

## รายงานรายวิชา SC187002 Practical skills in Biology (PSB)

หัวข้อ NCBI database and BLAST

ตอนที่ 1 : ศึกษาความสัมพันธ์เชิงวิวัฒนาการของยีนที่นักศึกษามีความสนใจใช้เครื่องมือ และข้อมูลในฐานข้อมูล NCBI เท่านั้น

ข้อมูลผู้ทำการทดลอง นางสาวสาธิตา แสงพระจันทร์ รหัสนักศึกษา 675020077-0

### ที่มาและความสำคัญของยีนที่สนใจ

ยีน EGFR (Epidermal Growth Factor Receptor) คือยีนชนิดหนึ่งที่อยู่บนเยื่อหุ้มเซลล์ ซึ่งมีบทบาทสำคัญในการควบคุมการเจริญเติบโต การแบ่งเซลล์ และการอยู่รอดของเซลล์ การกลายพันธุ์หรือการแสดงออกมากเกินไปของ EGFR เกี่ยวข้องกับมะเร็งหลายชนิด เช่น มะเร็งปอดชนิด NSCLC (non-small cell lung cancer), มะเร็งสมอง (glioblastoma), มะเร็งลำไส้ใหญ่ และมะเร็งเต้านมบางชนิด

ยาต้านมะเร็งหลายชนิดที่พุ่งเป้าไปที่ EGFR เช่น gefitinib, erlotinib, Osimertinib, cetuximab และ panitumumab

ยีน EGFR มีจำนวนลำดับเบสทั้งหมด 3848 bp และมีจำนวนลำดับกรดอะมิโน 1091 aa

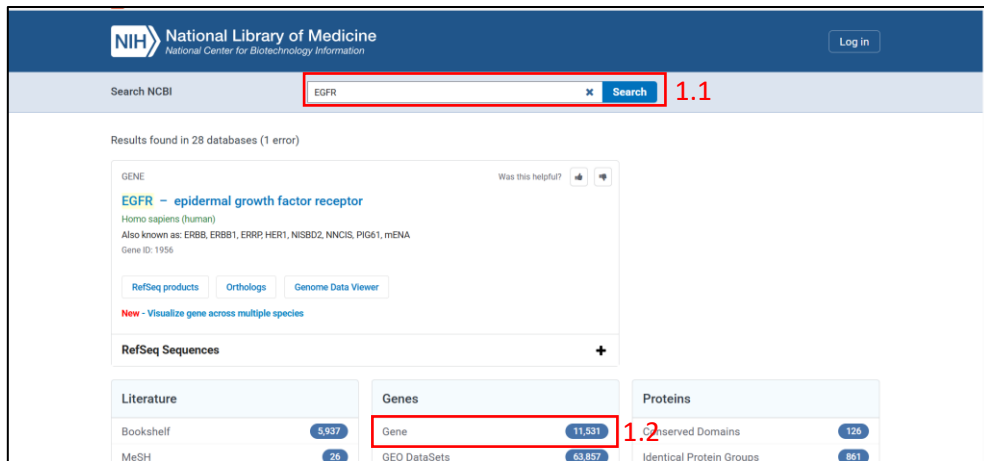
The screenshot displays the NCBI Protein database entry for the epidermal growth factor receptor isoform e precursor [Homo sapiens]. The page includes the following sections:

- FASTA:** The protein sequence is shown in a monospaced font.
- Protein 3D Structure:** A 3D ribbon diagram of the protein structure is displayed, with a caption indicating it is the crystal structure of EGFR 696-1022 WT in complex with AZD9291 prepared by PDB: 6JXT.
- Articles about the EGFR gene:** A list of recent articles is provided, including "Hypertrophic cardiomyopathy-associated mutations drive stromal activation" (Sci Adv 2024) and "EGFR mutations and abnormal trafficking in cancers" (Mol Biol Rep 2024).

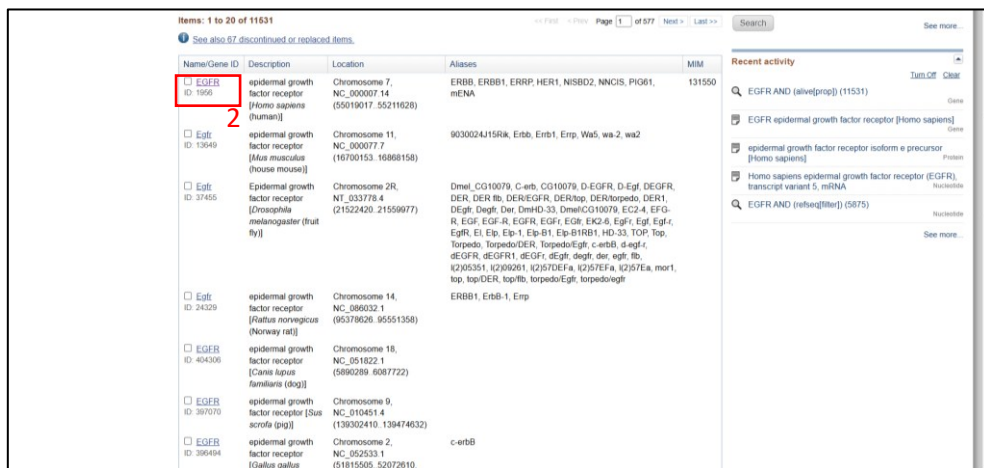
## ขั้นตอนการดำเนินงาน

## การค้นหาลำดับกรดอะมิโนของยีนที่สนใจในสิ่งมีชีวิตที่สนใจ

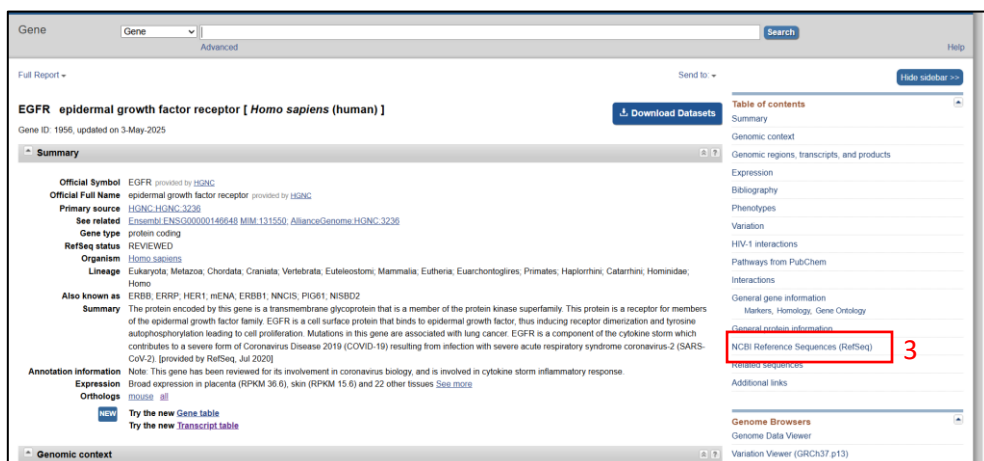
## 1. ใส่ชื่อยีน EGFR ในช่อง Search แล้วเลือก Gene



## 2. กดเลือก EGFR ID:1956



## 3. หน้าจอจะแสดง database ของยีน EGFR กดเลือก NCBI Reference Sequences (RefSeq) เพื่อดูข้อมูล Protein Sequences



## 4. กดเลือก NP\_001333826.1

NCBI Reference Sequences (RefSeq)

Try the new Transcript table

RefSeqs maintained independently of Annotated Genomes

These reference sequences exist independently of genome builds. [Explain](#)

**Genomic**

1. **NO\_00728.3 RefSeqGene**

Range: 5001,193307  
Download: [GenBank](#), [FASTA](#), [Sequence Viewer](#), [Graphics](#), [LRG\\_304](#)

**mRNA and Protein(s)**

1. **NM\_001346897.2 -- NP\_001333826.1 epidermal growth factor receptor isoform e precursor**

Status: REVIEWED

Description: Transcript Variant. This variant (5) lacks an in-frame exon in the 5' coding region and its 3' terminal exon extends past a splice site that is used in variant 1. The encoded isoform (e) is shorter and has a distinct C-terminus compared to isoform a.

Source sequence(s): [AC009877](#), [AC073324](#), [BC094781](#)  
Consensus CDS: [CCD887507.1](#)  
UniProtKB/TrEMBL: [F2Y1SGT](#), [Q56HJ8](#)  
Related: [ENST00000455089.1](#), [ENST00000455089.5](#)

2. **NM\_001346899.2 -- NP\_001333827.1 epidermal growth factor receptor isoform f precursor**

Status: REVIEWED

Description: Transcript Variant. This variant (6) has a 3' terminal exon that extends past a splice site that is used in variant 1. The encoded isoform (f) has a shorter and distinct C-terminus compared to isoform a.

Source sequence(s): [AC009877](#), [AC073324](#), [BC094781](#), [GU255993](#), [HG912715](#)  
UniProtKB/TrEMBL: [ASC880](#), [ETBSV0](#)

3. **NM\_001346899.2 -- NP\_001333828.1 epidermal growth factor receptor isoform g precursor**

Status: REVIEWED

## 5. ข้อมูลของยีน EGFR แสดงจำนวนลำดับกรดอะมิโน 1091 aa กดเลือก FASTA

Gene: **epidermal growth factor receptor isoform e precursor [Homo sapiens]**

NCBI Reference Sequence: [NP\\_001333826.1](#)

Identical Protein: [FASTA](#)

Go to: [GenBank](#)

LOCUS: **NP\_001333826**  
DEFINITION: **epidermal growth factor receptor isoform e precursor [Homo sapiens]**  
ACCESSION: **NP\_001333826**  
VERSION: **NP\_001333826.1**  
OBSSOURCE: **REFSEQ**; accession **NM\_001346897.2**  
KEYWORDS: **RefSeq**  
SOURCE: **Homo sapiens (human)**  
ORGANISM: **Homo sapiens**  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhina; Catarrhini; Hominidae; Homo.  
1 (residues 1 to 1091)  
AUTHORS: **Huang, X., Men, C., Shan, S., Yang, J., Zhang, S., Ji, X., Li, C. and Wang, Y.**  
TITLE: **EGFR upregulates miRNA subset to inhibit CYBB and cause DOP resistance in gastric cancer**  
JOURNAL: **Gene** 933, 149005 (2025)  
PUBMED: **32612328**  
REMARK: **GeneID: EGFR upregulates miRNA subset to inhibit CYBB and cause DOP resistance in gastric cancer.**  
REFERENCE: **2 (residues 1 to 1091)**  
AUTHORS: **Liu, J., Fang, R., Ma, R., Long, Y., Luo, H., Chen, L. and Wen, Y.**  
TITLE: **Amphiregulin promotes activated regulatory T cell-suppressive function via the AREG/EGFR pathway in laryngeal squamous cell carcinoma**  
JOURNAL: **Head Face Med** 20 (1), 62 (2024)  
PUBMED: **32654888**  
REMARK: **GeneID: Amphiregulin promotes activated regulatory T cell-suppressive function via the AREG/EGFR pathway in laryngeal squamous cell carcinoma.**  
Publication Status: Online-Only  
REFERENCE: **3 (residues 1 to 1091)**

Protein 3D Structure

Crystal structure of EGFR 696-1022 WT in complex with AZD5363 prepared by PDB: 6JXT  
Source: Homo sapiens  
Method: X-ray Diffraction  
Resolution: 2.307 Å

See all 174 structures...

Articles about the EGFR gene

Hyperphosphorylation-associated mutations drive stromal activation [Sci Adv. 2024]

EGFR mutations and abnormal trafficking in cancers [Mol Biol Rep. 2024]

EGFR mutations induce the suppression of CD8(+) T cell and anti-PD-1 [J Transl Med. 2024]

See all...

## 6. จะได้ลำดับกรดอะมิโนเพื่อนำมาศึกษาเปรียบเทียบความสัมพันธ์ทางวิวัฒนาการของยีน โดยใช้โปรแกรม blastp

Protein: **epidermal growth factor receptor isoform e precursor [Homo sapiens]**

NCBI Reference Sequence: [NP\\_001333826.1](#)

FASTA: [FASTA](#)

Send to: [Change region shown](#)

Analyze this sequence

Run BLAST

Identify Conserved Domains

Show in Genome Data Viewer

Protein 3D Structure

Crystal structure of EGFR 696-1022 WT in complex with AZD5363 prepared by PDB: 6JXT  
Source: Homo sapiens  
Method: X-ray Diffraction  
Resolution: 2.307 Å

See all 174 structures...

Articles about the EGFR gene

Hyperphosphorylation-associated mutations drive stromal activation [Sci Adv. 2024]

EGFR mutations and abnormal trafficking in cancers [Mol Biol Rep. 2024]

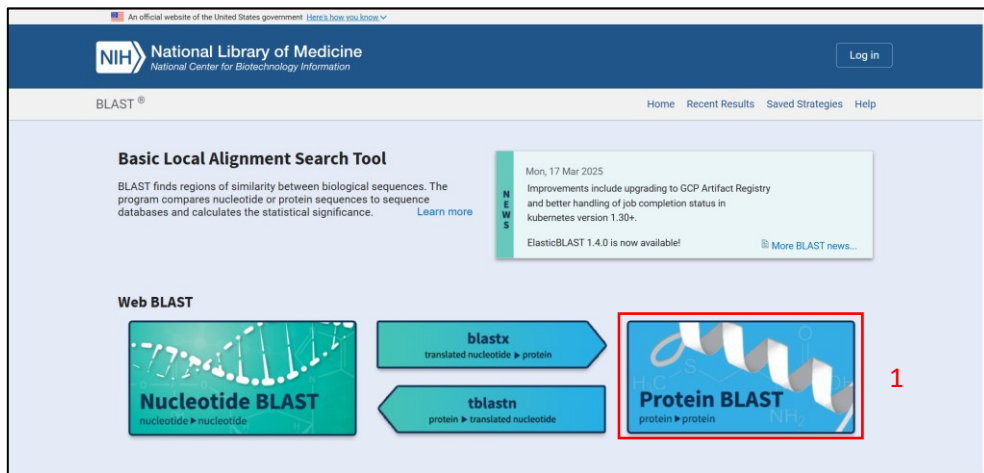
EGFR mutations induce the suppression of CD8(+) T cell and anti-PD-1 [J Transl Med. 2024]

See all...

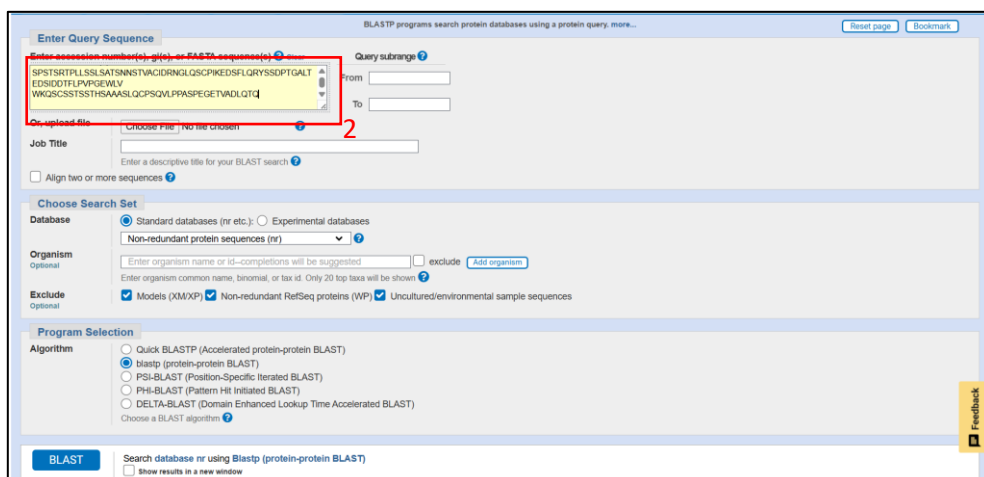
ลำดับกรดอะมิโน

## การใช้ blastp โดยเลือก database = RefSeq protein และ max target sequences = 50

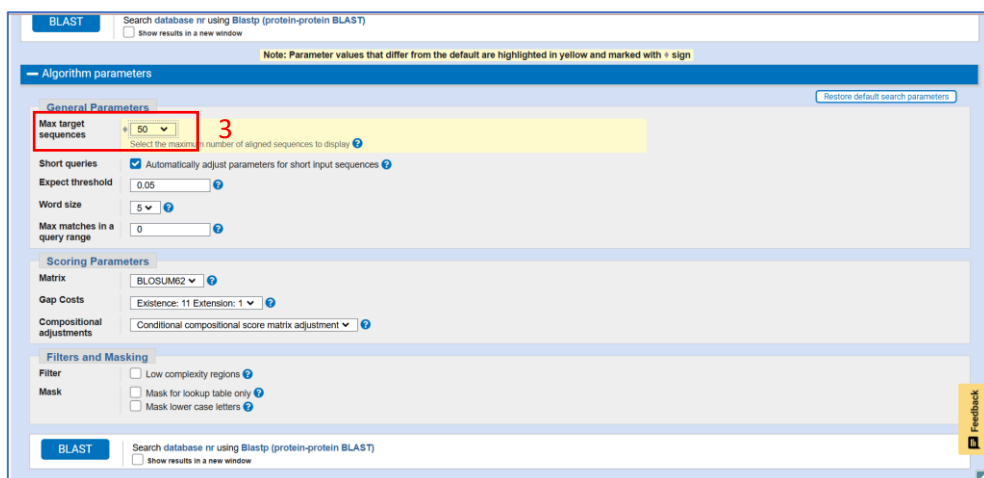
### 1. กดเลือกโปรแกรม blastp



### 2. นำลำดับกรดอะมิโนเติมลงในช่อง Enter accession number(s) และกด blast



### 3. เลือก Algorithm parameters และตั้งค่า Max target sequences = 50



## 4. กดเลือก BLAST

## 5. หน้าจอจะแสดงข้อมูล Protein Sequence ให้เลือกข้อมูลยื่นเปรียบเทียบกับกรดอะมิโนของสิ่งมีชีวิต 10 ชนิด

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> epidermal growth factor receptor isoform a precursor (Homo sapiens)	Homo sapiens	2267	2267	100%	0.0	100.00%	1091	NP_001333626.1
<input type="checkbox"/> epidermal growth factor receptor (Homo sapiens)	Homo sapiens	2265	2265	100%	0.0	99.91%	1091	KAA4013877.1
<input type="checkbox"/> EGFR, partial (synthetic construct)	synthetic construct	2263	2263	100%	0.0	99.82%	1091	AIC61960.1
<input checked="" type="checkbox"/> LOW QUALITY PROTEIN, EGFR isoform 4 (Pan troglodytes)	Pan troglodytes	2259	2259	100%	0.0	99.54%	1091	ENIG26286.1
<input type="checkbox"/> LOW QUALITY PROTEIN, EGFR isoform 1 (Pongo abelii)	Pongo abelii	2246	2246	100%	0.0	98.90%	1091	PNJ39098.1
<input type="checkbox"/> epidermal growth factor receptor isoform 1 precursor (Homo sapiens)	Homo sapiens	2242	2242	100%	0.0	95.95%	1136	NP_001333627.1
<input type="checkbox"/> epidermal growth factor receptor (Homo sapiens)	Homo sapiens	2174	2174	96%	0.0	100.00%	1167	A0275481.1
<input type="checkbox"/> epidermal growth factor receptor isoform g precursor (Homo sapiens)	Homo sapiens	2174	2174	96%	0.0	100.00%	1091	NP_001333628.1
<input type="checkbox"/> epidermal growth factor receptor (Homo sapiens)	Homo sapiens	2173	2173	96%	0.0	99.90%	1165	KAA4013879.1
<input type="checkbox"/> EGFR isoform 3 (Pan troglodytes)	Pan troglodytes	2172	2172	96%	0.0	99.71%	1165	ENIG26285.1
<input checked="" type="checkbox"/> epidermal growth factor receptor isoform X2 (Theropithecus gelada)	Theropithecus gelada	2162	2162	96%	0.0	99.23%	1165	XP_025233945.1
<input type="checkbox"/> EGFR isoform 2 (Pongo abelii)	Pongo abelii	2161	2161	96%	0.0	99.33%	1165	PNJ39099.1
<input checked="" type="checkbox"/> epidermal growth factor receptor isoform X2 (Ptilinopus leucostictus)	Ptilinopus leucostictus	2160	2160	96%	0.0	99.23%	1165	XP_023087808.1
<input checked="" type="checkbox"/> epidermal growth factor receptor isoform X2 (Symphalangus syndactylus)	Symphalangus syndactylus	2152	2152	96%	0.0	99.04%	1164	XP_055149091.1
<input type="checkbox"/> epidermal growth factor receptor isoform a precursor (Homo sapiens)	Homo sapiens	2150	2150	96%	0.0	95.78%	1210	NP_005219.2

6. กดเลือก Distance tree of results เพื่อสร้าง phylogenetic tree

**Sequences producing significant alignments**

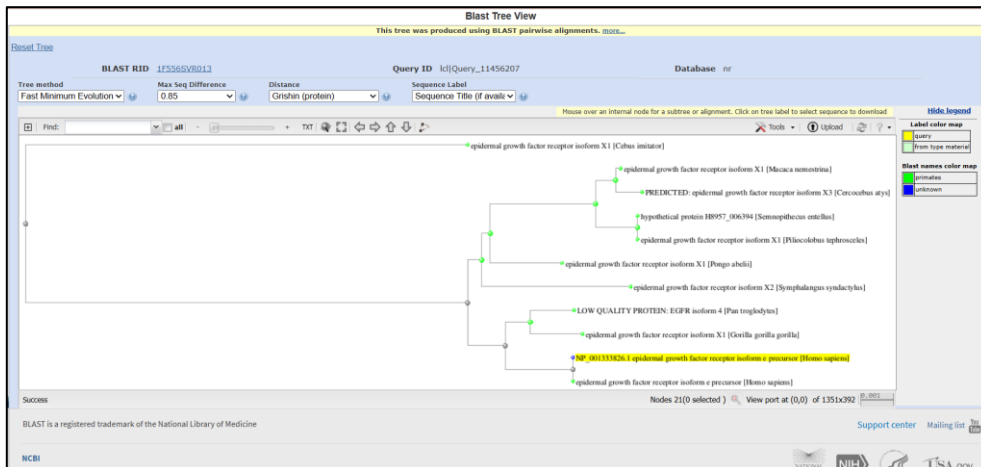
Download ▾ Select columns ▾ Show 50 ▾

☐ select all 12 sequences selected

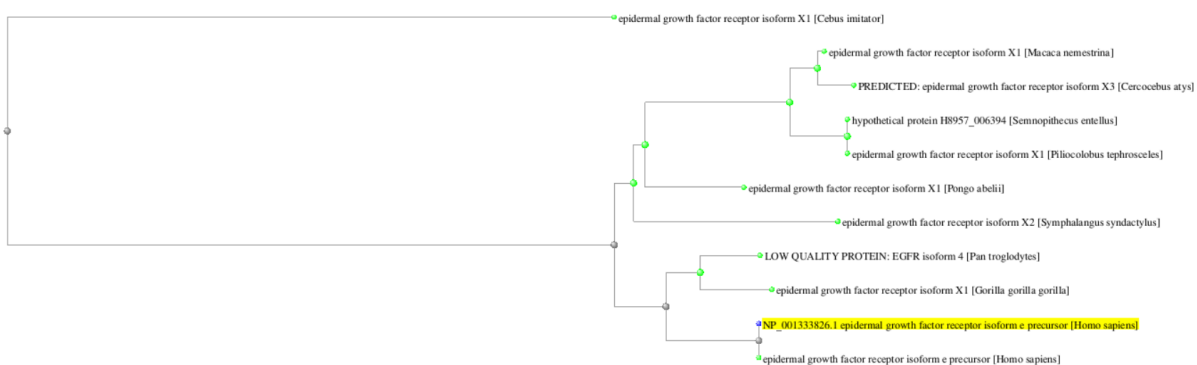
GenPept **6** Graphics **Distance view of results** **Multiple alignment** **MSA Viewer**

	Description ▾	Scientific Name ▾	Max Score ▾	Total Score ▾	Query Cover ▾	E value ▾	Per. Ident ▾	Acc. Len ▾	Accession
<input checked="" type="checkbox"/>	epidermal growth factor receptor isoform 1 precursor [Homo sapiens]	Homo sapiens	2267	2267	100.0%	0.0	100.00%	1091	NP_001333826.1
<input type="checkbox"/>	epidermal growth factor receptor [Homo sapiens]	Homo sapiens	2265	2265	100%	0.0	99.91%	1091	KAA4013877.1
<input type="checkbox"/>	EGFR_partial [synthetic construct]	synthetic co...	2263	2263	100%	0.0	99.82%	1091	AIC61360.1
<input checked="" type="checkbox"/>	LOW QUALITY PROTEIN: EGFR isoform 4 [Pan troglodytes]	Pan troglody...	2259	2259	100%	0.0	99.54%	1091	PN26286.1
<input checked="" type="checkbox"/>	LOW QUALITY PROTEIN: EGFR isoform 1 [Pongo abelii]	Pongo abelii	2246	2246	100%	0.0	98.90%	1091	PNJ39609.1
<input type="checkbox"/>	epidermal growth factor receptor isoform 1 precursor [Homo sapiens]	Homo sapiens	2242	2242	100%	0.0	95.95%	1136	NP_001333827.1
<input type="checkbox"/>	epidermal growth factor receptor [Homo sapiens]	Homo sapiens	2174	2174	96%	0.0	100.00%	1167	A0Z75461.1
<input type="checkbox"/>	epidermal growth factor receptor isoform 1 precursor [Homo sapiens]	Homo sapiens	2174	2174	96%	0.0	100.00%	1165	NP_001333828.1
<input type="checkbox"/>	epidermal growth factor receptor [Homo sapiens]	Homo sapiens	2173	2173	96%	0.0	99.90%	1165	KAA4013879.1
<input type="checkbox"/>	EGFR isoform 3 [Pan troglodytes]	Pan troglody...	2172	2172	96%	0.0	99.71%	1165	PN26285.1
<input type="checkbox"/>	epidermal growth factor receptor isoform X2 [Theropithecus gelada]	Theropithec...	2162	2162	96%	0.0	99.23%	1165	XP_025233945.1
<input type="checkbox"/>	EGFR isoform 2 [Pongo abelii]	Pongo abelii	2161	2161	96%	0.0	99.33%	1165	PNJ39609.0
<input type="checkbox"/>	epidermal growth factor receptor isoform X2 [Ptilocobus leithroscotes]	Ptilocobus...	2160	2160	96%	0.0	99.23%	1165	XP_023067808.1
<input checked="" type="checkbox"/>	epidermal growth factor receptor isoform X2 [Symphalangus syndactylus]	Symphalang...	2152	2152	96%	0.0	99.04%	1164	XP_055149091.1
<input type="checkbox"/>	epidermal growth factor receptor isoform 1 precursor [Homo sapiens]	Homo sapiens	2150	2150	96%	0.0	95.78%	1210	NP_005219.2
<input type="checkbox"/>	epidermal growth factor receptor, partial [synthetic construct]	synthetic co...	2149	2149	96%	0.0	95.78%	1211	AA441033.1
<input type="checkbox"/>	epidermal growth factor receptor, partial [synthetic construct]	synthetic co...	2149	2149	96%	0.0	95.69%	1210	BA446646.1
<input type="checkbox"/>	Chain A: Epidermal growth factor receptor [Homo sapiens]	Homo sapiens	2149	2149	96%	0.0	95.69%	1210	ZSYD_A
<input type="checkbox"/>	epidermal growth factor receptor [Homo sapiens]	Homo sapiens	2149	2149	96%	0.0	95.69%	1210	KAA4013876.1

7. ได้ phylogenetic tree รูปแบบ Rectangle Cladogram



ผลการทดลอง



## ภาคผนวก

NCBI accession number : EGFR ID: 1956 : NP\_001333826.1

ลำดับกรดอะมิโน :

>NP\_001333826.1 epidermal growth factor receptor isoform e precursor [Homo sapiens]

MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTLGLTFEDHFLSLQRMFNNCEVVLGNLEITYV  
QRNYDLSFLKTIQEVAGYVLIALNTVERIPLLENLQIIRGNMYEENSALAVLSNYDANKTGLKELPMRNL  
QGQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGCTGPRESDCLVCRKFR  
DEATCKDTCPPMLLYNPPTYQMDVNPPEGKYSFGATCVKKCPRNYVVTDHGSCVRACGADSYEMEEDGVRK  
CKKCEGPCRKVCNGIGIGIEFKDLSINATNIKHFKNCTSIISGDLHILPVAFRGDSFTHTPPLDPQEELDIL  
KTVKEITGFLLIQAWPENRTDLHAFENLEIIRGRTKQHGGFSLAVVSLNITSLGLRSLKEISDGDVLIISG  
NKNLCYANTINWKKLFGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGREC  
VDKCNLLEGEPPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVMGENNT  
LVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPNGPKIPSIATGMVGALLLLLVVALGIGLFMRRRHIVR  
KRTLRLQLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIK  
ELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLL  
NWCVQIAKGMNYLEDRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEKEYHAEGGKVPIKWMALLES  
ILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDA  
DSRPKFRELIIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVDDADEYLIPQQGFFS  
SPSTSRTPLLSSLATSNNSTVACIDRNLQSCPIKEDSFLQRYSSDPTGALTEDSIDDTFLPVPGEWL  
WKQSCSSTSSTHSAASLQCPSQVLPPASPEGETVADLQTQ