

# เอกสารประกอบปฏิบัติการ Basic Bioinformatics

## การใช้ฐานข้อมูลชีวภาพ NCBI และโปรแกรม BLAST

อ.ดร.ทศพล เตโช

### พื้นฐานความรู้ที่ควรรู้และเกี่ยวข้องกับปฏิบัติการ

#### 1.1 BLAST: เครื่องมือค้นหาแห่งโลกชีววิทยา

BLAST (Basic Local Alignment Search Tool) เป็นโปรแกรมคอมพิวเตอร์ที่ได้รับการพัฒนาโดย Stephen Altschul และทีมงานในปี ค.ศ. 1990 ถือเป็นหนึ่งในเครื่องมือที่สำคัญที่สุดในโลกของชีวสารสนเทศ

ลองคิดว่า BLAST เป็นเหมือน "Google สำหรับดีเอ็นเอ" หากเรามีลำดับดีเอ็นเอหรือโปรตีนลำดับหนึ่งที่ไม่ทราบว่าคืออะไร เราสามารถใช้ BLAST ค้นหาในฐานข้อมูลขนาดยักษ์ในฐานข้อมูลเพื่อหาลำดับที่คล้ายคลึงกัน

#### หลักการการทำงานของ BLAST: เหมือนการเปรียบเทียบประโยค

สมมติเรามีข้อความที่ไม่ชัดเจน: "นกกา\_\_ใหญ่\_\_าน" และเราต้องการหาว่าประโยคเต็มควรจะเป็น "นกกาเหว่าใหญ่บินผ่าน"

#### ขั้นตอนที่ 1: แบ่งเป็นคำสั้นๆ (Word Breaking)

- เราจะแบ่งข้อความที่มี "นกกา\_\_ใหญ่\_\_าน" เป็นชิ้นเล็กๆ
- เช่น: "นกกา", "ใหญ่", "าน" (คำที่อ่านได้ชัดเจน)
- ในกรณีของดีเอ็นเอ: ลำดับ ATGGCTAAG จะถูกแบ่งเป็น ATG, TGG, GGC, GCT, CTA, TAA, AAG

#### ขั้นตอนที่ 2: ค้นหาคำที่เหมือนในฐานข้อมูล (Database Search)

- เราจะเอาคำ "นกกา" ไปค้นหาในพจนานุกรมหรือหนังสือทั้งหมดที่มี
- พบว่ามีประโยคในฐานข้อมูล: "นกกาเหว่าใหญ่บินผ่าน", "นกกาตัวเล็กๆ กิน", "นกกาขาวสวยงาม"
- ในกรณีของดีเอ็นเอ: ค้นหาว่าชุด ATG ปรากฏในลำดับใดบ้างในฐานข้อมูล GenBank

#### ขั้นตอนที่ 3: ขยายการเปรียบเทียบ (Alignment Extension)

- เมื่อเจอ "นกกา" แล้ว เราจะดูต่อไปว่าคำถัดไปตรงกันไหม
- เปรียบเทียบ: "นกกา\_\_ใหญ่\_\_าน" กับ "นกกาเหว่าใหญ่บินผ่าน"
- "นกกา" ✓ ตรงกัน
- "ใหญ่" ✓ ตรงกัน
- "าน" จากคำ "ผ่าน" ✓ ตรงกัน
- ในกรณีของดีเอ็นเอ: ขยายไปดูว่าลำดับข้างเคียงตรงกันแค่ไหน

#### ประเภทของ BLAST: เครื่องมือสำหรับงานต่างๆ

เหมือนกับที่เรามีเครื่องมือต่างๆ สำหรับงานต่างๆ ใช้ใหม่ มีค้อนสำหรับตอกตะปู มีเลื่อยสำหรับตัดไม้ BLAST ก็มีประเภทต่างๆ สำหรับงานต่างๆ เช่นกัน:

**blastn** (nucleotide-nucleotide): เปรียบเทียบดีเอ็นเอกับดีเอ็นเอ

- ใช้เมื่อ: ต้องการหาชิ้นที่คล้ายกัน หรือระบุชนิดของสิ่งมีชีวิต
- ตัวอย่าง: เปรียบเทียบลำดับดีเอ็นเอจากเลือดในที่เกิดเหตุกับฐานข้อมูลเพื่อหาว่าเป็นเลือดของสัตว์ชนิดไหน

**blastp** (protein-protein): เปรียบเทียบโปรตีนกับโปรตีน

- ใช้เมื่อ: ต้องการหาหน้าที่ของโปรตีน หรือโปรตีนที่มีโครงสร้างคล้ายกัน

- ตัวอย่าง: ค้นหาว่าโปรตีนใหม่ที่พบมีหน้าที่อะไร โดยเปรียบเทียบกับโปรตีนที่รู้หน้าที่แล้ว  
**blastx** (nucleotide query vs protein database): แปลดีเอ็นเอเป็นโปรตีนแล้วเปรียบเทียบ
- ใช้เมื่อ: มีลำดับดีเอ็นเอแต่ต้องการหาว่าจะแปลเป็นโปรตีนอะไร  
**tblastn** (protein query vs nucleotide database): เปรียบเทียบโปรตีนกับดีเอ็นเอที่แปลแล้ว
- ใช้เมื่อ: มีโปรตีนแต่ต้องการหาโปรตีนที่สร้างโปรตีนนั้น

## 1.2 เข้าใจผลลัพธ์ BLAST: การอ่านคะแนน

เมื่อใช้ BLAST แล้ว เราจะได้ผลลัพธ์ที่มีค่าต่างๆ เหล่านี้:

1. **E-value (Expect value):** ความน่าจะเป็นที่จะพบผลลัพธ์นี้โดยบังเอิญ
  - E-value = 0.001 หมายความว่า มีโอกาส 1 ใน 1,000 ที่ผลนี้เกิดขึ้นโดยบังเอิญ
  - ยิ่ง E-value ต่ำ = ยิ่งมั่นใจได้ว่าผลลัพธ์ถูกต้อง
  - E-value < 0.05 ถือว่าเชื่อถือได้
2. **Identity (%):** เปอร์เซ็นต์ความเหมือนกัน
  - 90% Identity = เหมือนกัน 90% ต่างกัน 10%
  - ยิ่งสูงยิ่งดี แต่ 100% ไม่จำเป็นเสมอไป
3. **Query Coverage (%):** เปอร์เซ็นต์ของลำดับที่เราค้นหาที่ถูกจับคู่ได้
  - 80% Coverage = ลำดับของเราจับคู่ได้ 80% เหลือ 20% ที่ไม่จับคู่
  - ควรสูงกว่า 70% จึงจะเชื่อถือได้
4. **Bit Score:** คะแนนที่ปรับแล้ว ใช้เปรียบเทียบผลลัพธ์ต่างๆ
  - คะแนนสูง = การจับคู่ดี
  - ใช้เปรียบเทียบว่าผลลัพธ์ไหนดีกว่ากัน

---

## การเตรียมตัวสำหรับปฏิบัติการ

ในปฏิบัติการนี้ เราจะได้เรียนรู้การใช้เครื่องมือเหล่านี้จริงๆ ผ่านการแก้ปัญหา "ลำดับลึกลับ" 3 ลำดับ:

1. ลำดับลึกลับ 1: ลำดับดีเอ็นเอที่ไม่ทราบที่มา
2. ลำดับลึกลับ 2: ลำดับโปรตีนที่ไม่ทราบหน้าที่
3. ลำดับลึกลับ 3: ลำดับ 16S rRNA สำหรับระบุชนิดแบคทีเรีย

เมื่อจบปฏิบัติการ นักเรียนจะสามารถใช้เครื่องมือเหล่านี้ในการแก้ปัญหาทางชีววิทยาได้ด้วยตนเอง และเข้าใจว่าโลกแห่งข้อมูลชีววิทยานั้นกว้างใหญ่และน่าตื่นตาตื่นใจแค่ไหน!

---

## วัตถุประสงค์ของปฏิบัติการ

### วัตถุประสงค์หลัก

1. เพื่อให้นักเรียนสามารถใช้เครื่องมือ BLAST ในการเปรียบเทียบลำดับและค้นหาลำดับที่มีความคล้ายคลึง
2. เพื่อฝึกทักษะการวิเคราะห์และตีความผลลัพธ์จาก BLAST อย่างถูกต้อง

### ผลการเรียนรู้ที่คาดหวัง (Learning Outcomes)

เมื่อสิ้นสุดปฏิบัติการแล้ว นักเรียนจะสามารถ:

- อธิบายหลักการและความสำคัญของชีวสารสนเทศได้

- ใช้งานเว็บไซต์ NCBI ได้อย่างมีประสิทธิภาพ
- ค้นหาและดาวน์โหลดลำดับดีเอ็นเอและโปรตีนจากฐานข้อมูลได้
- ใช้โปรแกรม BLAST ได้อย่างถูกต้องและเหมาะสม
- วิเคราะห์และตีความผลลัพธ์ BLAST ได้อย่างถูกต้อง

## วัสดุ และอุปกรณ์

### อุปกรณ์หลัก

1. คอมพิวเตอร์หรือแท็บเล็ต (1 เครื่องต่อนักเรียน 1-2 คน)
  - ระบบปฏิบัติการ: Windows, macOS, หรือ Linux
  - หน่วยความจำ: อย่างน้อย 4 GB RAM
  - พื้นที่เก็บข้อมูล: เหลือพื้นที่ว่างอย่างน้อย 1 GB
2. การเชื่อมต่ออินเทอร์เน็ต
  - ความเร็วอย่างน้อย 10 Mbps ต่อคอมพิวเตอร์
  - การเชื่อมต่อที่เสถียรตลอดระยะเวลาปฏิบัติการ
3. เว็บเบราว์เซอร์ (อัปเดตเป็นเวอร์ชันล่าสุด)
  - Google Chrome (แนะนำ)
  - Mozilla Firefox
  - Microsoft Edge
  - Safari (สำหรับ macOS)

### ซอฟต์แวร์และเว็บไซต์

1. NCBI Website: <https://www.ncbi.nlm.nih.gov/>
2. BLAST Suite: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
3. Text Editor (สำหรับบันทึกข้อมูล):
  - Notepad++ (Windows)
  - TextEdit (macOS)
  - gedit (Linux)

### ข้อมูลตัวอย่างสำหรับปฏิบัติการ

ลำดับลิกลับ 1: ลำดับดีเอ็นเอที่ไม่ทราบที่มา

```
>Mystery_Sequence_1
TTTGCCCCAGCTGCGAGCAGGGACAGGTCTGGCCACCGGGCCCTGGTAAAGACTCTAATAACCCGCTGGCCCTGAGGAAGAGGTGCTGACGACC
AAGGAGATCTTCCACAGACCCAGCACCAGGAAATGGTCCGGAATTCAGGCCTCAGCCCCAGCCATCTGCCGACCCCCACCCAGGCCCTA
ATGGGCCAGGCGGAGGGTTGACAGGCAGGGGAGATGGGCTCTGAGACTATAAAGCCAGTGGGGACCCAGCAGCCCTCAGCCCTCAGGACAGGC
TGATCAGAAAGAGGCCATCAAGCAGGTCTGTTCCAAGGGCCTTTGCGTCAGATCACTGTCCTTCTGCCATGGCCCTGTGGATGCGCCTCCTGCCCC
TGCTGGCGCTGCTGGCCCTCTGGGGACCTGACCCAGCCGCGGCCTTTGTGAACCAACACCTGTGCGGCTCCACCTGGTGGAAGCTCTCTACCTAG
TGTGCGGGGAACGAGGCTTCTTACACACCCAAGACCCGCCGGGAGGCAGAGGACCTGCAGGTGGGGCAGGTGGAGCTGGGCGGGGCCCTGGTG
CAGGCAGCCTGCAGCCCTGGCCCTGGAGGGTCCCTGCAGAAAGCGTGGCATCGTGGAACAGTGTGTACCAGCATCTGCTCCCTCTACCAGCTGG
AGAATACTGCAACTAGATGCAGCCGACGAGCCACACCCCTCCGCTCCTGCACCGAGAGACATGGAATAAAGCCCTGAACCAGC
```

ลำดับลิกลับ 2: ลำดับโปรตีนที่ไม่ทราบหน้าที่

```
>Mystery_Sequence_2
MSKRKAPQETLNGGITDMLTELANFEKNVSAIHKYNAYRKAASVIAKYPHKIKVGTGKIAEKIDFLATGKLRKLEKIRQDDTSSSINFLTRVSGI
GPSAARKFVDEGIKTLEDLRKNEDKLNHHQRIGLKYFGDFEKIRPREEMLQMQDIVLNEVKKVDSEYIATVCGSFRRGAESSGDMVLLTHPSFTS
ESTKQPKLLHQVVEQLQKVHFITDTLSKGETKFMGVQCQLPSKNDEKEYPHRRIDIRLIPKDYQYCGVLYFTGSDIFNKNMRAHALEKGFITNEYTI
RPLGVTGVAGEPLPVDSEKDFDYIQWKYREPKDRSE
```

ลำดับที่กลับ 3: ลำดับ 16S rRNA สำหรับระบุชนิดแบคทีเรีย

>Mystery\_Sequence\_3

```
AGAGTTTGTCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAACGGTTGATGGGAGCTTGCTCCCTGATGTTAGCG
ACGGACGGGTGAGTAACACGCTGGGTAACTGCTGAAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGATGCTTTGTTGAACCGCATGG
TTCGAAGATGAAAGCGGCTTTGCTAGCCTTGGATGGATCCCGCGCGTATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCATAG
CCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGT
CTGACGGAGCAACGCCCGCTGAGTGATGAAGGCTTTCGGGTCGTAAGACTCTGTTGTTAGGGAAGAACAAGTACCGTTTCAATAAGGTGGTACCTT
GACGGTACCTGACCAGAAAGCCACGGCTAATACGTGCCAGCAGCCGCGTAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGC
GCGCGCAGGTGGTTTCTTAAGTCTGATGTGAAAGCCACGGCTCAACCGTGGAGGGTCATTGGAAGTGGGAGACTTGAGTGCAGAAGAGGAAAGT
GGAATTCATGTGTAGCGGTGAAATGCGCAGAGATATGGAGGAACACCAAGTGGCGAAGGCGACTTCTGGTCTGTAACCTGACGCTGATGTGCGAAA
GCGTGGGGATCAACAGGATTAGATACCTGGTAGTCCACGCGCTAAACGATGAGTGCTAAGTGTAGAGGGTTTCCGCCCTTTAGTGCTGAAGTT
AACGCATTAAGCACTCCGCTGGGGAGTACGGCCGAAGGCTGAACTCAAAGGAATTGACGGGGGCCGACAAGCGGTGGAGCATGTGGTTTAA
TTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTGACAATCCTAGAGATAGGACGTCCCTTCGGGGGCGAGAGTGACAGGTGGTGCA
TGGTTGTCTGTCAGCTCGTGTCTGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGTCTTTGTTGCCAGCGGTCCGGCCGGGAACCAA
AGGAGACTGCCAGTGATAAATGGAGGAAGTGGGGATGACGTCAAATCATCATGCCCTTATGATTGGGCTACACACGTGCTACAATGGACGGT
ACAAAGGGCTGCAAGACCGCGAGGTGGAGCTAATCCCATAAACCGATCGTAGTCCGGATCGCAGTCTGCAACTCGACTGCGTGAAGTCGGAATCG
CTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCTTGTACACACCGCCGTCACACCAGAGAGTTGTAACACCCGAAGTCGG
TGAGGTAACCTTTTAGGAGCCAGCCGCCGAAGGTGGGGCAGATGATTGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAAGGTGCGGCTGGATCA
CTCTCT
```

## กิจกรรมที่ 1: การใช้งาน BLAST - blastn

### ขั้นตอนที่ 1.1: การเตรียมข้อมูล query

1. ไปที่ <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
2. เลือก "Nucleotide BLAST" (blastn)
3. ในช่อง "Enter Query Sequence" ให้ copy-paste ลำดับ Mystery\_Sequence\_1

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query. more...

**Enter Query Sequence**

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

TGGAACAGTGCTGTACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTAC  
TGCAACTAGATGCAGCCCGCAGGCGAGCCCGCACACCTCTCGCTCTGCA  
CCGAGAGACATGGAATAAGCCCTGAACCCAGC

Query subrange [?](#)

From

To

Or, upload file  No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

**Choose Search Set**

Database ☒ Standard databases (nr etc.): ☐ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus ☐ Experimental databases

Core nucleotide database (core\_nt) [?](#)

Organism  ☐ exclude [Add Organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

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

Limit to ☐ Sequences from type material

Entrez Query  [YouTube](#) [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 core\_nt using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window

**+ Algorithm parameters**

### ขั้นตอนที่ 1.2: การตั้งค่าพารามิเตอร์

1. เลือกฐานข้อมูล:
  - Core nucleotide database (core\_nt) สำหรับการค้นหาทั่วไป

- หรือ **Reference RNA sequences (refseq\_rna)** สำหรับลำดับที่ใช้ในการอ้างอิงความแม่นยำสูง
2. เลือก
    - megablast (ความแม่นยำสูงจะมีโอกาสเจอลำดับที่เหมือนกันน้อย)
    - blastn (ความแม่นยำต่ำจะมีโอกาสเจอลำดับที่เหมือนกันหลายลำดับ)
  3. คลิก "BLAST"

### ขั้นตอนที่ 1.3: การวิเคราะห์ผลลัพธ์

1. Graphic Summary: สังเกตแถบสีที่แสดงระดับความคล้าย:
  - แดง: > 200 bits (identity สูงมาก)
  - ชมพู: 80-200 bits (identity สูง)
  - เขียว: 50-80 bits (identity ปานกลาง)
  - น้ำเงิน: 40-50 bits (identity ต่ำ)
  - ดำ: < 40 bits (identity ต่ำมาก)
2. Descriptions: ดูรายการผลลัพธ์ที่จัดเรียงตาม Max Score:
  - สังเกต Query Cover (%)
  - E-value
  - Per. Ident (%)
3. Alignments: คลิกที่ผลลัพธ์แรกเพื่อดูการจัดเรียงลำดับ:
  - บันทึก Score และ E-value
  - นับจำนวน identities, gaps
  - สังเกตพื้นที่ที่ match (||||), mismatch, gaps (-)

## กิจกรรมที่ 2: การใช้งาน BLAST - blastp

### ขั้นตอนที่ 2.1: การเตรียม query โปรตีน

1. ไปที่หน้าหลัก BLAST แล้วเลือก "Protein BLAST" (blastp)
2. ใน query box ให้ paste ลำดับ Mystery\_Sequence\_2

**Standard Protein BLAST**

blastn
**blastp**
blastx
tblastn
tblastx

BLASTP programs search protein databases using a protein query. more...

### Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file  No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

### Choose Search Set

Database ?

Organism Optional  ☐ exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude Optional ☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

### Program Selection

Algorithm

☐ Quick BLASTP (Accelerated protein-protein BLAST)  
☒ blastp (protein-protein BLAST)  
☐ PSI-BLAST (Position-Specific Iterated BLAST)  
☐ PHI-BLAST (Pattern Hit Initiated BLAST)  
☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

**BLAST** Search database nr using Blastp (protein-protein BLAST)

☐ Show results in a new window

## ขั้นตอนที่ 2.2: การตั้งค่าและค้นหา

- เลือกฐานข้อมูล:
  - Non-redundant protein sequences (nr)
  - ClusteredNR (nr\_cluster\_seq)
  - Reference Proteins (refseq\_protein)
- เลือก Algorithm: **blastp**
- คลิก "BLAST"

## กิจกรรมที่ 3: การวิเคราะห์ 16S rRNA เพื่อระบุชนิดของแบคทีเรีย

### ขั้นตอนที่ 3.1: การใช้ blastn กับ 16S database

- เข้า blastn อีกครั้ง
- Paste ลำดับ Mystery\_Sequence\_3
- เปลี่ยนฐานข้อมูลเป็น **rRNA/ITS database**
  - เลือก Database เป็น **16S ribosomal RNA sequences (Bacteria and Archaea)**
- คลิก "BLAST"

blastn

blastp

blastx

tblastn

tblastx

Standard Nucleotide BLAST

BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter Query Sequence

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

AGGAATCTACCTAGTAGTGGGGGATAACACCTGGAAACGGATGCTAATAC  
CGCATAACATCTTCCGCATGGAGGGGATTGAATTCAAATCTTGCTTTTCAGT  
ACCGCGACGAGTACCTAAGGAAAGCACATAGTTGCAACAAGG

Query subrange ?

From

To

Or, upload file

Choose File No file chosen ?

Job Title

Mystery\_Sequence\_3

Enter a descriptive title for your BLAST search ?

☐ Align two or more sequences ?

Choose Search Set

Database

☐ Standard databases (nr etc.):
☒ rRNA/ITS databases
☐ Genomic + transcript databases
☐ Betacoronavirus
☐ Experimental databases

16S ribosomal RNA sequences (Bacteria and Archaea) ? Targeted Loci Project Information

Organism

Optional

☐ exclude
Add Organism

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

YouTube 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 16S ribosomal RNA sequences (Bacteria and Archaea) using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window

### ขั้นตอนที่ 3.2: การระบุชนิดแบคทีเรีย

- ดูผลลัพธ์ที่มี Max Identity สูงที่สุด (ควรจะ > 97%)
- บันทึกข้อมูล:
  - ชื่อสปีชีส์ที่น่าจะเป็น
  - % Identity
  - Query Coverage
  - E-value
- เปรียบเทียบผลลัพธ์ 3-5 อันดับแรก:
  - ชื่อสปีชีส์เหมือนกันหรือไม่
  - ความแตกต่างของ % Identity

#### กิจกรรมที่ 4: จงตรวจสอบข้อมูลลำดับต่อไปนี้โดยใช้โปรแกรม BLAST

##### >Mystery\_Sequence\_D

```
ATGGAGGAGCCGCGAGTCAGATCCTAGCGTCGAGCCCCCTCTGAGTCAGGAAACATTTTCAGACCTATGGAACTACTTCTGAAAACACAGTTCTG
TCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATTGAACAATGGTTCACCTGAAGACCCAGGTCCAGATGAAGCTCCC
AGAATGCCAGAGGCTGCTCCCGCGTGGCCCTGCACCAGCAGCTCCTACACCGCGGCCCCCTGCACCAGCCCCCTCTGGCCCCCTGTCATCTTCT
GTCCCTTCCAGAAAACCTACCAGGGCAGCTACGGTTTCCGTCTGGGCTTCTTGCACTTCTGGGACAGCCAAGTCTGTGACTTGCACGTACTCCCT
GCCCTCAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCTGTGACGCTGTGGGTTGATTCCACACCCCCGCGCCGACCCGCGTCCGCGCCATG
GCCATCTACAAGCAGTCACAGCACATGACGGAGGTTGTGAGGCGCTGCCCCACCATGAGCGCTGCTCAGATAGCGATGGTCTGGCCCCCTCTCAG
CATCTTATCCGAGTGGAAGGAAATTTGCGTGTGGAGTATTTGGATGACAGAAACACTTTTCGACATAGTGTGGTGGTGCCCTATGAGCCGCTGAG
GTTGGCTCTGACTGTACCACCATCCACTACAACACATGATGTAACAGTTCTTGATGGGCGGCATGAACCGGAGGCCATCCTCACCATCATCACA
CTGGAAGACTCCAGTGGTAATCTACTGGGACGGAACAGCTTTGAGGTGCGTGTTTGTGCTGTCTGGGAGAGACCGGCGCACAGAGGAAGAGAAT
CTCCGCAAGAAAGGGGAGCCTCACCACGAGCTGCCCCAGGGAGCACTAAGCGAGCACTGCCCAACAACACCAGCTCCTCTCCCGAGCCAAAGAAG
AAACCACTGGATGGAGAATATTTACCCCTTCAGATCCGTGGGCGTGAGCGCTTCGAGATGTTCCGAGAGCTGAATGAGGCCTTGGAACCTCAAGGAT
GCCAGGCTGGGAAGGAGCCAGGGGGGAGCAGGGCTCACTCCAGCCACCTGAAGTCCAAAAAGGGTCAGTCTACCTCCCGCCATAAAAACTCATG
TTCAAGACAGAAGGCGCTGACTCAGACTGA
```

##### >Mystery\_Sequence\_E

```
MADKKAVEKSVQEIQATFFYEHPQDLQTPKDGKGLNKLQKDIHKDRIEIQRLKYKVDEPLTKGKIQEIDKLLLELDSVDGKTIHHGGAGMLS
SLAHALAANPGVQIGSANWKEEMQKFREEAMKILKSLPSRLQIDEQERRIFPNKDFEAKTHKEIQAAKSGFGMKSIDFFEKVLGTEQFIAQRSKT
LKELEKVAFPQGVFPEKLRSHQFAVTTDPKVADFEMADLTKADILLKDPKVPDQSKFPDLFENSFAQFSEPMKTLQPQIYVGSIDAFALQQF
GDLALFTIKGLPGKEGIIGLPDFSMQGSALKLVDPKAIQONPDGFQAPQRNMEAAIEAIRKAGAPIGAAGIATVRMAQDGVKGYTMGGVDWISK
PDDVMQWKKGDLDDAAKDKVDKLAGSAVKLSPLSYEDPVQALNSLGLSANQSLDQVKSVFPEFEGKLTKEAEKHQKVASGPKSIDGLKQKLDLFR
TNSFDKAFGGSYESLTPDGIKKLGNATFKDGASALQVFEVKKPKGKGQKFEFNKYKPNVFKDLTGKPVTNKKVK
```

##### >Mystery\_Sequence\_F

```
TGGCTCAGGATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGAAGCACTTTATTTGATTTGCTTGAACCATTGATTCAATCCATGATTC
AATACTCGAGTGGCGAACGGGTGAGTAACACGTGGGCAACCTGCCCATAGACTGGGATAAATCCGGGAAACCGGGGCTAATACCGGATGGTTGTT
TGAACCGCATGGTTCAAGGATGAAAGGCGGCTTTGCTATCATTATAGATGGATCCGCGCTGCATTAGCTAGTTGGTGGGGTAACGGCTCACCAAG
GCAACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCG
CAATGGACGAAAAGTCTGACGGAGCAACGCCGCTGAGTGATGAAGGCTTTCGGGTCTGTAATACTCTGTTGTTAGGGAAGAACAAGTCTAGTTGAA
TAAGCTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGTAATACGTAGGTGGCAAGCGTTGTCCGGAATT
ATTGGGCGTAAAGGGCTCGCAGGCGGTTTCTTAAGTCTGATGTGAAAGCCCCGGCTCAACCGGGGAGGGTCAATTGGAACCTGGGGAACCTGAGTG
CAGAAGAGGAGAGTGGAATCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAACACCAAGTGGCGAAGGCGACTCTCTGGTCTGTAACGTAC
GCTGAGGAGCGAAAGCGTGGGGAGCAACAGGATTAGATACCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTAGGGGGTTTCCGCCCT
TTAGTGCTGAAGTTAACGCATTAAGCACTCCGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGG
AGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTGAAAACCTAGAGATAGGGCTTCTCCTTGGGAGCAGAG
TGACAGGTGGTGCATGGTTGTCGTGAGTCTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCTTGATCTTAGTTGCCATCATTAA
GTTGGGCACTCTAAGGTGACTGCCGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCTTATGACCTGGGCTACACACGTGC
TACAATGGACAGAAACAAAGGGCTGCGAGACCGCGAGGTTTAGCCAATCCATAAATCTATTCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACAT
GAAGCCGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGATACACACCGCCGTCACACCACGAGAGTTTGTA
CACCCGAAGTCGGTGGGGTAACCTTTTAGGAACAGCCGCCCTAAGGTGGGACAGATGATTGGGGTGAAGTCGTAAACAAGGTAGCCGTATCGGAAG
TGGCGCTGGATCACCTCCTTT
```



ตารางบันทึกผลการทดลอง

ตาราง 1: ผลการวิเคราะห์ Mystery\_Sequence\_1 (blastn)

ลำดับที่	Scientific Name	Max Score	Query Cover (%)	E-value	Per. Ident (%)	Accession
1						
2						
3						

ข้อสันนิษฐานเกี่ยวกับฟังก์ชันของลำดับนี้: .....

ตาราง 2: ผลการวิเคราะห์ Mystery\_Sequence\_2 (blastp)

ลำดับที่	Protein Description	Max Score	Query Cover (%)	E-value	Per. Ident (%)	Organism
1						
2						
3						

ข้อสันนิษฐานเกี่ยวกับฟังก์ชันของโปรตีนนี้: .....

ตาราง 3: การระบุชนิดจาก 16S rRNA (Mystery\_Sequence\_3)

ลำดับที่	Scientific Name	Max Score	Query Cover (%)	E-value	Per. Ident (%)
1					
2					
3					
4					
5					

สปีชีส์ที่น่าจะเป็น: .....

ความเชื่อมั่นในการระบุ (1-5): ..... เหตุผล: .....

ตาราง 4: ผลการวิเคราะห์ Mystery\_Sequence\_D

ลำดับที่	Scientific Name	Max Score	Query Cover (%)	E-value	Per. Ident (%)	Accession
1						
2						
3						

ข้อสันนิษฐานเกี่ยวกับฟังก์ชันของลำดับนี้: .....

ตาราง 5: ผลการวิเคราะห์ Mystery\_Sequence\_E

ลำดับที่	Protein Description	Max Score	Query Cover (%)	E-value	Per. Ident (%)	Organism
1						
2						
3						

ข้อสันนิษฐานเกี่ยวกับฟังก์ชันของโปรตีนนี้: .....

ตาราง 6: ผลการวิเคราะห์ Mystery\_Sequence\_F

ลำดับที่	Scientific Name	Max Score	Query Cover (%)	E-value	Per. Ident (%)
----------	-----------------	-----------	-----------------	---------	----------------

1					
2					
3					
4					
5					

สปีชีส์ที่น่าจะเป็น: .....

ความเชื่อมั่นในการระบุ (1-5): ..... เหตุผล: .....