

# 生物資訊學-第一次作業

## 一、前往 *NCBI* 查詢基因、解釋原由

#uuid1

### 尋找基因

我利用關鍵詞的搜尋法來尋找

```
1 ACTB[Gene Name] AND refseq[Filter] AND biomol_mrna[Properties]
```

並且有確實的找到以下資訊，可以看到在 *AnnotationProvider* 這邊，顯示出該資料是來自 *RefSeq*，而 *NCBIRefSeq* (*Reference Sequence Database*) 是美國國家生物技術資訊中心 (*NCBI*) 維護的一個基因組、轉錄組和蛋白質序列的標準數據庫。它提供了經過整理和標準化的基因組數據，並進行基因註釋，以確保數據的準確性和一致性。

### PREDICTED: Pithys albifrons albifrons actin beta (ACTB), mRNA

NCBI Reference Sequence: XM\_071570841.1

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

Go to:

LOCUS XM\_071570841 1848 bp mRNA linear VRT 26-FEB-2025  
DEFINITION PREDICTED: Pithys albifrons albifrons actin beta (ACTB), mRNA.  
ACCESSION XM\_071570841  
VERSION XM\_071570841.1  
DBLINK BioProject: [PRJNA1227525](#)  
KEYWORDS RefSeq.  
SOURCE Pithys albifrons albifrons  
ORGANISM [Pithys albifrons albifrons](#)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archelosauria; Archosauria; Dinosauria; Saurischia; Theropoda;  
Coelurosauria; Aves; Neognathae; Neoaves; Telluraves; Australaves;  
Passeriformes; Thamnophilidae; Pithys.  
COMMENT MODEL [REFSEQ](#): This record is predicted by automated computational  
analysis. This record is derived from a genomic sequence  
([NC\\_092473](#)) annotated using gene prediction method: Gnomon.  
Also see:  
[Documentation](#) of NCBI's Annotation Process  
  
##Genome-Annotation-Data-START##  
Annotation Provider :: NCBI RefSeq  
Annotation Status :: Full annotation  
Annotation Name :: [GCF\\_047495875.1-RS\\_2025\\_02](#)  
Annotation Pipeline :: NCBI eukaryotic genome annotation  
pipeline  
Annotation Software Version :: [10.3](#)  
Annotation Method :: Gnomon; cmsearch; tRNAscan-SE  
Features Annotated :: Gene; mRNA; CDS; ncRNA  
Annotation Date :: 02/25/2025  
##Genome-Annotation-Data-END##

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequences

Find in this Sequence

Show in Genome

Reference sequence

RefSeq protein product  
See the reference  
cytoplasmic 1 (Xenopus laevis)

More about the

ACTB gene

Related information

BioProject

Protein

Taxonomy

Annotated Genomes

利用 *Ctrl*<sup>^</sup> / *Cmd*<sup>⌘</sup> + *f* 標注，並標注 *GGTACC* 序列。

為什麼要尋找GGTACC序列呢？

因為 *GGTACC* 是限制性內切酶 *KpnI* 的識別位點。*KpnI* 是一種來源於 *Klebsiella pneumoniae* (肺炎克雷伯氏菌) 的限制酶，會專一性地識別 *GGTACC* 並在 *GGTAC* | *C* 之間進行切割。

那我尋找的類型為：*Pithys albifrons albifrons* x 的 *ACTB* ( $\beta$ -肌動蛋白) 基因。

```

/db_xref="GeneID:139679266"
/translation="MDDDTAALVVDNGSGMCKAGFAGDDAPRAVFPISIVGRPRHQGVM
VGMGQKDSYVGDEAQSQRGILTLYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHP
VLLTEAPLNPKANREKMTQIMFETNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGVT
HTVPIYEGYALPHAILRLDLAAGRLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCY
VALDFEQEMATASSSSLEKSYELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHE
TTFNSIMKCDVDIRKLDYANTVLSGGTMYPGIADRMQKEITALAPSTMKIKIIAPPE
RKYSVWIGGSILASLSTFQQMWISKQEYDESGPSIVHRKCF"
1848
/gene="ACTB"
/experiment="COORDINATES: polyA evidence [ECO:0006239]"

polyA_site

ORIGIN
1  gggcgggcggg  cgggagccgc  cgcgagctcc  gctgccctcc  ccgcgcgcgc  gccccgctcc
61  cgccgcccgc  tcgcgcccgc  cccggctccg  accgcgcgcg  tccacacaaa  gccagccatg
121 gatgatgata  ttgccgcgct  cgttgttgac  aatggttccg  gtatgtgcaa  ggccggtttt
181 gccggggagc  atgccccccg  tgctgtcttc  ccatccatcg  tggggcgctc  caggcatcag
241 ggtgtgatgg  ttggtatggg  ccagaaagac  agctatgttg  gtgatgaagc  ccagagcaaa
301 agaggatcgc  tgacctgaa  gtacccatt  gaacacgcta  ttgtcaccaa  ctgggatgac
361 atggagaaga  tctggcacca  cactttctac  aatgagctga  gagtagcccc  cgaggagcac
421 cctgtgctgc  ttacagaggc  tccccgaac  cccaaagcca  acagagagaa  gatgacacag
481 atcatgttcg  agaccttcaa  caccacagcc  atgtatgtag  ccatccagcg  tgttgtgtcc
541 ctgtatgcct  ctggtcgta  cactggtatt  gtgatggact  ctggtgatgg  tgttaccac
601 actgtgcccc  tctatgaagg  ctatgccctc  ccccatgcc  tctctcgct  ggtatctggc
661 ggcggtgacc  tgacggacta  cctcatgaag  atcctgacag  agagaggcta  cagcttcacc
721 accacagctg  agagagaaat  tgtgcgtgac  atcaaggaga  agctgtgcta  tgttgcctg
781 gattttgagc  aggagatggc  cacagctgcc  tctagctctt  ccctggagaa  gagctatgag
841 ctccctgatg  gccaggtcat  caccattggc  aatgagaggt  tcaggtgcc  tgaggccctc
901 ttccagccat  ctttcttggg  tatggagtcc  tgtggtatcc  atgaaactac  cttcaactcc
961 atcatgaagt  gtgatgtgga  tatccgtaag  gacctgtatg  ccaacacagt  gctgtctggt
1021 ggtaccacaa  tgtaccctgg  cattgctgac  aggatgcaga  aggagatcac  agccctggca
1081 ccagcacaa  tgaaaatcaa  gatcattgcc  ccactgagc  gcaaatactc  tgtctggatc
1141 ggaggctcca  tcctggcctc  cctgtccacc  ttccagcaga  tgtggatcag  caagcaggag
1201 tatgatgaat  cgggaccctc  cattgtccac  cgcaaatgct  tctaaacgg  actgttacca
1261 acaccacac  ctttgtgacg  aaatgaaacc  cataaatgcg  cataaaacaa  gaagagattg
1321 gcatggcttt  attgttttt  tttttggcgc  ttgactcagg  attaaaaaac  tggaacggtg
1381 aaggtgtcag  cagcagcttt  aaaatgaaac  atgttgagc  gaacgcccc  aaagtcttac
1441 aatgcattcg  aggactttga  ttgtacattt  gtttctttt  ttaatatgca  ttccaaatat
1501 tgttataatg  cattgttaca  ggaagttaact  gcctctgtg  aaggcaacag  ccagctcg
1561 agggagccga  tgccagttac  tgggtttagg  ttataattgc  ttgtctgtaa  attatgtaac
1621 ccaacaagtg  tctttttgta  tctttccgcc  ttaaaaacaa  aacacacttg  atccttttc
1681 tcttcaattt  gtcaagcaag  caggctgtgt  tccccagtg  tagatgggaa  tgaaggcttt
1741 acagcccccc  acaggttcac  agaaggcca  gtatgtggg  gagggagggg  ctacctgtac
1801 actgacttaa  gaccagttca  aataaaagt  cacaccatag  aggcctga
//
```

## 二、下載該 *mRNA* 的 *GenBank*，import 進 *CLC* 並進行 *KpnI* 限制酶的加入

#uuid2

點擊Send to

Send to: ▼

Complete Record

Coding Sequences

Gene Features

Choose Destination

File

Collections

Clipboard

Analysis Tool

Pick Primers

↓

點擊Complete Record to

A

- ☒ Complete Record
- ☐ Coding Sequences
- ☐ Gene Features

### Choose Destination

- ☒ File
- ☐ Clipboard
- ☐ Collections
- ☐ Analysis Tool

Download 1 item.



選取GenBank

A

- ☒ Complete Record
- ☐ Coding Sequences
- ☐ Gene Features

### Choose Destination

- ☒ File
- ☐ Clipboard
- ☐ Collections
- ☐ Analysis Tool

Download 1 item.

Format

GenBank

Show GI ☐

Create File



再來將載入完成的字體 *import* 至 *CLC Sequence Viewer 7*

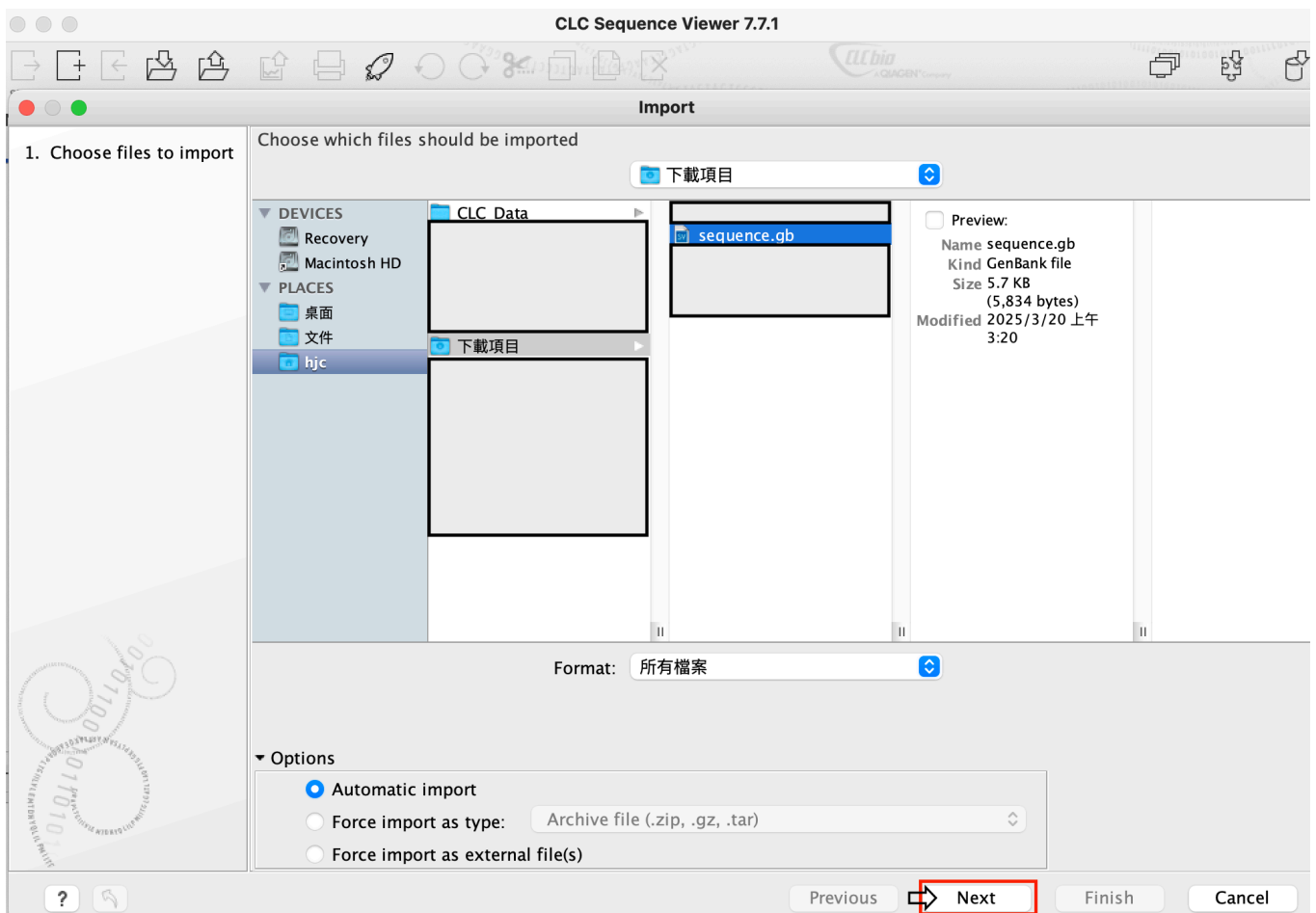
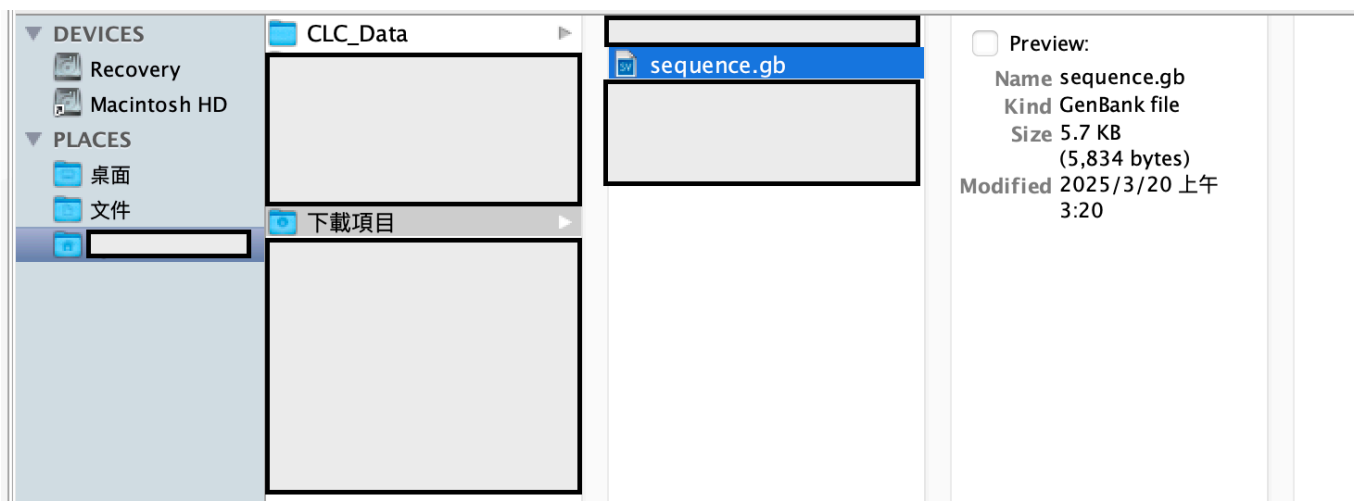


sequence.gb

6 KB GenBank file

今天 凌晨 3:20





點擊 *toolbox* 中的 *Restriction Site Analysis*

### Restriction Site Analysis

1. Select DNA/RNA sequence(s)

Select DNA/RNA sequence(s)  
Navigation Area  

CLC\_Data  
XM\_071570841

Selected elements (1)  

XM\_071570841

?

Previous
Next
Finish
Cancel



### Restriction Site Analysis

1. Select DNA/RNA sequence(s)  
2. Enzymes to be considered in calculation

Enzymes to be considered in calculation  
Enzyme list  
☐ Use existing enzyme list

All enzymes  
Filter: Kpn

Name	Overhang	Methylati...	Popul...
KpnI	3' - gtac	5': N6-...	****
Kpn2I	5' - ccgg	5': 5-m...	*
Kpn378I	3' - gc		*
Kpn2kl	5' - ccngg	5': 5-m...	*
Kpn49kl	5' - aatt		*
Kpn49klI	5' - ccsgg		*

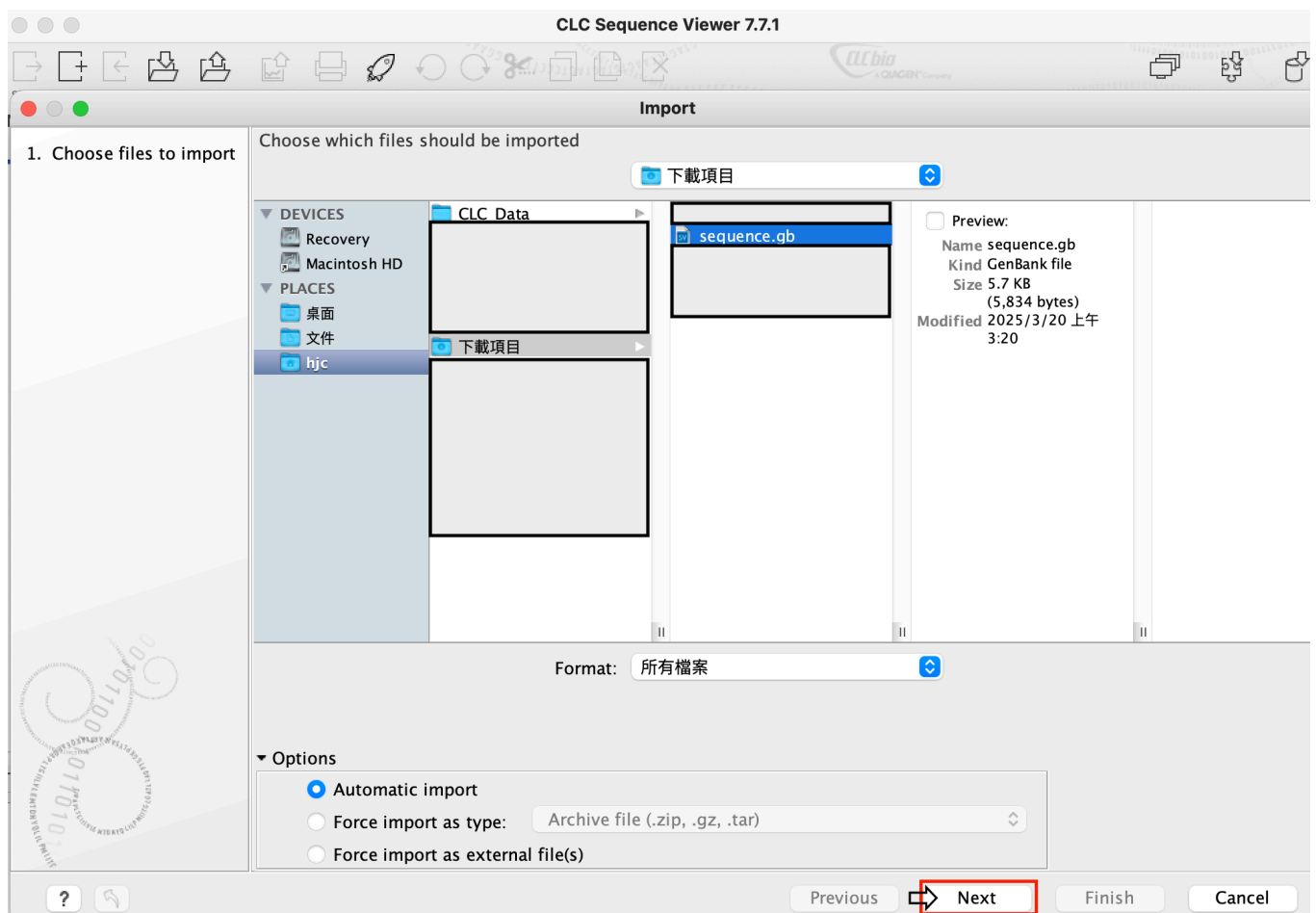
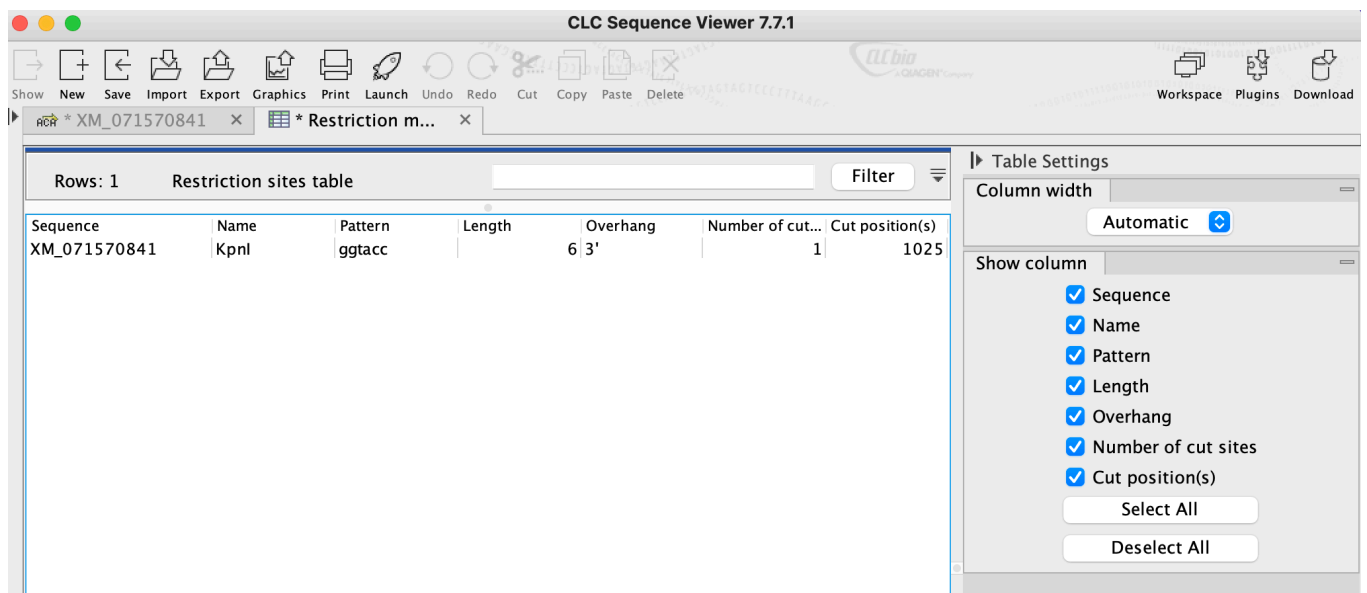
Enzymes to be used  
Filter:

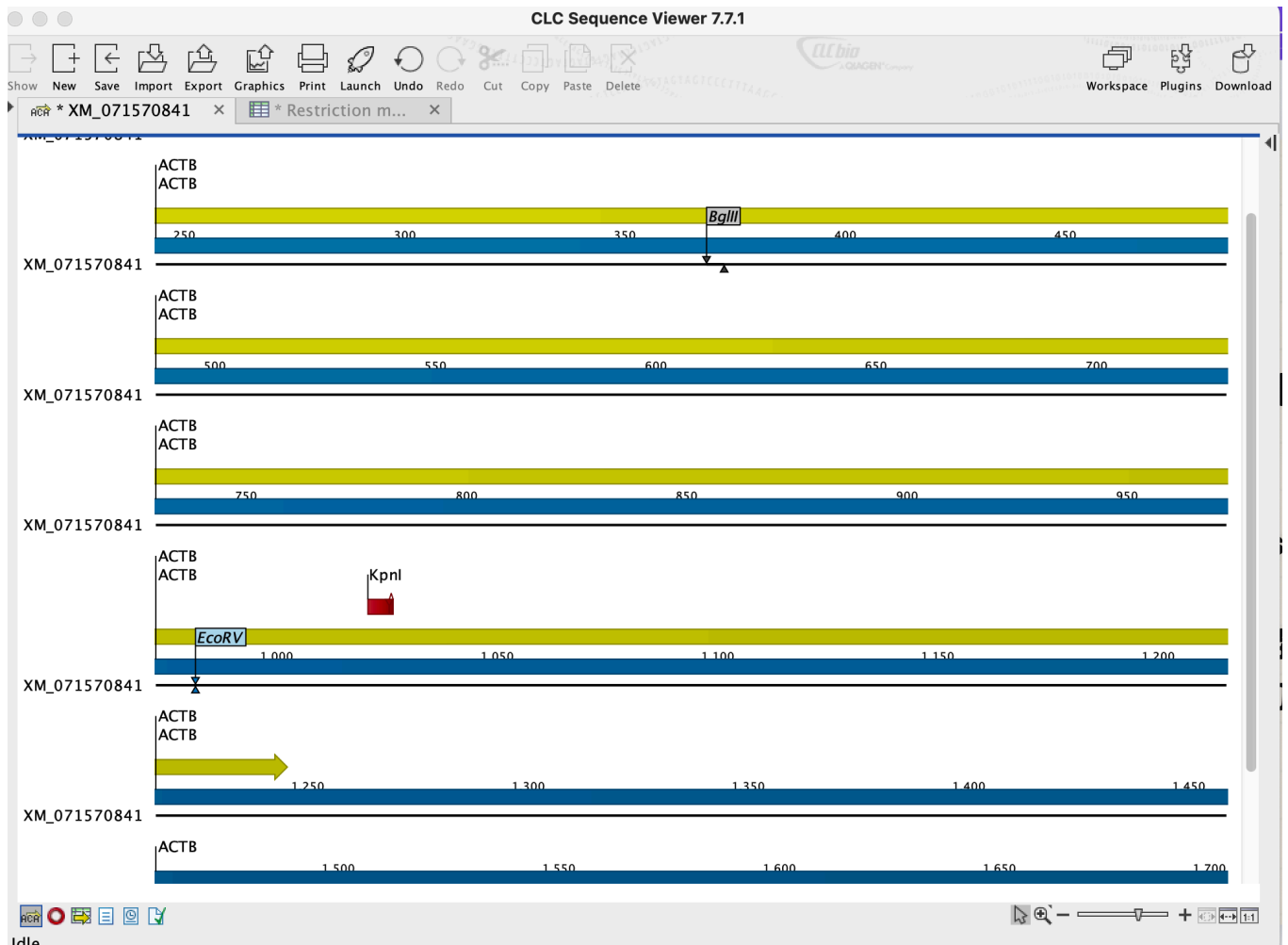
Name	Overhang	Methylat...	Popul...
KpnI	3' - gtac	5': N6-...	****

?

Previous
Next
Finish
Cancel







這是我們的 *KpnI* 限制酶的位置

### 三、下載該 *protein* 的 *Sequence* 並將載好的 *GenePept* 資料 *import* 進 *CLC*

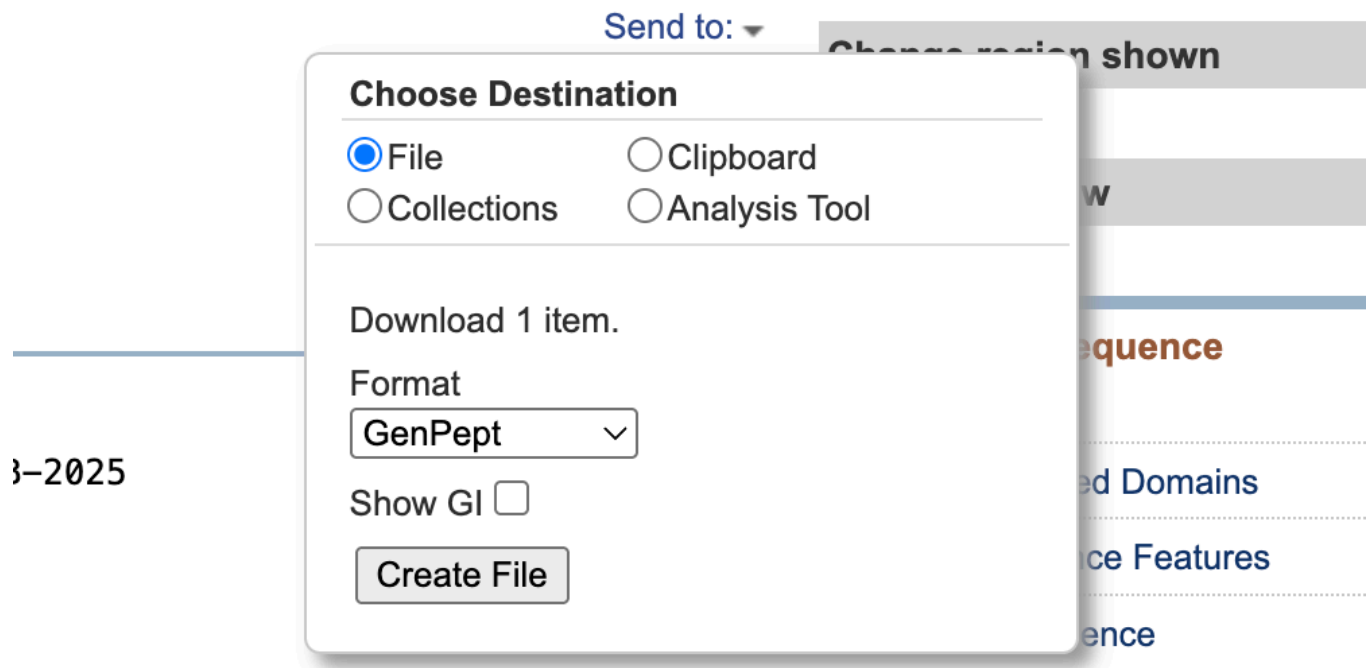
#uuid3

再來，回到 *NCBI* 尋找特徵中的 *Protein\_id*，並將他下載下來

[CDS](#)

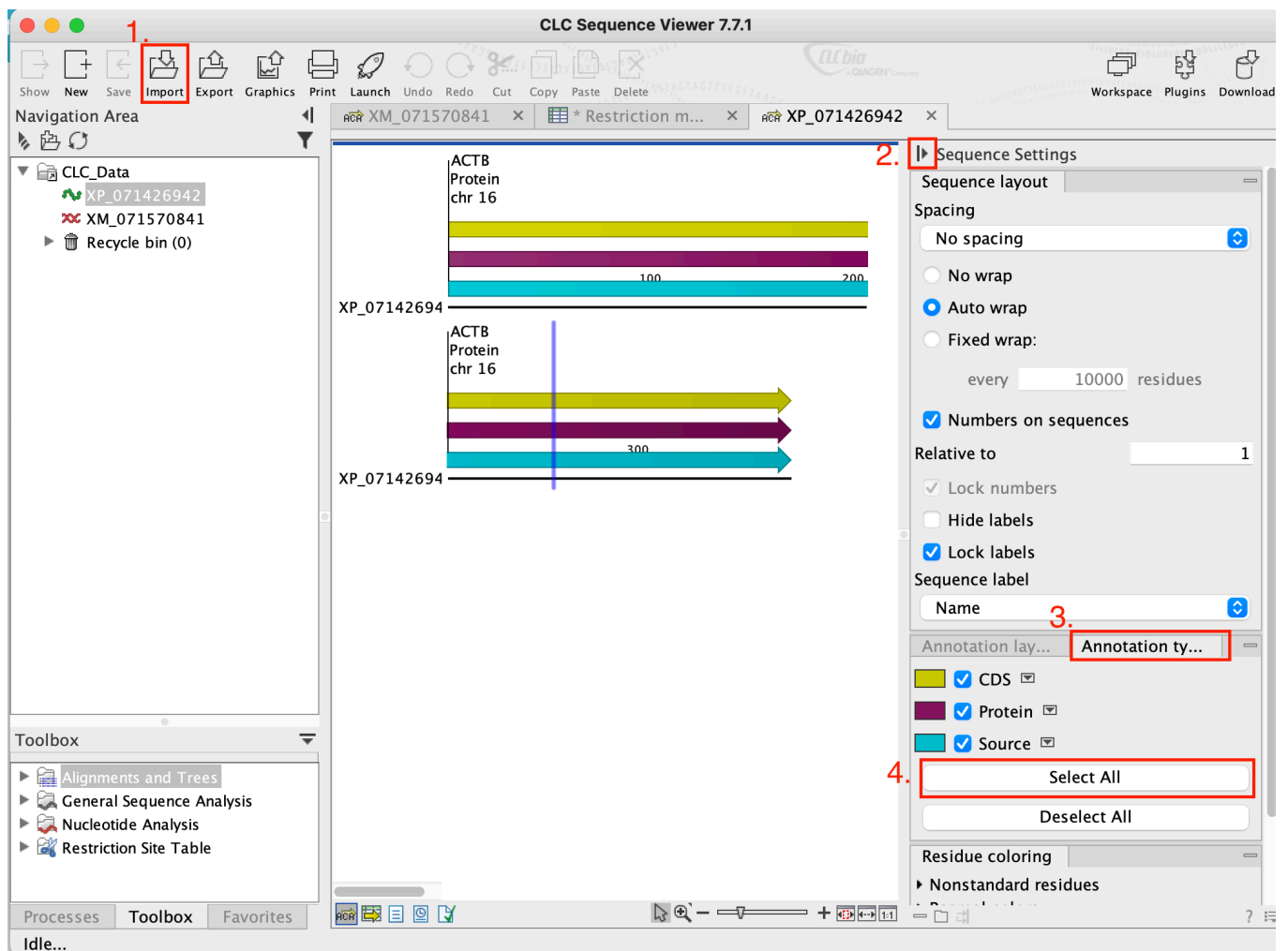
```
118..1245
/gene="ACTB"
/codon_start=1
/product="actin, cytoplasmic 1"
/protein_id="XP_071426942.1"
/db_xref="GeneID:139679266"
/translation="MDD DIAALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVM
VGMGQKDSYVGDEAQSKRGILT LKYP I EHGIVTNWDDMEKIWHHTFYNELRVAPEEHP
VLLTEAPLNPKANREKMTQIMFETFNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGVT
HTVPIYEGYALPHAILRLDLAGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCY
VALDFEQEMATAASSSSLEKSYELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHE
TTFNSIMKCDVDIRKDL YANTVL SGGTTMYPGIADRMQKEITALAPSTMKIKI IAPPE
RKYSVWIGGSILASLSTFQQMWISKQEYDESGPSIVHRKCF"
```





做和前面 [步驟二](#) 的 *Import* 類似，*Import* 後打開右邊的 *Sequence Settings*，點擊 *Annotation Type*，可根據圖片上的步驟實現。





以下為結果呈現

