบบ shRNA ออกมาให้เลือกจำนวนหลายโมเลกุล (รูปที่ 8.11) โดยแสดงค่า ตำแหน่งของ siRNA และ GC%

## Find siRNA sequences - Standard search

### Instructions

1. Enter your siRNA name for reference (optional)
  2. Paste the sequence of your target gene. (Spaces and numbers will be ignored)
  3. Enter the desired motif size (21 nt by default).
  4. If you want to increase the siRNA's specificity by removing non-unique sequences, choose the database corresponding to the target gene. (optional)
  5. To avoid off target effect, choose miRNA SEED database corresponding to the target gene. (optional)
- Only siRNA presenting homology with less than 5 Unigene sequences are shown, and the list of homolog genes found is displayed

siRNA Name :

Coding Sequence (max: 9600bp):

Motif size :  nt

mRNA Database:

miRNA SEED Database:

รูปที่ 8.10 การใส่ข้อมูลและตั้งค่าสำหรับโปรแกรม InvivoGen: siRNA Wizard

### 12 siRNA candidate target sequences of 21 nt found

Sequence	Start	GC%	Design	Scrambled
GCCCATATCGAACCCCTAATGA	22	47.62	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
GAACCCATATGAGCTGTCCTA	31	47.62	<input type="checkbox"/>	<input type="checkbox"/>
GTGAAGTTCCTCTCCATACA	55	47.62	<input type="checkbox"/>	<input type="checkbox"/>
GTTCCCTCCATACAACCTCA	60	47.62	<input type="checkbox"/>	<input type="checkbox"/>
GTTAAAGTACAACGGCATAAT	179	33.33	<input type="checkbox"/>	<input type="checkbox"/>
GCATAATAACAGACACTATCA	193	33.33	<input type="checkbox"/>	<input type="checkbox"/>
GTGATGGACAGGCCCTCATA	304	52.38	<input type="checkbox"/>	<input type="checkbox"/>
GACAGGCCCTCATAAGATCT	310	47.62	<input type="checkbox"/>	<input type="checkbox"/>
GAAATGAATGCCCTAATTAT	366	33.33	<input type="checkbox"/>	<input type="checkbox"/>
GCTCCTGTTATCCTGATTCTA	400	42.86	<input type="checkbox"/>	<input type="checkbox"/>
GGTGTCTTTCAACCAGAATCT	473	42.86	<input type="checkbox"/>	<input type="checkbox"/>
GACAATCCAGCCCTAATGAT	534	47.62	<input type="checkbox"/>	<input type="checkbox"/>

#### For generating the hairpin insert :

Loop sequence :  psiRNA Vector for cloning :

#### For generating scrambled siRNA, select database :

รูปที่ 8.11 ตัวอย่างผลการออกแบบ siRNA ด้วยโปรแกรม InvivoGen: siRNA Wizard

- Design hairpin insert ผู้ใช้สามารถแปลงลำดับนิวคลีโอไทด์ของ siRNA ไปเป็นลำดับของ oligonucleotides ที่จะใช้ในการสร้าง shRNA expression vector โดยสามารถกำหนดลำดับนิวคลีโอไทด์ของ Loop sequence และเลือกชนิดของ vector ที่ต้องการใช้ได้ จากรูปที่ 8.11 ถ้าคลิกที่คำว่า Design hairpin ก็จะได้แสดงผลลำดับของ oligonucleotides (รูปที่ 8.12)
- Scramble siRNA/shRNA เป็นการออกแบบ scramble siRNA negative control สำหรับ siRNA/shRNA ซึ่งจะไม่สามารถยับยั้งการแสดงออกของยีนได้ จากรูปที่ 8.11 ถ้าคลิกที่คำว่า Generate scrambled siRNA ก็จะได้แสดงผลลำดับของ target siRNA และ scrambled siRNA (รูปที่ 8.12)

## Design hairpin insert

[Printer Friendly](#)

Hairpin inserts for **psiRNA-h7SK G1 (cloning sites: BbsI/BbsI)** expression vector

Construct siRNA #0 siRNA GC%: 47.62 Position: 22

Oligo 1

5' **ACCTCGCCCATATCGAACCTAATGATCAAGAGTCATTAGGGTTCGATATGGGCTT** 3'

Oligo 2

5' **CAAAAAGCCCATATCGAACCTAATGACTCTTGATCATTAGGGTTCGATATGGGCG** 3'

## Scramble siRNA

Rescramble the siRNA, if needed

Construct

siRNA target:

5' GCCCATATCGAACCTAATGA 3'

siRNA scrambled:

5' GCACAGATTCAACCGTACTTA 3'

รูปที่ 8.12 ตัวอย่างผลการออกแบบ hairpin insert และ scramble siRNA ด้วยโปรแกรม

InvivoGen: siRNA Wizard

➤ **BLOCK-iT RNAi Designer** [<https://rnaidesigner.thermofisher.com/rnaiexpress/>]

เป็นโปรแกรม web-based สำหรับใช้ในการออกแบบ siRNA และ shRNA ของบริษัท Thermo Fisher Scientific โดยผู้ใช้งานสามารถใส่ accession no. ของ reference sequence หรือลำดับนิวคลีโอไทด์ในรูปแบบ FASTA format ของยีนที่สนใจเข้าสู่โปรแกรม จากนั้นตั้งค่า parameters ต่างๆ ในการออกแบบ (รูปที่ 8.13) แล้วคลิกที่คำว่า RNAi Design โปรแกรมจะทำการออกแบบลำดับของ siRNA พร้อมระบุตำแหน่ง ค่า GC% และ Rank (รูปที่ 8.14)



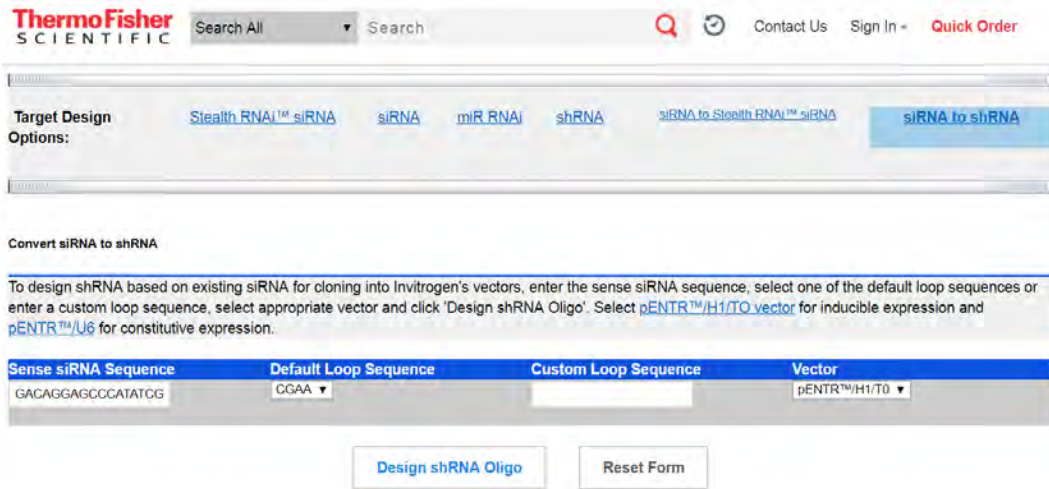
รูปที่ 8.13 BLOCK-iT RNAi Designer [<https://maidesigner.thermofisher.com/maiaexpress>]

10 Stealth RNAi™ siRNA target sequences (Up to 10 top scoring Stealth RNAi™ siRNA sequences are reported, sorted by the Start position and ranked as ★★★★★ to ★★★★★ to indicate knockdown probability). Select the sequence to order and click "Continue".

Select	No.	Start	Sequence(DNA)	Region	GC%	Rank <sup>1</sup>
<input type="checkbox"/>	1	15	GACAGGAGCCCATATCGAACCCCTAA		52.0	★★★★★
<input type="checkbox"/>	2	192	GGCATAATAACAGACACTATCAAGA		36.0	★★★★★
<input type="checkbox"/>	3	257	TGCATGTGTAATGGTTCTTGCTTT		36.0	★★★★★
<input type="checkbox"/>	4	258	GCATGTGTAATGGTTCTTGCTTTA		36.0	★★★★★
<input type="checkbox"/>	5	384	TATTACTATGAGGAATGCTCCTGTT		36.0	★★★★★
<input type="checkbox"/>	6	440	CAGGGATAACTGGCATGGCTCGAAT		52.0	★★★★★
<input type="checkbox"/>	7	460	CGAATCGACCGTGGGTGCTTTCAA		52.0	★★★★★
<input type="checkbox"/>	8	530	CGGAGACAATCCACGCCCTAATGAT		52.0	★★★★★
<input type="checkbox"/>	9	545	CCCTAATGATAAGACAGGCAGTTGT		44.0	★★★★★
<input type="checkbox"/>	10	563	CAGTTGTGGTCCAGTATCGTCTAAT		44.0	★★★★★

รูปที่ 8.14 ตัวอย่างผลการออกแบบ siRNA ด้วยโปรแกรม BLOCK-iT RNAi Designer

นอกจากนี้สามารถเลือกที่ siRNA to shRNA เพื่อสั่งให้โปรแกรมเปลี่ยนลำดับนิวคลีโอไทด์ของ siRNA ไปเป็นลำดับของ oligonucleotides ที่จะใช้ในการสร้าง shRNA expression vector โดยสามารถกำหนดลำดับนิวคลีโอไทด์ของ Loop sequence และเลือกชนิดของ vector ที่ต้องการใช้ได้ (รูปที่ 8.15) จากนั้นคลิกที่คำว่า Design shRNA Oligo ก็จะได้แสดงผลลำดับของ oligonucleotides (รูปที่ 8.16)



รูปที่ 8.15 การออกแบบ hairpin insert จาก siRNA ด้วยโปรแกรม BLOCK-iT RNAi Designer

**shRNA Oligo Features:**

	Linker	Sense Sequence	Loop Sequence	Antisense Sequence
Top Strand	5'	-CACCCGACAGGAGCCCATATCGAACCCCTAACGAA	TAGGGTTCGATATGGGCTCCTGTC	-3'
Bottom Strand	5'	-AAAAGACAGGAGCCCATATCGAACCCCTAATTCGTTAGGGTTCGATATGGGCTCCTGTC	-3'	
ds Oligo	5'	-CACCCGACAGGAGCCCATATCGAACCCCTAACGAA	TAGGGTTCGATATGGGCTCCTGTC	-3'
	3'	-CTGTCCTCGGGTATAGCTTGGGATTGCTTAATCCCAAGCTATACCCGAGGACAGAAAA	-5'	

รูปที่ 8.16 ตัวอย่างผลการออกแบบ hairpin insert ด้วยโปรแกรม BLOCK-iT RNAi Designer

## โปรแกรมการออกแบบ siRNA/shRNA สำหรับเชื้อไวรัส

จากที่นำเสนอไปแล้วข้างต้น จะเห็นได้ว่าโปรแกรมคอมพิวเตอร์สำหรับการออกแบบ siRNA และ shRNA ไม่ว่าจะเป็น DSIR, BIOSETTIA, InvivoGen และ BLOCK-iT RNAi Designer จะเริ่มต้นการออกแบบ siRNA/shRNA จากลำดับนิวคลีโอไทด์เพียงสายเดียว แล้วคำนวณลักษณะ thermodynamics ของ siRNA/shRNA ได้แก่ GC%, Tm และ structure เพื่อให้ได้ siRNA/shRNA ที่เหมาะสม โดยมีได้ค่านึงว่าลำดับนิวคลีโอไทด์ส่วนที่เลือกมาเป็น siRNA/shRNA นั้นอยู่ที่บริเวณใดของยีน และบริเวณดังกล่าวมีการกลายพันธุ์เกิดขึ้นหรือไม่ ทั้งนี้การออกแบบ siRNA/shRNA สำหรับการยับยั้งการแสดงออกของยีนของเชื้อไวรัส ควรมีการคำนึงถึงตำแหน่งของ siRNA/shRNA ด้วย เนื่องจากเชื้อไวรัสบางชนิดมีอัตราการกลายพันธุ์สูงมาก ดังนั้นจึงมีความเป็นไปได้ว่า siRNA/shRNA ที่ใช้จะเกิด mismatch กับสารพันธุกรรมของเชื้อไวรัสเนื่องจากเกิดการกลายพันธุ์ ซึ่งจะส่งผลให้ siRNA/shRNA ดังกล่าวไม่สามารถยับยั้งการแสดงออกของยีนของเชื้อไวรัสได้ ดังนั้นในการออกแบบ siRNA/shRNA สำหรับเชื้อไวรัส จึงนิยมออกแบบให้จำเพาะกับบริเวณอนุรักษ์ (conserved regions) ที่ได้จาก nucleotide sequences alignment เพื่อให้สามารถยับยั้งการแสดงออกของยีนในสายพันธุ์ต่างๆ (subtypes, genotypes, serotypes) ของเชื้อไวรัสชนิดเดียวกัน หรือยับยั้งการแสดงออกของยีนในเชื้อไวรัสชนิดต่างๆ ที่จัดอยู่ในตระกูลเดียวกัน เป็นต้น ในที่นี้จึงขอกกล่าวถึงโปรแกรมที่น่าสนใจ ดังนี้

### ➤ siVirus [<http://sivirus.rnai.jp/>]

เป็นโปรแกรม web-based ที่มุ่งเน้นการออกแบบ siRNA สำหรับเชื้อไวรัส 4 ชนิดที่มีอัตราการกลายพันธุ์สูง ได้แก่ human immunodeficiency virus (HIV), hepatitis C virus (HCV), SARS coronavirus (SARS-CoV) และ Influenza virus (รูปที่ 8.17) โดยผู้ใช้สามารถเลือกสายพันธุ์ต่างๆ และยีนของเชื้อไวรัสที่สนใจ (รูปที่ 8.18) เพื่อให้โปรแกรมออกแบบ siRNA จากส่วนที่เป็น highly conserved target site ระหว่างเชื้อไวรัสสายพันธุ์ต่างๆ แล้วแสดงผล ได้แก่ siRNA target, efficacy prediction, off-target hits และ conservation (รูปที่ 8.19)



Website for efficient antiviral siRNA design.

- Find highly conserved target sites.
- Design highly effective siRNAs.
- Design off-target minimized siRNAs.

<http://siVirus.RNAi.jp/>





**HIV-1**



**HCV**



**SARS**



**HBV**  
Coming soon



**Influenza**

รูปที่ 8.17 โปรแกรม siVirus [<http://sivirus.rnai.jp/>]



**HIV-1**

[siVirus top](#) → **HIV-1** :

- [List by subtype](#)
- [List by sampling country](#)
- [Import list from text](#)



→ Home  
<http://siVirus.RNAi.jp/>

**HIV-1 - check/uncheck all**

Subtype: A1 - check/uncheck all

<input checked="" type="checkbox"/> 1	A1	BY	1997	97BL006	<a href="#">AF193275</a>	<input checked="" type="checkbox"/> 14	A1	KE	2000	MSA4079	<a href="#">AF457086</a>
<input checked="" type="checkbox"/> 2	A1	KE	2000	KER2008	<a href="#">AF457052</a>	<input checked="" type="checkbox"/> 15	A1	KE	2000	NKU3005	<a href="#">AF457089</a>
<input checked="" type="checkbox"/> 3	A1	KE	2000	KER2009	<a href="#">AF457053</a>	<input checked="" type="checkbox"/> 16	A1	KE	1986	ML170_1986	<a href="#">AF539405</a>
<input checked="" type="checkbox"/> 4	A1	KE	2000	KER2012	<a href="#">AF457055</a>	<input checked="" type="checkbox"/> 17	A1	KE	1994	Q23_17	<a href="#">AF004885</a>
<input checked="" type="checkbox"/> 5	A1	KE	2000	KNH1207	<a href="#">AF457068</a>	<input checked="" type="checkbox"/> 18	A1	KE	1999	KNH1088	<a href="#">AF457063</a>
<input checked="" type="checkbox"/> 6	A1	KE	2000	KNH1209	<a href="#">AF457069</a>	<input checked="" type="checkbox"/> 19	A1	KE	1999	KNH1135	<a href="#">AF457065</a>
<input checked="" type="checkbox"/> 7	A1	KE	2000	KNH1211	<a href="#">AF457070</a>	<input checked="" type="checkbox"/> 20	A1	KE	1999	KSM4021	<a href="#">AF457075</a>
<input checked="" type="checkbox"/> 8	A1	KE	2000	KSM4024	<a href="#">AF457077</a>	<input checked="" type="checkbox"/> 21	A1	SE	1994	SE7253	<a href="#">AF069670</a>
<input checked="" type="checkbox"/> 9	A1	KE	2000	KSM4030	<a href="#">AF457079</a>	<input checked="" type="checkbox"/> 22	A1	SE	1994	SE7535	<a href="#">AF069671</a>
<input checked="" type="checkbox"/> 10	A1	KE	2000	MSA4069	<a href="#">AF457080</a>	<input checked="" type="checkbox"/> 23	A1	SE	1995	SE8538	<a href="#">AF069669</a>
<input checked="" type="checkbox"/> 11	A1	KE	2000	MSA4070	<a href="#">AF457081</a>	<input checked="" type="checkbox"/> 24	A1	SE	1995	SE8891	<a href="#">AF069673</a>
<input checked="" type="checkbox"/> 12	A1	KE	2000	MSA4072	<a href="#">AF457083</a>	<input checked="" type="checkbox"/> 25	A1	SE	1995	UGSE8131	<a href="#">AF107771</a>
<input checked="" type="checkbox"/> 13	A1	KE	2000	MSA4076	<a href="#">AF457084</a>	<input checked="" type="checkbox"/> 26	A1	TZ	1997	97TZ02	<a href="#">AF361872</a>

Subtype: A2 - check/uncheck all

<input checked="" type="checkbox"/> 39	A2	CD	1997	97CDKFE4	<a href="#">AF286240</a>	<input checked="" type="checkbox"/> 41	A2	CD	1997	97CDKTB48	<a href="#">AF286238</a>
<input checked="" type="checkbox"/> 40	A2	CD	1997	97CDKS10	<a href="#">AF286241</a>	<input checked="" type="checkbox"/> 42	A2	CY	1994	94CY017_41	<a href="#">AF286237</a>

รูปที่ 8.18 ตัวอย่างการเลือกสายพันธุ์ต่างๆ ของเชื้อไวรัส HIV ในโปรแกรม siVirus





รูปที่ 8.20 ฐานข้อมูล VIRsiRNAdb [http://crdd.osdd.net/servers/virsimadb/search.php]

VIRsiRNAid	Sense Sequence	Virus Name	Target Region	Cell Line	Test Object	Efficacy	PMID	Offtarget #	Align With	ALIGN0 Result
<a href="#">virsi1774</a>	gaaauuggcugcaauuuuu	<a href="#">Enterovirus [EV]</a>	3D Pol	RD	mRNA	100	<a href="#">17712333</a>		<a href="#">refseqs</a>	
<a href="#">virsi1718</a>	ccacaacgucacacaauugu	<a href="#">Human Papillomavirus [HPV]</a>	E7	HeLa	Protein	100	<a href="#">17589817</a>		<a href="#">refseqs</a>	
<a href="#">virsi1292</a>	aaggacaugaccuaccguagac	<a href="#">SARS Coronavirus</a>	RdRp	NIH 3T3	mRNA	100	<a href="#">16757801</a>		<a href="#">refseqs</a>	
<a href="#">virsi1296</a>	aaugucaaccgcuucaauguggc	<a href="#">SARS Coronavirus</a>	RdRp	NIH 3T3	mRNA	100	<a href="#">16757801</a>		<a href="#">refseqs</a>	
<a href="#">virsi1024</a>	cccgcucaaugcccg/cga	<a href="#">Hepatitis C Virus [HCV]</a>	5'UTR	Huh-7	Protein	100	<a href="#">19173711</a>		<a href="#">refseqs</a>	
<a href="#">virsi1025</a>	auuugggcgugcccccga	<a href="#">Hepatitis C Virus [HCV]</a>	5'UTR	Huh-7	Protein	100	<a href="#">19173711</a>		<a href="#">refseqs</a>	
<a href="#">virsi1497</a>	ggcugcggacuguuuggaa	<a href="#">West Nile Virus [WNV]</a>	E	Monkey Kidney Cells	Cell Count	100	<a href="#">16464133</a>		<a href="#">refseqs</a>	
<a href="#">virsi1713</a>	gagguauaugacuauugcuuuu	<a href="#">Human Papillomavirus [HPV]</a>	E6	CaSKI	mRNA	100	<a href="#">19174559</a>		<a href="#">refseqs</a>	
<a href="#">virsi1715</a>	agggaggauagaauagauggu	<a href="#">Human Papillomavirus [HPV]</a>	E7	CaSKI	RNA	100	<a href="#">19174559</a>		<a href="#">refseqs</a>	
<a href="#">virsi1658</a>	gcuugauucuguaugacaaba	<a href="#">Influenza A Virus</a>	NP	MDCK	RNA	100	<a href="#">19479060</a>		<a href="#">refseqs</a>	

รูปที่ 8.21 ตัวอย่างผลการค้นหา siRNA/shRNA จากฐานข้อมูล VIRsiRNAdb

## วิธีการขนส่ง siRNA เข้าสู่เซลล์

เมื่อสังเคราะห์ siRNA หรือสร้าง shRNA expression vector เสร็จแล้ว ในขั้นตอนต่อไปจะต้องคำนึงถึงวิธีการนำ siRNA หรือ shRNA expression vector เข้าสู่เซลล์เป้าหมาย โดยมีหลักการทั่วไปดังนี้

- High transfection efficiency: สามารถนำพา siRNA เข้าสู่เซลล์ได้อย่างมีประสิทธิภาพ
- Low toxicity: ไม่มีความเป็นพิษ หรือมีความเป็นพิษต่อเซลล์น้อย
- Low immunogenicity: ไม่กระตุ้นระบบภูมิคุ้มกัน หรือกระตุ้นเพียงเล็กน้อย
- Sustained silencing effect: ทำให้เกิด gene silencing ได้นาน
- Cells type specific: มีความจำเพาะต่อชนิดของเซลล์ที่สนใจ ซึ่งเป็นคุณสมบัติที่สำคัญในการใช้ในสิ่งมีชีวิต (*in vivo*) หรือการประยุกต์ใช้ทางคลินิก (clinical application)

## ประเภทของพาหะสำหรับขนส่ง siRNA

โดยทั่วไปสามารถแบ่งพาหะ (Vehicles) สำหรับนำพา siRNA ได้เป็น 2 ประเภท ดังนี้

- Viral vectors: อาศัยลักษณะโครงสร้างของ virus particle ในการนำพา siRNA เข้าสู่เซลล์ โดยสามารถกำหนดความจำเพาะต่อเซลล์แต่ละชนิดได้ โดยการดัดแปลง surface หรือ envelop protein ของไวรัส ให้จำเพาะกับ receptor บนผิวเซลล์แต่ละชนิด ซึ่งสามารถแบ่งได้เป็น 3 ชนิด ได้แก่ adenovirus, adeno-associated virus และ lentiviruses
- Non-viral vectors: อาศัยคุณสมบัติของโมเลกุลที่ลักษณะคล้ายไขมัน (lipid-based vehicles) ในการจับกับ siRNA และนำเข้าสู่เซลล์ได้ เนื่องจากเยื่อหุ้มเซลล์มีลักษณะคล้ายไขมัน (hydrophobic) เช่นกัน ซึ่งสามารถแบ่งได้เป็น 4 ชนิด ได้แก่ lipid-raft compartments, liposomes, lipidoids และ lipoplexes



## การประยุกต์ใช้เทคนิค RNAi ในการยับยั้งการเพิ่มจำนวนของเชื้อไวรัส

เนื่องจากเทคนิค RNAi สามารถยับยั้งการแสดงออกของยีนโดยกระตุ้นให้เกิดการทำลาย mRNA (mRNA degradation) หรือยับยั้งการสร้างโปรตีน (translational repression) อย่างจำเพาะ ดังนั้นจึงมีการประยุกต์ใช้เทคนิค RNAi ในการยับยั้งการเพิ่มจำนวนของเชื้อไวรัสภายในเซลล์ อย่างไรก็ตามประสิทธิภาพในการยับยั้งการเพิ่มจำนวนของเชื้อไวรัส นอกจากจะขึ้นกับการออกแบบ siRNA / shRNA ที่มีประสิทธิภาพแล้ว ยังขึ้นอยู่กับลักษณะทางชีววิทยาของเชื้อไวรัส แต่ละชนิด ดังนี้

- ลักษณะสารพันธุกรรมของเชื้อไวรัส เนื่องจากสารพันธุกรรมของเชื้อไวรัส มีหลายรูปแบบ โดยจำแนกได้เป็น 7 ประเภท ได้แก่ double strand DNA viruses, single strand DNA viruses, double strand RNA viruses, positive single strand RNA viruses, negative single strand RNA viruses, RNA reverse transcribing viruses (retroviruses) และ DNA reverse transcribing viruses ทั้งนี้กลไก RNAi จะสามารถทำลาย mRNA และ viral RNA ที่เป็นสารพันธุกรรมของ RNA virus ได้ ในขณะที่ RNAi จะสามารถยับยั้งการแสดงออกของยีน (mRNA) แต่ไม่สามารถทำลาย viral DNA ที่เป็นสารพันธุกรรมของ DNA virus ได้ ดังนั้นกลไก RNAi จึงมีประสิทธิภาพสูงในการยับยั้งการเพิ่มจำนวนของเชื้อไวรัสที่มีสารพันธุกรรมเป็น RNA ได้ดีกว่าเชื้อไวรัสที่มีสารพันธุกรรมเป็น DNA
- ตำแหน่งในการเพิ่มจำนวนของเชื้อไวรัสภายในเซลล์ เนื่องจากเชื้อไวรัสบางชนิดเพิ่มจำนวนภายในนิวเคลียส และเชื้อไวรัสบางชนิดเพิ่มจำนวนในไซโตพลาสซึม แต่กลไก RNAi จะทำงานในส่วนไซโตพลาสซึมของเซลล์ ดังนั้นกลไก RNAi จึงมีประสิทธิภาพในการยับยั้งเชื้อไวรัสที่มีการเพิ่มจำนวนในไซโตพลาสซึมได้ดีกว่า เชื้อไวรัสที่มีการเพิ่มจำนวนภายในนิวเคลียส

ตัวอย่างการพัฒนา siRNA เพื่อใช้เป็นยาด้านไวรัสชนิดต่างๆ ได้แก่ human immunodeficiency virus (HIV), respiratory syncytial virus (RSV), hepatitis C virus (HCV) และ hepatitis B virus (HBV) เป็นต้น

### เอกสารอ้างอิง และแนะนำอ่านเพิ่มเติม

1. Amarzguioui M, Prydz H. An algorithm for selection of functional siRNA sequences. *Biochem Biophys Res Commun.* 2004 Apr 16;316(4):1050-8.
2. Dorsett Y, Tuschl T. siRNAs: applications in functional genomics and potential as therapeutics. *Nat Rev Drug Discov.* 2004 Apr;3(4):318-29. Review.
3. Echeverri CJ, Perrimon N. High-throughput RNAi screening in cultured cells: a user's guide. *Nat Rev Genet.* 2006 May;7(5):373-84.
4. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature.* 1998 Feb 19;391(6669):806-11.
5. Haasnoot J, Westerhout EM, Berkhout B. RNA interference against viruses: strike and counterstrike. *Nat Biotechnol.* 2007 Dec;25(12):1435-43. Review.
6. Ivacic D, Ely A, Arbuthnot P. Countering hepatitis B virus infection using RNAi: how far are we from the clinic? *Rev Med Virol.* 2011 Nov;21(6):383-96.
7. Lares MR, Rossi JJ, Ouellet DL. RNAi and small interfering RNAs in human disease therapeutic applications. *Trends Biotechnol.* 2010;28(11):570-9.
8. Matveeva OV, Kang Y, Spiridonov AN, et al. Optimization of duplex stability and terminal asymmetry for shRNA design. *PLoS ONE.* 2010;5(4):e10180.
9. Mittal V. Improving the efficiency of RNA interference in mammals. *Nat Rev Genet.* 2004 May;5(5):355-65. Review.
10. Naito Y, Ui-Tei K, Nishikawa T, Takebe Y, Saigo K. siVirus: web-based antiviral siRNA design software for highly divergent viral sequences. *Nucleic Acids Res.* 2006 Jul 1;34(Web Server issue):W448-50.

11. Panjaworayan N, Payungporn S, Poovorawan Y, Brown CM. Identification of an effective siRNA target site and functional regulatory elements, within the hepatitis B virus posttranscriptional regulatory element. *Virology*. 2010 Sep 8;7:216.
12. Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorov A. Rational siRNA design for RNA interference. *Nat Biotechnol*. 2004 Mar;22(3):326-30.
13. Thakur N, Qureshi A, Kumar M. VIRsiRNAdb: a curated database of experimentally validated viral siRNA/shRNA. *Nucleic Acids Res*. 2012 Jan;40(Database issue):D230-6.
14. Thongthae N, Payungporn S, Poovorawan Y, T-Thienprasert NP. A rational study for identification of highly effective siRNAs against hepatitis B virus. *Exp Mol Pathol*. 2014 Aug;97(1):120-7.
15. Ui-Tei K, Naito Y, Takahashi F, Haraguchi T, Ohki-Hamazaki H, Juni A, Ueda R, Saigo K. Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res*. 2004 Feb 9;32(3):936-48.
16. Vert JP, Foveau N, Lajaunie C, Vandenbrouck Y. An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics*. 2006 Nov 30;7(1):520.



ดัชนี

Index













เว็บไซต์ โปรแกรม และฐานข้อมูลสาธารณะที่เกี่ยวข้อง

Related public websites, programs and databases

## ดัชนี

## Index

คลิกที่ชื่อ หรือสแกน QR code เพื่อไปสู่เว็บไซต์ โปรแกรม หรือฐานข้อมูลสาธารณะที่สนใจ

BioEdit	BioSettia	BLAST	BLOCKit
			
bNAber	DSIR	euHCVdb	HBVdb
			
HBVRegDB	HepSEQ	HPV T cell Antigen Database	ICTV
			
IEDB	IRD	IVDB	LANL HCV Database
			

## ดัชนี (ต่อ)

## Index (Cont.)

คลิกที่ชื่อ หรือสแกน QR code เพื่อไปสู่เว็บไซต์ โปรแกรม หรือฐานข้อมูลสาธารณะที่สนใจ
















LANL HIV database	MEGA	NCBI	NCBI-Genome
			
NCBI-Nucleotide	NCBI-OMIM	NCBI-Protein	NCBI-PubMed
			
NCBI-Structure	NCBI-Taxonomy	NCBI-VWR	Oligo Calc
			
Oligo Analyzer	OpenFluDB	PaVE	PCR Primer Stats
			



## ดัชนี (ต่อ)

## Index (Cont.)

คลิกที่ชื่อ หรือสแกน QR code เพื่อไปสู่เว็บไซต์ โปรแกรม หรือฐานข้อมูลสาธารณะที่สนใจ

Primer3Plus	Primer3web	PrimerBank	Primer-BLAST
			
PrimerQuest Tool	RNAi central-siRNA	SeqHepB	siRNAwizard
			
siVirus	Stanford HIV Drug Resistance DB	VIPERdb	ViPR
			
ViralZone	VirHostnet	VirOligo	VIRsiRNadb
			

ประวัติผู้พิมพ์

Biography

## ประวัติผู้นิพนธ์ (Biography)

ผู้นิพนธ์ รองศาสตราจารย์ ดร.สัชชัย พยุงพร

Associate Professor Sunchai Payungporn, Ph.D.



### คุณวุฒิ

- พร.ต. (ชีวเวชศาสตร์) จุฬาลงกรณ์มหาวิทยาลัย พ.ศ. 2550
- วท.บ. (ชีวเคมี เกียรตินิยมอันดับ 1) จุฬาลงกรณ์มหาวิทยาลัย พ.ศ. 2545

สังกัดหน่วยงาน ภาควิชาชีวเคมี คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

E-mail address: sp.medbiochemcu@gmail.com

**ความเชี่ยวชาญพิเศษ** อดุชีววิทยาของเชื้อไวรัสและเทคนิคทางด้านอณูชีววิทยาในการศึกษา  
ลักษณะทางพันธุกรรม จำแนกสายพันธุ์ วิเคราะห์การกลายพันธุ์ ศึกษาความสัมพันธ์เชิง  
วิวัฒนาการ และพัฒนาเทคนิคในการตรวจวินิจฉัยเชื้อไวรัสอุบัติใหม่และอุบัติซ้ำในประเทศไทย

### ผลงานวิจัยระดับนานาชาติ

มีผลงานวิจัยระดับนานาชาติมากกว่า 90 เรื่องในฐานข้อมูล Scopus (h-index =23)



### ประวัติการได้รับรางวัลและการประกาศเกียรติคุณ

รางวัลและการประกาศเกียรติคุณ	ปี พ.ศ.	จากหน่วยงาน
รางวัลนักวิจัยรุ่นเยาว์ ประจำปี 2555	2556	กองทุนรัชดาภิเษกสมโภช จุฬาลงกรณ์มหาวิทยาลัย
Award of Completion: BGI International Bioinformatics Workshop (2013)	2556	Beijing Genomics Institute (BGI), China
ทุนส่งเสริมบัณฑิต	2555	มูลนิธิ "อานันทมหิดล" แผนกแพทยศาสตร์
รางวัลเพื่อส่งเสริมความเป็นเลิศในการปฏิบัติงาน ด้านการเรียนการสอนสำหรับอาจารย์รุ่นใหม่ ประจำปี 2554	2555	กองทุนกาญจนาภิเษกเฉลิมพระเกียรติ จุฬาลงกรณ์มหาวิทยาลัย
Award of Completion: 2010 Bioinformatics and Comparative Genome Analysis Course	2553	Institut Pasteur, France
Award of Completion: Biosafety Level 3 Safety Training	2551	The Emory University Onsite BSL3 Program
รางวัลทะกุจิ ประเภทวิทยานิพนธ์ดีเด่น ระดับปริญญาเอก	2550	สมาคมเทคโนโลยีชีวภาพแห่งประเทศไทย
รางวัลการเสนองานแบบปากเปล่าดีเด่น ระดับปริญญาเอก	2550	คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
รางวัลการเสนองานแบบโปสเตอร์ ดีเด่น	2547	RGJ-Ph.D. Congress ครั้งที่ 5 (สกว.)
รางวัล Senior project ดีเด่น	2545	ภาควิชาชีวเคมี คณะวิทยาศาสตร์ จุฬาฯ

หนังสือเล่มนี้จัดพิมพ์เพื่อประโยชน์ทางการศึกษา  
ห้ามจำหน่ายเชิงพาณิชย์

Free E-book is available at >>  
ดาวน์โหลดหนังสืออิเล็กทรอนิกส์ฟรี >>



หนังสือเล่มนี้เหมาะสำหรับ		
ผู้เริ่มต้น	ระดับกลาง	ระดับสูง
	▲	▲
สาขา วิทยาศาสตร์ชีวภาพ/จุลชีววิทยา/ไวรัสวิทยา		

ISBN 978-616-468-907-7



9 786164 689077