

บทปฏิบัติการที่ 9

การหาลำดับเบสและการวิเคราะห์โดยคอมพิวเตอร์

หลักการ

ในปัจจุบันความก้าวหน้าทางด้านวิทยาการต่างๆ มีการพัฒนาขึ้นมาก ทำให้ความต้องการต่อการสืบค้นหาและการวิเคราะห์ข้อมูลต่างๆ มีความจำเป็นมากยิ่งขึ้น ดังนั้น คอมพิวเตอร์จึงเข้ามายึดบทบาทสำคัญอย่างมากในการที่จะช่วยค้นหาข้อมูล ซึ่งสารสนเทศศาสตร์ (Bioinformatics) จึงเป็นศาสตร์ในสาขาใหม่ที่ถูกพัฒนาขึ้นมา โดยเป็นศาสตร์ที่เกี่ยวข้องกับการใช้คอมพิวเตอร์ในการสืบค้นหาและการวิเคราะห์ข้อมูลทางวิทยาศาสตร์ชีวภาพที่ได้มาจากการวิจัยและค้นคว้า ซึ่งในปัจจุบันสามารถทำได้ง่าย โดยการสืบค้นผ่านเครือข่ายอินเตอร์เน็ต (world wide web) ซึ่งสารสนเทศจึงเป็นศาสตร์ที่ต้องใช้ความรู้ในหลายสาขาวิชามาประกอบกัน เช่น สหศิลป์ อนุชีววิทยา โปรแกรมสำเร็จรูปในการวิเคราะห์ข้อมูล เป็นต้น การประยุกต์ใช้ศาสตร์นี้จะเป็นประโยชน์ต่อการนำไปสู่ความเข้าใจในหลักการเบื้องต้นของการทางชีววิทยาของสิ่งมีชีวิต

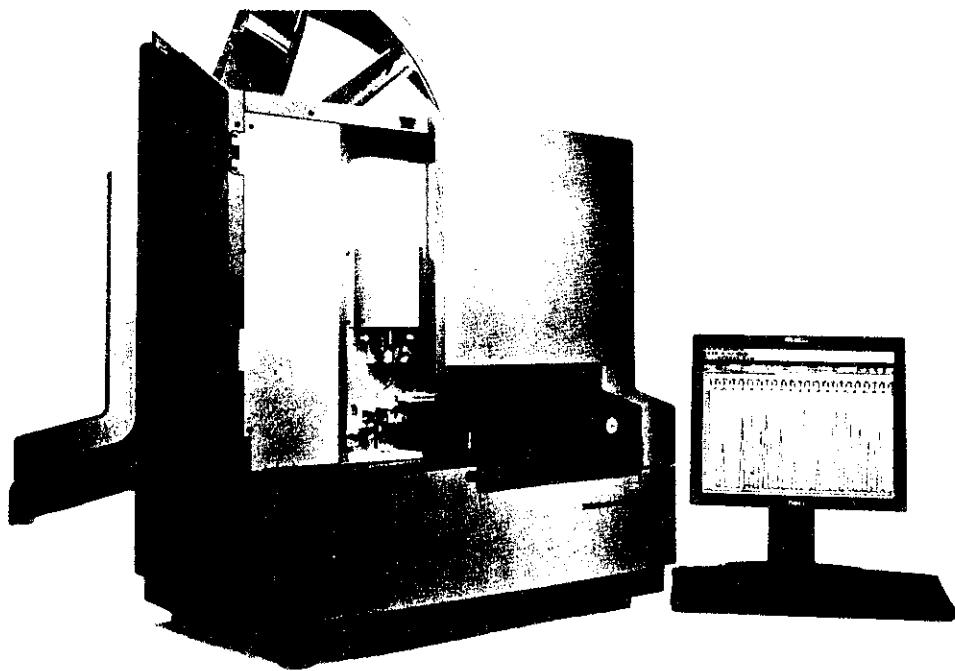
วัตถุประสงค์

1. เพื่อให้เข้าใจการค้นหาข้อมูลทางอณูชีววิทยาจากฐานข้อมูลบนเครือข่ายอินเตอร์เน็ต

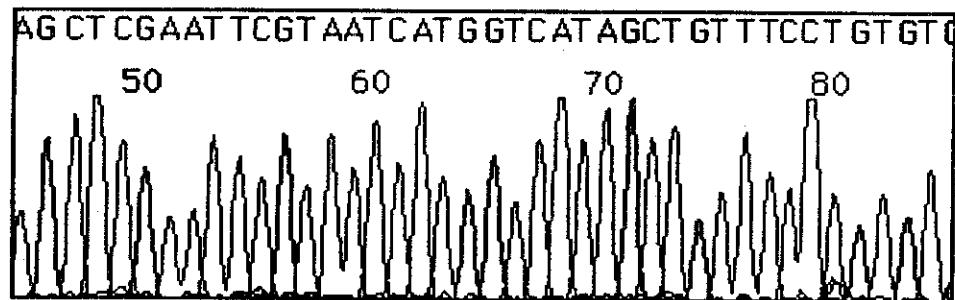
2. เพื่อศึกษาการหาลำดับเบสและการวิเคราะห์โดยคอมพิวเตอร์

ข้อมูลและอุปกรณ์

- ตัวอย่างข้อมูลนิวคลีโอไทด์ จำนวน 2 ข้อมูล (sequence1.txt และ sequence2.txt)
- โจทย์ข้อมูลนิวคลีโอไทด์ จำนวน 1 ข้อมูล (unknown 1-15.txt)
- เครื่องคอมพิวเตอร์ที่มีระบบปฏิบัติการ Windows และสามารถติดต่อกับเครือข่ายอินเตอร์เน็ต



รูปที่ 9-1 เครื่องหาลำดับเบสอัตโนมัติ (Autamatic DNA Sequencer;
PE Applied Biosystems)



รูปที่ 9-2 ตัวอย่างการอ่านลำดับเบสจากเครื่องหาลำดับเบสแบบอัตโนมัติ
(จาก PE Applied Biosystems)

วิธีการทดลอง

1. การอ่านลำดับเบสจากเครื่องอัตโนมัติและการเก็บข้อมูลลำดับเบสในรูปแบบ FASTA
 1. อ่านลำดับเบสจากข้อมูลจริงที่ได้จากเครื่องอัตโนมัติ
 2. เปิดโปรแกรม microsoft word หรือ notepad และสร้างไฟล์ใหม่ด้วยคำสั่ง "File-New"
 3. พิมพ์เครื่องหมาย ">" ตามตัวย่อหรือตั้งชื่อข้อมูลลำดับเบสนั้นๆ ที่บรรทัดแรกของไฟล์ เช่น >a1 ,>a2 เป็นต้น
 4. พิมพ์ลำดับเบสที่ได้จากการอ่าน หรือใช้คำสั่ง copy ลำดับเบส และ past ลงในบรรทัดถัดมาไม่ต้องเว้นวรรค ได้ผลดังแสดงในรูปที่ 1

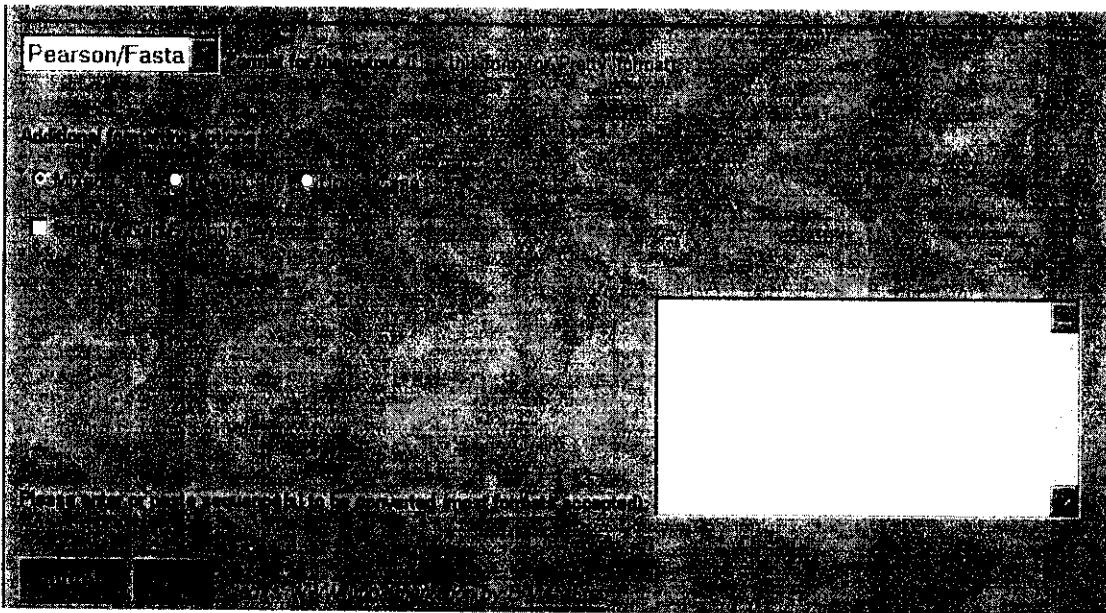
```
>a1
1 gtttactcta gtgcgacgggt gagtacacgt gagatggcct tagagggat acaatttagaa
61 tgaatgtcaa tacccatatg ctttgagtga aatggatTA tatttatctg cctaaagaga
121 agctcgccgc tgatttagttt gttggtaagg taagagctt ccaaggcgac gatcgtatc
181 tggtttgaga ggacgatcaG acacacttggA acttagacac ggtccagact cttacgggag
241 gcagcagtgg ggaattttcc gcaatggggg aaagcgttgc ggagcaatac cgcgtgaggg
301 aagactgcct atgggttgta aacctttt ttcagggagg aataaaaatgt a cgtgtacccgt
361 aagaataaa

>a2
1 cgggaagcgg ggcnccttcg gaccctcngct attggatgag cctngtcgga ttagctagg
61 ggtggggtaa tggcctacca aggccnnngat ccgtatctgg tctgagaggg tgatcagcna
121 cactgggact gagacacggc ccagactct acgggaggca gcagtgggaa atattgegca
181 atgggggaaa ccctgacgca gccatggcgc gtgtgtgaag aaggccttag ggttgtaaag
241 cactttcaat gggcacaaaa cgcctgttagt taatagctgc aggacttgat gttacctata
301 gaagaag

>a3
1 cttcggtcan naagggannn cggcccatta tttaatccta gttggtaagg taacggccta
61 ccaagggtac gatgggtacg tggctctgaga ggtatgtacg ccacactggg actgagacac
121 ggcccagact cctacggag gcagcagtgg ggaatattgc gcaatggacg aaagtctgac
181 gcagccacgc cgcgtgagtg atgaaggccct tagggttgt aagctctgtc gggagggnccg
241 aagtagttt actacttgac ggtacctcgC aagaaaa
```

รูปที่ 1 โดยใช้โปรแกรม notepad

5. เปิด web server ที่มีการให้บริการใช้โปรแกรม FASTA ในที่นี้เข้า web READSEQ Sequence Conversion ได้ผลดังรูปที่ 2



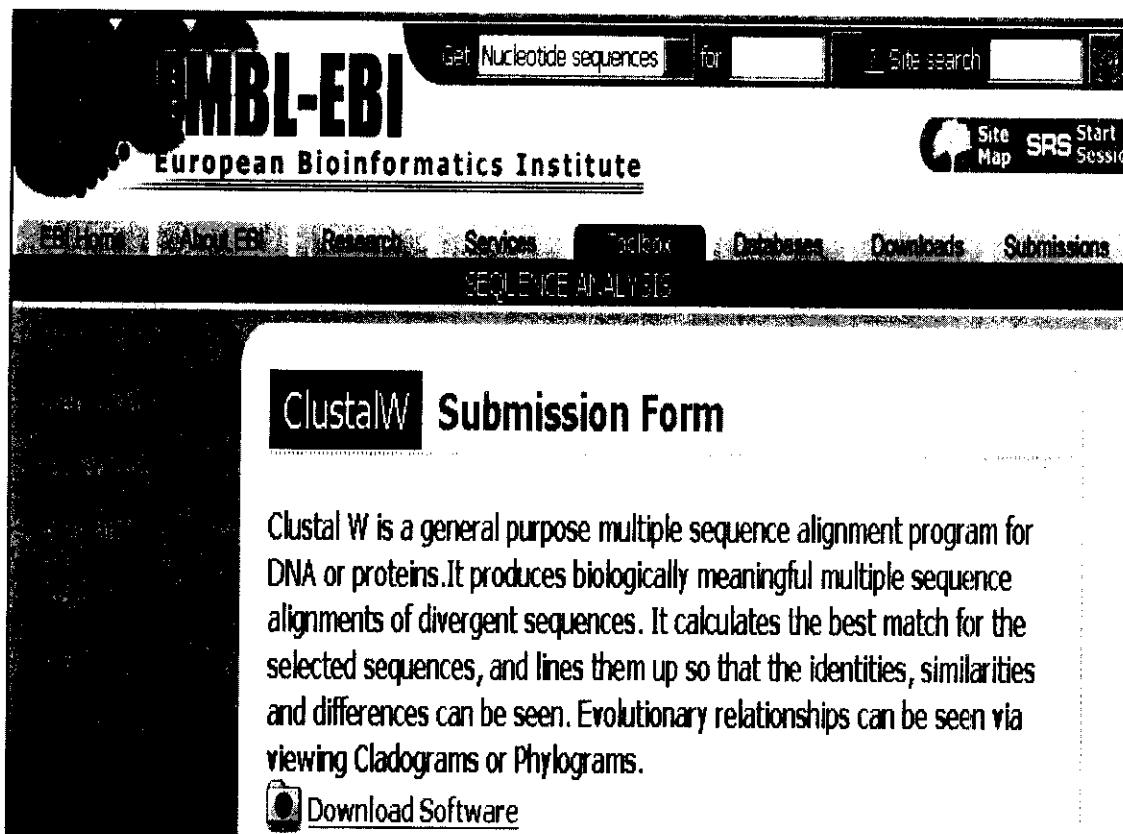
รูปที่ 2 web READSEQ Sequence Conversion

6. ใช้คำสั่ง copy สำลับเบสทั้งหมดใน notepad และ paste ลงในช่องว่างใน web READSEQ Sequence Conversion และกด submit ได้ผลดังรูปที่ 3

```
>a1, 307 bases, 1E66 checksum.  
cgggaagccgggcncctcgccctngctatggatgacccctngtcgg  
ttatctatgttgtgggttaatggcttaccaggcgcnngatccgtatcg  
tctgegeggatgtatcagccatctggactgtggacscggccccatctct  
acggggggcagcgtggaaatitgcgcataatggggaaaccttgcacga  
gcctatggccgttgtlgaegaesggctttaggttgtaaaggcacitcaat  
ggggccaaaaacgccttgtatgttgcggatcttgtatgtttaccatia  
gaagaag  
>a2, 276 bases, 19AD checksum.  
cttcggccatccatggggnnnnnccatattatccatccatgttgcagg  
taacggcttaccaegggtatgtgggtatgttgcgttgtatgtatgtatgt  
cccatgtggactgtatgtatgtatgtatgtatgtatgtatgtatgtatgt  
ggatgttgcgtatgtatgtatgtatgtatgtatgtatgtatgtatgt  
atgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt  
atgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt  
atgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt  
>a3, 311 bases, 16A0 checksum.  
aaatctcgccaaatggggatctggccctctngtcaatgtatgtatgtatgt
```

รูปที่ 3

7. หลังจาก format ข้อมูลแล้ว สามารถนำมาใช้กับโปรแกรมที่เกี่ยวกับการหาลำดับเบสที่ให้บริการอยู่ทั่วไปได้ รวมไปถึงการใช้ในการออกแบบ primers เช่น clustalW ดังตัวอย่างในรูปที่ 4

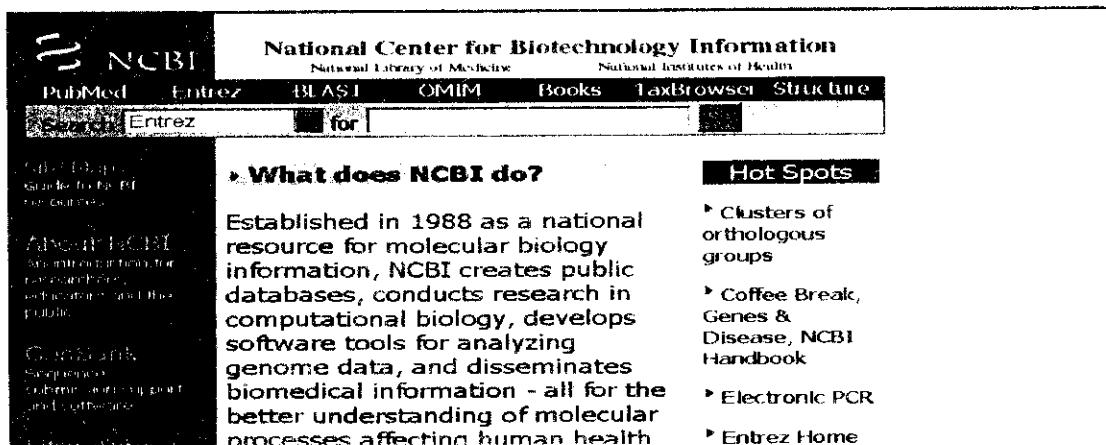


รูปที่ 4 clustalW

II. การวิเคราะห์ความเหมือน (Homology Search) ของข้อมูลนิวคลีโอไทด์ด้วยโปรแกรมคอมพิวเตอร์ที่มีบริการบนเครือข่ายอินเตอร์เน็ต

การใช้ Blastn โดยใช้บริการจาก NCBI

1. เมื่อโปรแกรม Netscape หรือ Internet Explorer โดยการทำ double click บนไอคอน แล้วเข้าสู่ระบบอินเตอร์เน็ตของ web server ที่ <http://www.ncbi.nlm.nih.gov/> ซึ่งเป็นแหล่งเก็บข้อมูลด้านอนุชีววิทยาจำนวนมาก เรียกอีกชื่อว่า GenBank ได้ผลลัพธ์ที่ 1



รูปที่ 1

2. กดตรงปุ่ม “BLAST” จะพบ webpage แสดงกลุ่มโปรแกรม BLAST ดังแสดงในรูปที่ 2

รูปที่ 2

3. จากนั้นเลือก nucleotide (nr) และ blastn จากนั้นกำหนดค่าตามความเหมาะสม ดังแสดงในรูปที่ 3,4,5, และ 6 ตามลำดับ

รูปที่ 3

Options for advanced blasting

Limit by or select from: All organisms

Choose filter Low complexity Mask for lookup table only Mask lower case

Expect

Word Size

Matrix Gap Costs Existence: 11 Extension: 1

Other advanced

รูปที่ 4

Format

Show Graphical Overview Linkout Sequence Retrieval NCBI-gi Alignment in

Use new formatter Masking Character Default(X for protein, n for nucleotide) Masking Color Black

Number of: Descriptions Alignments

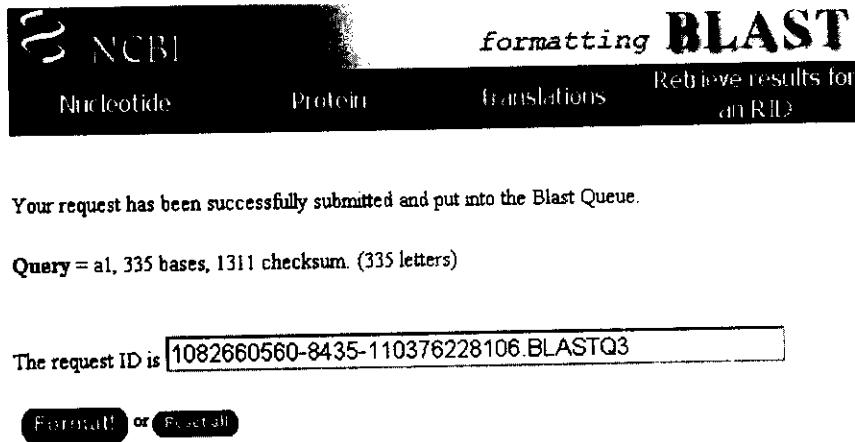
Alignment view

Limit results by or select from: All organisms

Expect value

รูปที่ 5

4. กด BLAST เพื่อเริ่ม run โปรแกรม รอสักครู่จะได้รับคำตอบจากโปรแกรม ดังแสดงในรูปที่ 6

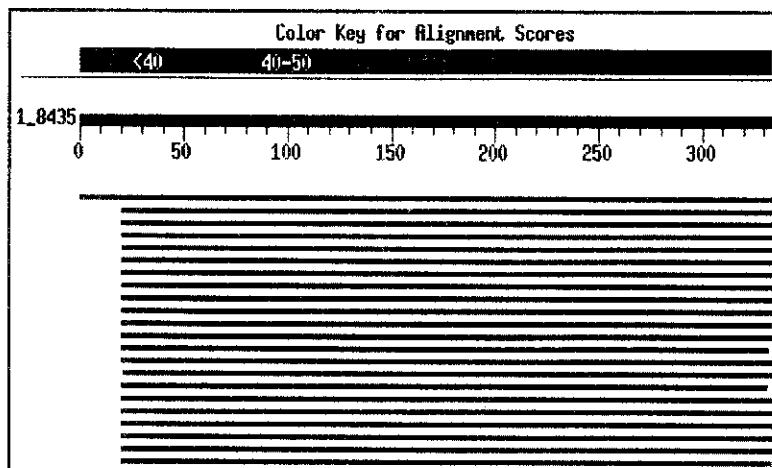


รูปที่ 6

5. เมื่อโปรแกรม run เสร็จแล้ว จะแสดงออกมานิรูปลักษณะที่แสดงในรูปที่ 7, 8 และ 9 ตามลำดับ โดยจะออกแบบอยู่ในรูปที่สามารถกดปุ่มเลือกเข้าไปดูในแต่ละส่วนที่สนใจจะศึกษาได้โดยสะดวกและเข้าใจง่าย

Distribution of 100 Blast Hits on the Query Sequence

Mouse-over to show defline and scores. Click to show alignments



§U¥1 7

Score	E
Sequences producing significant alignments:	
g 1465655 emb Z77638.1 BSPZ77638	Bacterium 2-400 C2.6 part... 646 0.0
g 28208627 gb AF382135.1	Uncultured bacterium clone ZA361... 591 e-166
g 28208626 gb AF382133.1	Uncultured bacterium clone ZA361... 591 e-166
g 1465643 emb Z77626.1 BSPZ77626	Bacterium 2-400 C2.16 par... 591 e-166
g 28208631 gb AF382139.1	Uncultured bacterium clone ZA363... 567 e-169
g 21743620 gb AF464307.1	Uncultured actinobacterium clone... 535 e-149
g 21743618 gb AF464305.1	Uncultured actinobacterium clone... 535 e-149
g 21743617 gb AF464304.1	Uncultured actinobacterium clone... 535 e-149
g 16326786 gb AF428625.1	Uncultured bacterium clone CRD99... 527 e-147
g 16326851 gb AF428690.1	Uncultured bacterium clone CRP99... 527 e-147
g 16327061 gb AF428690.1	Uncultured bacterium clone CR98... 525 e-146

§U¥1 8

>g1465855|emb|277638.1|BSP277638 Bacterium 2-400 C2 6 partial 16S rRNA gene
Length = 336

Score = 646 bits (326), Expect = 0.0
Identities = 336/336 (100%)
Strand = Plus / Plus

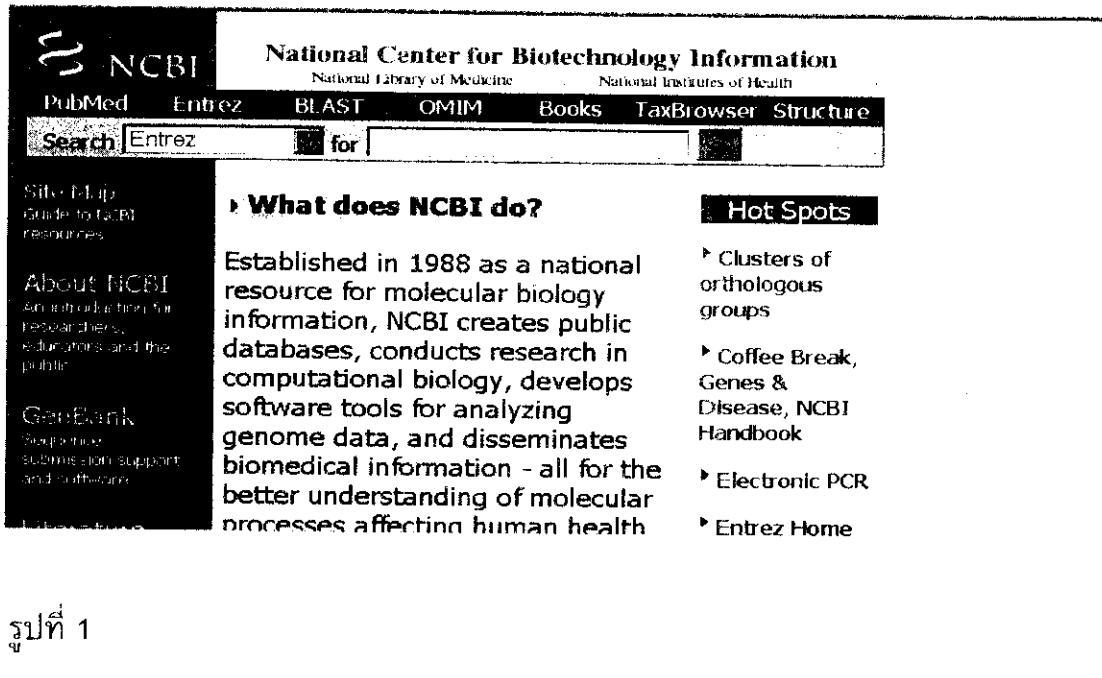
Query: 1 tgcttcggcttacacatgtgcattggctgtcgatcgatgtcgatgtggatgtgggt 60
[REDACTED]
Subject: 1 tgcttcggcttacacatgtgcattggctgtcgatcgatgtcgatgtggatgtgggt 60

Query: 61 taatcccgaaacggggcggcaaccctttatccatgtgccatcgatgtggggactc 120
[REDACTED]

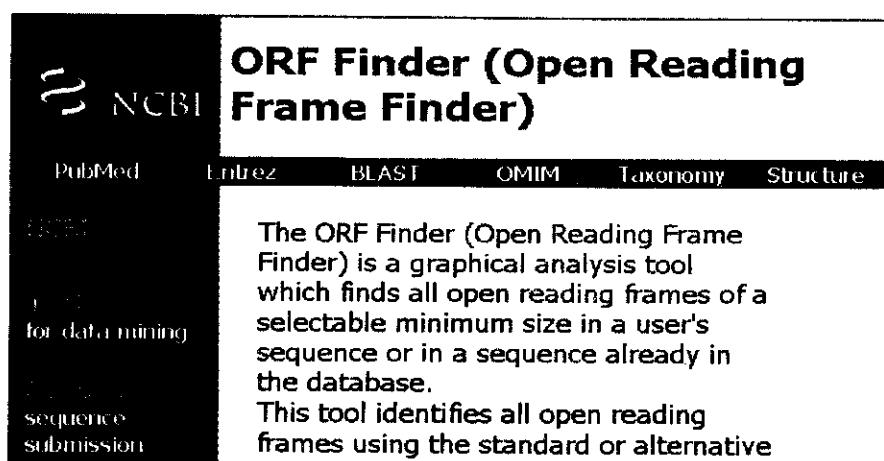
รูปที่ 9

III. การวิเคราะห์หาตำแหน่งที่อาจเป็นยีน (Open Reading Frame หรือ ORF) ของข้อมูลนิวคลีโอไทด์ด้วยโปรแกรมคอมพิวเตอร์ที่มีบริการบนเครือข่ายอินเตอร์เน็ต

1. เปิดโปรแกรม Netscape หรือ Internet Explorer ด้วยการทำ double click บนไอคอน
2. เปิดดู <http://www.ncbi.nlm.nih.gov/> (รูปที่ 8)
3. กดตรงปุ่ม “ORF Finder” ซึ่งอยู่ที่ด้านขวาเมื่อของหน้า web ได้ผลตั้งรูปที่ 1 และ 2



รูปที่ 1



รูปที่ 2

4. คัดลอกข้อมูลไฟล์จากแฟ้มดิสก์ที่ทำการบันทึกข้อมูลไว้แล้วแลกด้วยโปรแกรม Microsoft Word และเติมลงในช่องว่างของ webpage ดังรูปที่ 3 (สังเกตว่า ต้องให้ข้อมูลอยู่ในรูปแบบ FASTA)

Enter GI or ACCESSION

or sequence in FASTA format

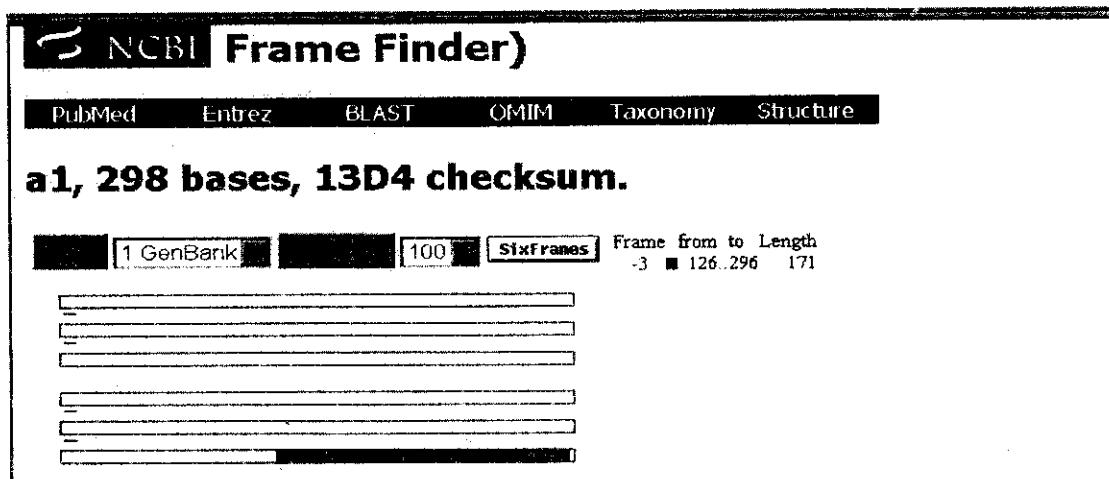
FROM: [] **TO:** []

Genetic codes

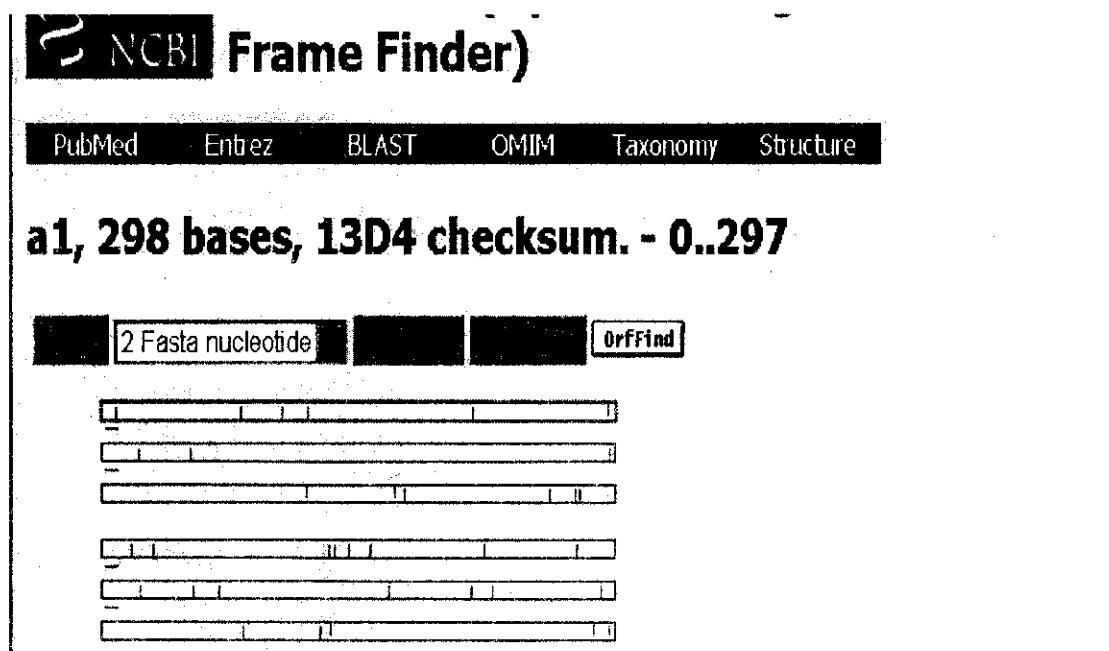
1 Standard

รูปที่ 3

5. คัดลอกข้อมูลไฟล์จากแฟ้มดิสก์ที่ทำการบันทึกข้อมูลไว้แต่แรกด้วยโปรแกรม Microsoft Word และเดิมลงในช่องว่างของ webpage ดังรูปที่ 15 (สังเกตว่า ต้องให้ข้อมูลอยู่ในรูปแบบ FASTA)
6. กดปุ่ม “ OffFind ” เพื่อทำการวิเคราะห์หาตำแหน่งซึ่งอาจเป็นยีนบนสายดีเอ็นเอ
7. รอสักครู่จะได้รับคำตอบรับจากโปรแกรม ดังแสดงในรูปที่ 4 และ 5 แสดงตำแหน่งซึ่งอาจเป็นยีนบนสายดีเอ็นเอต้นแบบ โดยพิจารณาจากทั้ง 6 เฟรม
8. สามารถกดปุ่ม viewall เพื่อดูเส้นดีเอ็นเอที่สนใจได้ ดังแสดงในรูปที่ 6
9. ถ้าต้องการรายละเอียดของลำดับกรดอะมิโนจากเฟรมใด ให้ double click บนตำแหน่งนั้นๆ จะปรากฏข้อมูลอยู่ด้านล่าง ดังแสดงในรูปที่ 7 และ 8



รูปที่ 4



รูปที่ 5

NCBI ORF Finder (Open Reading Frame Finder)

PubMed Entrez BLAST OMIM Taxonomy Structure

a1, 298 bases, 13D4 checksum. - 0..297

Program: blastp Database: nr BLAST with parameters Cognitor

2 Fasta nucleotide OrffFind

ญี่ปุ่นที่ 6

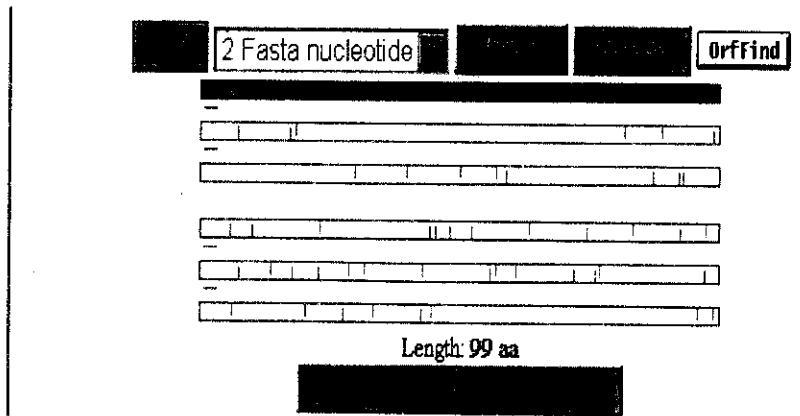
Length: 99 aa

```

1 ctcgtgcctgagggtgtcaggtaactctatacggcgaaaccttggtagttgcc
L V P * G V R L S P I R A Q P L L L V A
61 agcgattcaagtccggaaactctaacaagactgcccagtgcggaaactgtggaggatgggg
S D S S R E L * Q D C Q C K L * G R W G
121 tgacgtcaaatcatcacggcccttacgttctggctacacacgtgtacaaatggataga
* R Q I I T A L T S W A T H V L Q W Y R
181 cagagagcagccacttcgcgagaaggagcgaatctataaactatatacagttcgatcg
Q R A A T S R E G A N L * T I S Q F G S
241 gagttctgcactcgactccgtgaagctggaaatcgtagtaatcgatcatcagccatya 298
E S A T R L R E A G I A S N R I S A M

```

ญี่ปุ่นที่ 7



```

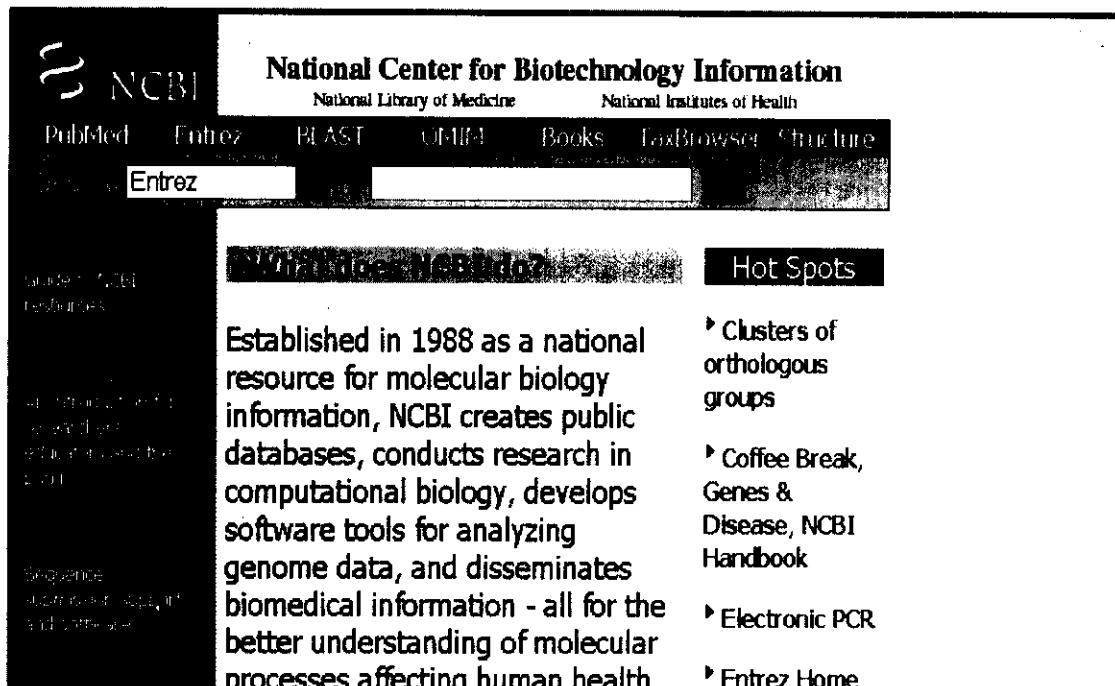
1 ctcgtgccgtgagggtcaggttaagtctatacgagcgcaacccctgtttagttgcc
L V P * G V R L S P I R A Q P L L L V A
61 agcgattcaagtccggaaactctaacaagactgccagtgcacaaactgtgaggaagggtgggaa
S D S S R E L * Q D C Q C K L * G R W G
121 tgacgtcaaatcatcacggccctteacgtctggctacacacgtgctacaatggtataga
* R Q I I T A L T S W A T H V L Q W Y R
181 caaaaaaaacccacttccacaaatctataaaactataatcacacttccaaatca

```

รูปที่ 8

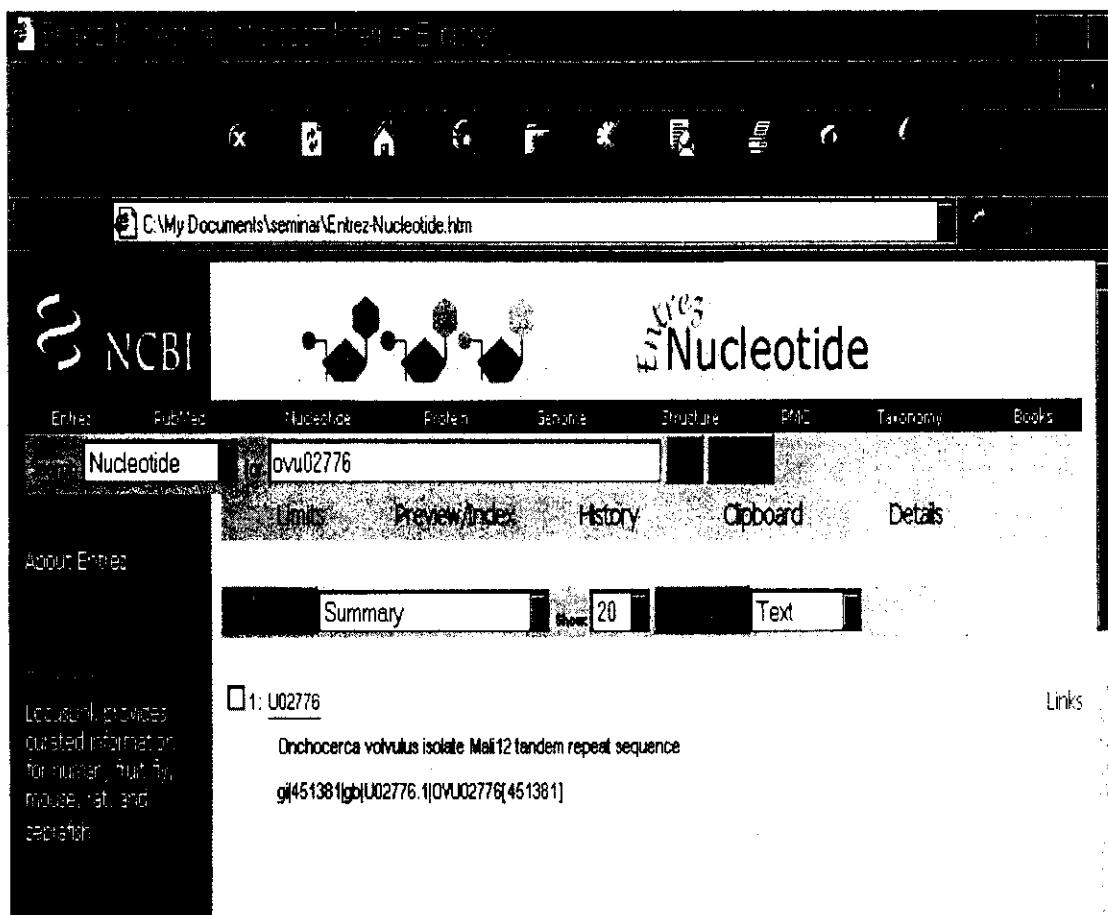
IV. การค้นหาข้อมูลทางอณูชีววิทยาจากฐานข้อมูลบนเครือข่ายอินเตอร์เน็ต

1. เปิดโปรแกรม Netscape หรือ Internet Explorer ด้วยการทำ double click บนไอคอน
2. เปิดดู <http://www.ncbi.nlm.nih.gov/> จะได้ผลดังรูปที่ 1



รูปที่ 1

3. พิมพ์รหัสหรือชื่อเชื้อแบคทีเรียที่ต้องการค้นหาสำาดับดีเอ็นเอ ลงในช่องว่างที่อยู่หลังคำว่า “search nucleotide”
4. กดตรงปุ่ม “GO” จะได้ผลดังรูปที่ 2



รูปที่ 2

5. เลือกเชื้อที่สนใจ โดยการ click เข้าไปจะได้รายละเอียดของเชื้อ ชื่อเจ้าของข้อมูล เชื่อหัวเรื่อง ชื่อวารสาร ชนิดของข้อมูล รวมไปถึงลำดับของเบสที่สนใจ ดังรูปที่ 3 และ รูปที่ 4 ตามลำดับ
6. ทดลองดูข้อมูลในลักษณะอื่นๆ เช่น FASTA หรือข้อมูลอื่นที่เกี่ยวข้องกับข้อมูลที่สนใจ

 NCBI

Search | Help | Log In | My NCBI | Helpdesk | Feedback | About NCBI | NCBI Home | Nucleotide

Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Ebo

Nucleotide

default 20 File

1: Z77548. Bacterium 2-400 C...[gi:1465565] Links

LOCUS BSPZ77548 368 bp DNA linear BCT 26-JUL-1996

DEFINITION Bacterium 2-400 C2/26 partial 16S rRNA gene.

ACCESSION Z77548

VERSION Z77548.1 GI:1465565

KEYWORDS 16S ribosomal RNA; 16S rRNA gene; small subunit ribosomal RNA.

SOURCE Cyanobacteria (blue-green algae)

ORGANISM Cyanobacteria

รูปที่ 3

source= EMBL 22/220
/db_xref="taxon:1117"
gene 1..368
/gene="16S rRNA"
rRNA <1..>368
/gene="16S rRNA"
/product="16S ribosomal RNA"
/note="obtained by PCR with universal eubacterial primers,
approximately homologous to *E. coli* 180-500"
ORIGIN
1 gttaccta ggcgacgg tggatcactg gagatgtcc ttagagggtt acaattggaa
61 tgatgtctta taccatcg cttagatgtt aatggatttt tttttatcg ccttseagaga
121 agtcgcggc tggatggc ttggggagg taagaccta cccaggcgc acatcgatc
181 tggtttggaa ggacgttccg acacactgg aatggacac ggtccagatc cttacgggg
241 gcggcgggg ggaatttcc gcaatgggg aatggatgc gggatcaccc cggttgggg
301 aatggatcc tgggttta aaccctttt ttcgggggg aataaaatgtt cttttttttt
361 aatggatcc
//

รูปที่ 4

ผลการทดลอง

- จงยกตัวอย่างฐานข้อมูลพร้อมวิธีคิดต่อ (URL ของ website) มาสัก 3 ฐานข้อมูล

.....

.....

.....

.....

.....

.....

.....

- ถ้านักศึกษาต้องการวิเคราะห์ข้อมูลทางอนุชีววิทยาโดยใช้โปรแกรมคอมพิวเตอร์ นักศึกษาคิดว่าอย่างไรให้โปรแกรมคอมพิวเตอร์วิเคราะห์ข้อมูลแบบได้บ้าง

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

แบบฝึกหัดบทปฎิบัติการที่ 9

- การหาลำดับเบสด้วยเครื่องอัตโนมัติ มีข้อดีอย่างไรบ้าง
- แนวโน้มหรือทิศทางการหาลำดับเบสและการวิเคราะห์โดยคอมพิวเตอร์ จะมีการเปลี่ยนแปลงในรูปแบบใด
- Bioinformatics จะมีบทบาทอย่างไรบ้างในอนาคต

