

## รายงานวิชา SC187002 Practical skill in Biology (PSB)

### หัวข้อ NCBI database and BLAST

ตอนที่ 1 ศึกษาความสัมพันธ์เชิงวิวัฒนาการของยีนที่นักศึกษามีความสนใจโดยใช้เครื่องมือ และข้อมูลในฐานข้อมูล NCBI เท่านั้น

#### 1. ที่มาและความสำคัญของยีนที่สนใจ

##### 1.1 ยีนที่สนใจ

Gene symbol : *CDKA1*

Gene description : hypothetical protein

Organism : *Phaeodactylum tricornutum*

##### 1.2 ที่มาและความสำคัญ

*CDKA1* (Cyclin-Dependent Kinase A1) เป็นยีนที่เข้ารหัสโปรตีนชนิดหนึ่งในกลุ่ม cyclin-dependent kinases (CDKs) ซึ่งมีบทบาทสำคัญในการควบคุม วัฏจักรเซลล์ (cell cycle) โดยเฉพาะในระยะที่เซลล์เตรียมตัวแบ่ง (G1-to-S phase) และระยะเข้าสู่การแบ่งเซลล์ (G2-to-M phase) หากยีนนี้หยุดทำงานเซลล์จะหยุดการเจริญเติบโตหรือไม่สามารถแบ่งตัวได้ ยีนนี้เป็นยีนที่พบได้ในสิ่งมีชีวิตกลุ่มพืชและสาหร่ายเป็นยีนหลักในการควบคุมวัฏจักรเซลล์ของสิ่งมีชีวิตเหล่านี้

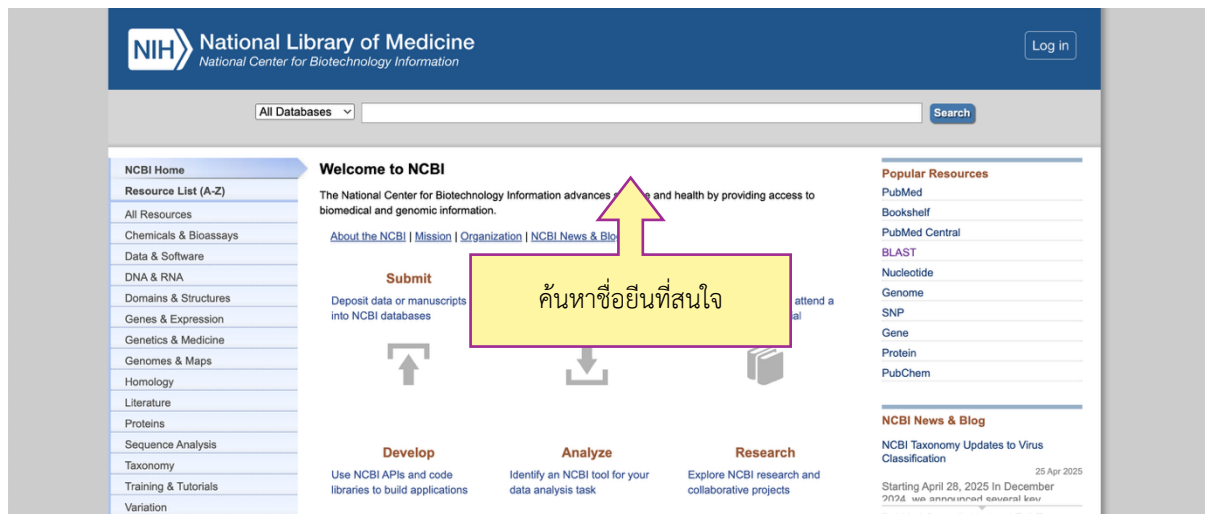
ในสาหร่ายและสิ่งมีชีวิตชั้นต่ำ เช่น *Phaeodactylum tricornutum* *CDKA1* มีบทบาทสำคัญในการตอบสนองต่อ สิ่งเร้าจากสิ่งแวดล้อม เช่น แสง โดยทำงานร่วมกับไซคลินเฉพาะกลุ่ม เช่น dsCYC2

##### 1.3 จำนวนลำดับเบสทั้งหมดของยีน : 3774 bp

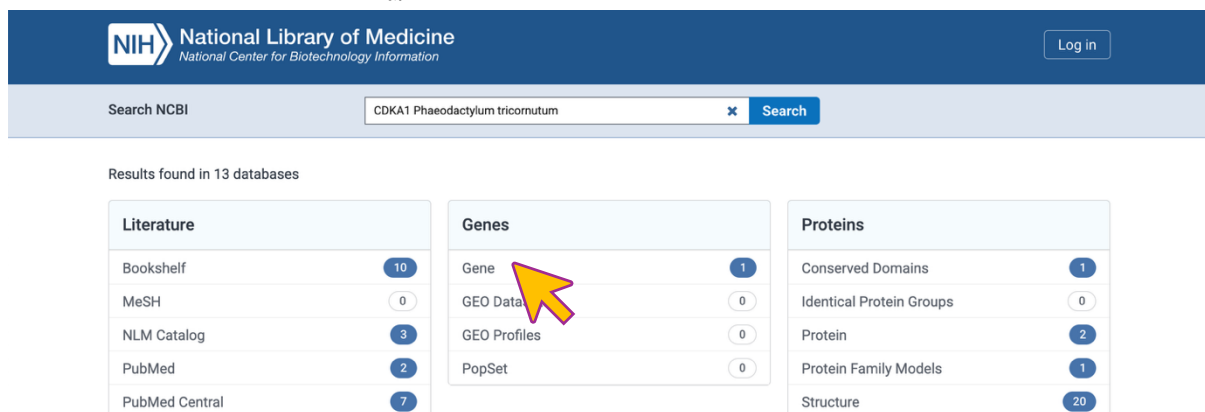
##### 1.4 จำนวนลำดับกรดอะมิโนของยีน : 290 aa

## 2. ขั้นตอนการดำเนินงาน

2.1 เข้าสู่เว็บไซต์ <https://www.ncbi.nlm.nih.gov/> จะปรากฏหน้าจอดังภาพ จากนั้นค้นหาที่น่าสนใจในช่องค้นหา



2.2 เมื่อกดค้นหาแล้วจะปรากฏหน้าจอดังภาพ ให้เลือกที่ “Gene”



2.3 เมื่อปรากฏหน้าจอแสดงรายละเอียดของยีนที่น่าสนใจ ให้คลิกที่แถบด้านขวา “NCBI Reference Sequences (RefSeq)”



## 2.4 เมื่อปรากฏหน้าจอดังภาพ สามารถคลิกเพื่อดูรายละเอียดของ Gene, mRNA และ Protein ได้

NCBI Reference Sequences (RefSeq)

NEW Try the new Transcript table

Genome Annotations

The following sequences are aligned to a specific genome build. [Explain](#)

Reference assembly

Genomic

1. NC\_011676.1 Reference

Range 494840..498613

Download [GenBank](#), [FASTA](#), [Sequence Viewer \(Graphics\)](#)

mRNA and Protein(s)

1. XM\_002180023.1 → XP\_002180059.1 predicted protein [Phaeodactylum tricornutum CCAP 1055/1]

[See identical proteins and annotated locations for XP\\_002180059.1](#)

Status: PROVISIONAL

UniProtKB/TrEMBL [B](#)

Conserved domains

LN00009; cyclin-dependent kinase A; Provisional

Kc\_like; Protein Kinases, catalytic domain

Related sequences

Nucleotide

Protein

predicted protein [Phaeodactylum tricornutum CCAP 1055/1]

Phaeodactylum tricornutum CCAP 1055/1 chromosome 8, whole genome shotgun sequence

CDKA1 [Solanum lycopersicum]

See more...

## 2.5 เมื่อคลิกที่ GenBank จะปรากฏข้อมูลจำนวนคู่เบสของยีน

NIH National Library of Medicine  
National Center for Biotechnology Information

Nucleotide

Search

Advanced

Help

GenBank

Phaeodactylum tricornutum CCAP 1055/1 chromosome 8, whole genome shotgun sequence

NCBI Reference Sequence: NC\_011676.1

[FASTA](#) [Graphics](#)

Go to:

LOCUS NC\_011676 3774 bp DNA linear CON 16-JUN-2017

DEFINITION Phaeodactylum tricornutum CCAP 1055/1 chromosome 8, whole genome shotgun sequence.

ACCESSION NC\_011676 REGION: 494840..498613

VERSION NC\_011676.1

Change region shown

Whole sequence (abbreviated view)

Selected region

from: 494840 to: 498613

Update View

Customize view

Analyze this sequence

Run BLAST

## 2.6 เมื่อคลิกที่ XP\_002180059.1 จะปรากฏข้อมูลจำนวนกรดอะมิโน จากนั้นสามารถกด Run Blast เพื่อวิเคราะห์ข้อมูลพันธุกรรมเปรียบเทียบความใกล้เคียงของลำดับกรดอะมิโนกับสิ่งมีชีวิตอื่น

NIH National Library of Medicine  
National Center for Biotechnology Information

Protein

Search

Advanced

Help

GenPept

predicted protein [Phaeodactylum tricornutum CCAP 1055/1]

NCBI Reference Sequence: XP\_002180059.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to:

LOCUS XP\_002180059 298 aa linear PLN 16-JUN-2017

DEFINITION predicted protein [Phaeodactylum tricornutum CCAP 1055/1].

ACCESSION XP\_002180059

VERSION XP\_002180059.1

DBLINK BioProject: [PRJNA33251](#)

BioSample: [SAMN02953727](#)

DBSOURCE REFSEQ: accession [XM\\_002180023.1](#)

Change region shown

Customize view

Analyze this sequence

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this database

Show in GenBank

## 2.7 การวิเคราะห์ข้อมูล

- เลือก Database เป็น “Reference. Protein (refseq\_protein)”
- เลือก Exclude model (XM/XP)
- ปรับค่า Max target sequence เป็น 50
- กด BLAST เพื่อวิเคราะห์ข้อมูล

BLAST® » blastp suite

Standard Protein BLAST

Enter Query Sequence

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

XP\_002180059.1

Query subrange

From

To

Or, upload file  [เลือกไฟล์](#) [ไม่ได้อัปโหลดไฟล์](#)

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database ☒ Standard databases (nr etc.): ☐ Experimental databases

[+](#) Reference proteins (refseq\_protein) [?](#) **1**

Organism

Optional ☐ exclude [Add organism](#)

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

Exclude ☒ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences **2**

Program Selection

Algorithm ☒ 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 refseq\\_protein using Blastp \(protein-protein BLAST\)](#)

[Show results in a new window](#)

Note: Parameter values that differ from the default are highlighted in yellow and marked with [+](#) sign

Algorithm parameters

General Parameters

Max target sequences [+](#) 50 **3**

Select the maximum number of aligned sequences to display

Short queries ☒ Automatically adjust parameters for short input sequences

Expect threshold  0.05

Word size  5

Max matches in a query range  0

2.8 เมื่อ Blast เสร็จเรียบร้อยแล้ว จะปรากฏหน้าจอดังภาพ จากนั้นเลื่อนลงมาที่ Description และเลือกสิ่งมีชีวิตที่ต้องการเปรียบเทียบ จากนั้นคลิกที่ “Multiple alignment”

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments

Download Select columns Show 50

☐ select all 8 sequences selected

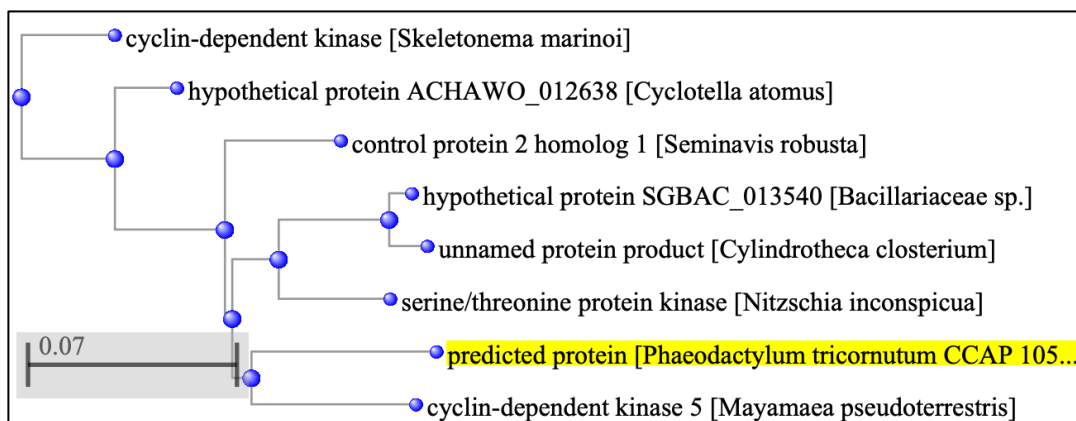
[GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) [MSA Viewer](#)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> predicted protein [Phaeodactylum tricornutum CCAP 1055/1]	<a href="#">Phaeodactylum tricornutum CCAP 1055/1</a>	593	593	100%	0.0	100.00%	290	<a href="#">EEC48250.1</a>
<input checked="" type="checkbox"/> cyclin-dependent kinase 5 [Mayamaea pseudoterrestris]	<a href="#">Mayamaea pseudoterrestris</a>	539	539	99%	0.0	89.58%	295	<a href="#">GKY93737.1</a>
<input checked="" type="checkbox"/> MAG: hypothetical protein SGBAC_013540 [Bacillariaceae sp.]	<a href="#">Bacillariaceae sp.</a>	531	531	99%	0.0	88.54%	296	<a href="#">KAL3926271.1</a>
<input checked="" type="checkbox"/> unnamed protein product [Cylindrotheca closterium]	<a href="#">Cylindrotheca closterium</a>	529	529	99%	0.0	88.19%	296	<a href="#">CAJ1941722.1</a>
<input checked="" type="checkbox"/> control protein 2 homolog 1 [Seminavis robusta]	<a href="#">Seminavis robusta</a>	523	523	97%	0.0	90.39%	289	<a href="#">CAB9526496.1</a>
<input checked="" type="checkbox"/> serine/threonine protein kinase [Nitzschia inconspicua]	<a href="#">Nitzschia inconspicua</a>	521	521	99%	0.0	89.58%	297	<a href="#">KAG7352640.1</a>
<input checked="" type="checkbox"/> cyclin-dependent kinase [Skeletonema marinoi]	<a href="#">Skeletonema marinoi</a>	517	517	99%	0.0	86.06%	310	<a href="#">KAK1732962.1</a>
<input checked="" type="checkbox"/> hypothetical protein ACHAWO_012638 [Cyclotella atomus]	<a href="#">Cyclotella atomus</a>	516	516	99%	0.0	86.81%	289	<a href="#">KAL3772240.1</a>

## 2.9 เมื่อปรากฏหน้าต่างดังภาพ ให้คลิกที่ “Phylogenetic Tree” เพื่อวิเคราะห์ข้อมูลเปรียบเทียบ

## 2.10 ได้ผลการทดลองดังภาพ จากนั้นคลิก “Tools” เพื่อดาวน์โหลดภาพที่สมบูรณ์

## 3. ผลการทดลอง



ภาพแสดงผลการเปรียบเทียบกรดอะมิโน 8 ชนิดของไดอะตอมที่แตกต่างกัน 8 ชนิด

#### 4. ภาคผนวก

*CDKA1* Genomic Reference assembly : NC\_011676.1

*CDKA1* Protein Reference assembly : XP\_002180059.1

ลำดับกรดอะมิโน

>XP\_002180059.1 predicted protein [Phaeodactylum tricornutum CCAP 1055/1]

MERYQRMEKIGEGTYGVVYKAKDRVTGEIIALKKIRLEAEDEGIPSTAIRESLLKELQHPNIVRLYDVV  
HTERKLTLVFEFLDQDLKKYLDICDAGLELPILKSFLYQLLTGVAYCHHHRVLHRDLKPPNLLINREGNL  
KLADFGLARAFGIPVRSYTHEVWTLWYRSPDVLMSRKYSTPVDIWSVGCIFAEMANGRPLVAGTSEADQ  
LDRIFRLLGTPKLEDYPTINELPEYYPDMPPYPPRGGLSALVPRLNPIGIDLLSRMLQYDPARRITAQA  
ALEHEYFQAA