

“实用生物信息技术”课程小组讨论总结报告

组：G2 次：3 组长：边汉青 执笔：高倩

1. 时间

2026.4.22

2. 方式

线上

3. 主题

关于多平台序列比对、CNCB BLAST 及 NCBI BLAST 实例练习的阐述和讨论

4. 内容

A 序列对比

A1 双序列对比

A 1.1 利用中国国家生物信息中心生物信息分析平台（BIT）进行双序列比对

A 1.2 利用欧洲生物信息学研究所双序列比对工具（PSA）进行双序列比对

A 1.3 利用瑞士生物信息学研究所点阵图网站 Dotlet 生成点阵图

A 1.4 利用 NCBI 进行序列比对和生成点阵图

A1.5 利用 dottup 生成点阵图

A2 多序列对比

A2.1 利用 Uniprot 进行多序列对比。

A2.2 利用 NCBI 进行多序列对比。

A2.3 利用 EBI 的 T-Coffee 进行多序列对比

A2.4 利用 Polydot 生成多序列比对点阵图

B BLAST

B1 利用 CNCB 中的 `blastp` 进行相似性比较

B2 NCBI BLAST 实例练习与拓展学习

B2.1 NCBI BLAST 简要介绍

B2.2 灵长目 `alpha` 血红蛋白搜索

B2.3 人的脑红蛋白序列搜索

B2.4 mRNA 序列搜索

B2.5 BLAST 算法部分概念强化

C 问题

D 个人总结

A 序列比对

A1 双序列对比

以感兴趣的棉蚜（cotton aphid）P450 家族基因中的 *CYP6CY3-1* 和 *CYP6CY3-2* 基因的蛋白质进行双序列比对。

A1.1 利用中国国家生物信息中心生物信息分析平台（BIT）进行双序列比对

1) 利用 Uniprot 查找出相关蛋白质序列。

```
>tr|H6WJF4|H6WJF4_APHGO Cytochrome P450 CYP6CY3-1 (Fragment) OS=Aphis gossypii OX=80765 GN=CYP6CY3-1 PE=2 SV=1
MIRDPEIINDVLIKEFASFPDRGIYSDFSVNPLSNLFFMENPQWKTIRSKLSPAFTSGK
LKNMYNQMKCECDILMKNIDILREKNNIEIVRDILGKYATDVIGACAFGLKLSINDDQS
KFRKYGKSIFTPSIRALFRELCLMINPSSLRVRVVKDFSTEATEFFHRVFKETIAYRLEN
KIIVRNDYVDYLLQARKDLVLNPNLPKHEKFTETQIVANAFVMFVAGFETVSTTVSFCLYE
LALNKSIQDKVREIQLKLSQNDGQIDNALLMDLNYLDMVIAETLRKYPPLVALFRKASK
TYQVLNDSLIEKGQKIIIPVYALHYDKQYYTDPEKFIPERFLPEEKAKRPSGII
```

```
>tr|H6WNF2|H6WNF2_APHGO Cytochrome P450 (Fragment) OS=Aphis gossypii OX=80765 GN=CYP6CY3-2 PE=2 SV=1
MIRDPEIINNVLIKDFSYFPNNGIYSDFSVNLLSNLFFMENPQWKLIRKALSPAFTSGK
LKLMYDQIKECSDELKNIHKNFMKIDDKIEVDRMLGKYSTDVIGTCIFGLKLNVAVDDN
STFRKYGKSLFVPSLRTHLRELSLMSPTLLNLIKFRDFPADATEFFHSAPHETIITYREK
NNIVRNDYVDYLLQARNDLVLNKSIPQGERFSESQIIANAFVMFVAGFETVSTAMSFCLY
ELALKKHIQDRVRCINLKLKNNGLINNELLIELNYLDMVIAETLRKYPPTFALFRKAS
QTYHVPNDSLTIEKDQKIIIPYSLHYDPKYFTDPEVFDPERFSPEEKAKRISGTYLPFG
DGPRICIGKRFAELEMKLALVEILTKFEVLPCE
```

2) 打开 BIT 网站选择 NEEDLE

The screenshot displays the BIT website interface with several tool options. The 'Needle' tool is selected, showing its description: 'Needleman-Wunsch global alignment of two sequences.' Other visible tools include BLAST, L00maccner, L00ipain, L00tup, PM-Bdiversity, PM-extract-ma, PM-parallel-meta, Polydot, and Water. A 'Water' logo is also visible in the bottom right of the screenshot.

3) 打开后，分别将 *CYP6CY3-1* 和 *CYP6CY3-2* 基因的蛋白质序列进行填充，后点击提交，便可得到对比结果；结果显示：identity 为 63.2%与 CNCB 中的 BIT 对比结果一致。

The screenshot shows the Emboss Needle web interface. On the left, there are sections for 'Sequences type' (set to Protein), 'Data' (with two sequence input boxes), and 'Parameters' (with fields for Gap open, Gap extend, End gap penalty, End gap open, End gap extend, Matrix, and Result format). A 'Submit' button is at the bottom left. On the right, the alignment results are displayed, including sequence identifiers (e.g., H6WNF2_APHGO, H6WJF4_APHGO), alignment scores (e.g., 251/397 (63.2%)), and a visual alignment of the two sequences with vertical bars indicating matches.

A1.2 利用欧洲生物信息学研究所双序列比对工具（PSA）进行双序列比对

1) 打开 PSA, 点击进入 **emboss needle**; 输入蛋白质序列, 然后点击提交。

The screenshot shows the Job Dispatcher website interface. At the top, there's a 'Job Dispatcher' header for the European Bioinformatics Institute (EMBL-EBI). Below the header, there are navigation links (Home, Help & Privacy, Recent Jobs) and a 'Feedback' button. A yellow banner welcomes users. The main content area features a 'Pairwise Sequence Alignment' section with a description of MSA and a 'Launch EMBOSS Needle' button. Below this is the 'EMBOSS Stretcher' section, which contains the tool's input form. The form has two sequence input boxes, each with a 'Paste your sequence here' instruction and a 'Choose file' button. The 'Parameters' section includes an 'OUTPUT FORMAT' dropdown set to 'pair' and a 'More options' link. At the bottom, there's a 'Submit' button and a 'Title' field containing 'EMBOSS NEEDLE's job'.

2) 结果如下:

A1.4 利用 NCBI 进行序列比对和生成点阵图

1) 将两个蛋白质序列进行填充;

Needleman-Wunsch Global Align Protein Sequences

Enter Query Sequence

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

>tr|H6WJF4|H6WJF4_APHGO Cytochrome P450 CYP6CY3-1 (Fragment)
OS=Aphis gossypii OX=80765 GN=CYP6CY3-1 PE=2 SV=1
MIRDPEIINDVLIKEFASFPDRGIYDFSVNPLSNLFFMENPQWKIRKLSLSP
AFTSGK

Query subrange

From

To

Or, upload file

Job Title
Enter a descriptive title for your BLAST search

Enter Subject Sequence

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

>tr|H6WJF2|H6WJF2_APHGO Cytochrome P450 (Fragment) OS=Aphis
gossypii OX=80765 GN=CYP6CY3-2 PE=2 SV=1
MIRDPEIINDVLIKDFSYFPNRIYDFSVNLLSNLQFFMENPQWKIRKALS
AFTSGK

Subject subrange

From

To

Or, upload file

BLAST Show results in a new window

+ Algorithm parameters

2) 点击 blast 便可得到比对结果和点阵图

Descriptions **Graphic Summary** **Alignments** Dot Plot

Alignment view Pairwise [Restore defaults](#) [Download](#)

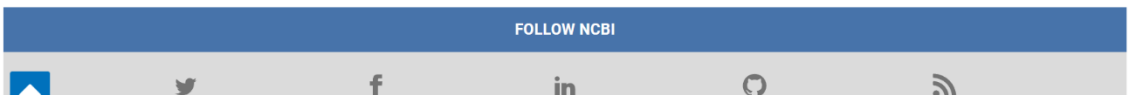
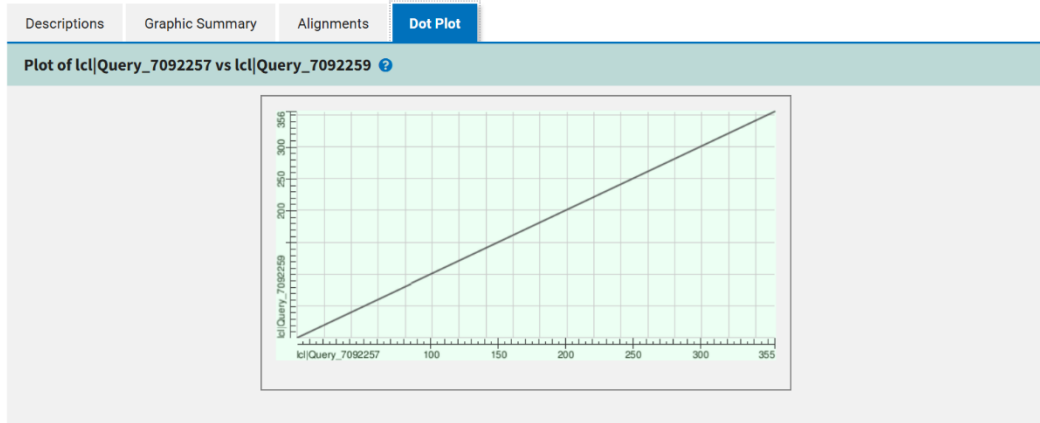
1 sequences selected

[Download](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

tr|H6WJF2|H6WJF2_APHGO Cytochrome P450 (Fragment) OS=Aphis gossypii OX=80765 GN=CYP6CY3-2 PE=2 SV=1
Sequence ID: Query_7092259 Length: 393 Number of Matches: 1

Range 1: 1 to 393 [Graphics](#) [Next Match](#) [Previous Match](#) [Related Information](#)

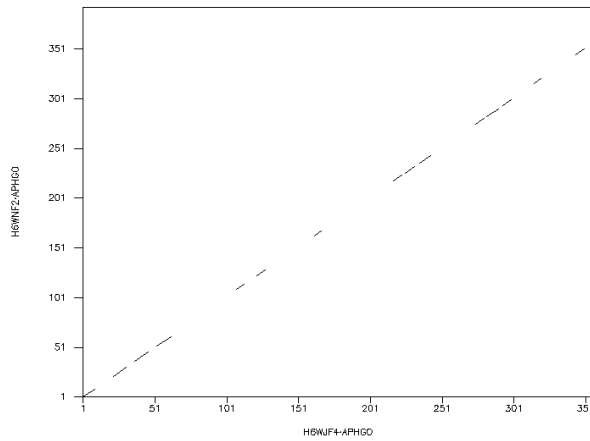
NW Score	Identities (64%)	Positives (75%)	Gaps (9%)
1243	250/393(64%)	297/393(75%)	38/393(9%)
Query 1	MIRDPEIINDVLIKEFASFPDRGIYDFSVNPLSNLFFMENPQWKIRKLSLSPFTSGK	60	
Sbjct 1	MIRDPEIIN+VLIK+FP+RGYSDFSVN LSN LFFMENPQWK IR LSPAFTSGK	60	
Query 61	LKMYNQMKCGDILMKNIDILREK--NNEIEVRDILGKYATDVIGACAFGLKLSINDDQ	119	
Sbjct 61	LKLMY+Q+KEC D LMKNI K +++IEVRD+LGKY+TDVIG C FGLKLN+++DD	120	
Query 120	SKFRKYGSIFTPSIRALFRELCLMINPSLLRVIRVKDFSTATEFFHRVFKETIAYRLE	179	
Sbjct 121	SFRKYKSLFVPSLRHLRELSMISPTLLNLKFRDFPADATEFFHSAPHEITTYREK	180	
Query 180	NKIVRNDYDYLQARKDLVLPNLPKHEKFTETQIVANAFVMFVAGFETVSTVTSFCLY	239	
Sbjct 181	NNIVRNDYQTLIQARNDLVLNKSPQGERFSESQIIANAFVMAAGFETVSTAMSFCLY	240	
Query 240	ELALNKSIQDQVREIQLKLSQNDGQIDNALLMDLNLDMVIAETLRKYPPLVALFRKAS	299	
Sbjct 241	ELALKKHIDQVRVQEIINLKLKSNGLINNELLIELNLDLWVLAETLRKYPPTALFRKAS	300	
Query 300	KTYQVNDLSLEIEGQKIIIPVYALHYDKQYTDPEKFIPEKFLPEEKAKRPSGII----	355	
Sbjct 301	+TY V NDSL IEK QKLIIP+Y+LHYD +Y+TDPE F PERF PEEKAKR SG	360	
Sbjct 361	DGPRICIGKRFAELEMLLALVEILTKFEVLPC 393		



A1.5 利用 dottup 生成点阵图

1) 将序列输入 dottup 序列框中之后点击“go”；

2) 从中我们可以直观的看出 CYP6CY3-1 和 CYP6CY3-2 蛋白质序列是有一定的相似性的；



A2 多序列对比

A2.1 利用 Uniprot 进行多序列对比

1) 在 Uniprot 中搜索 *CYP6CY3* 这个基因的蛋白质序列，将搜索到的添加至 basket，选择 tool 中的 align，进行序列对比；

UniProtKB 5 results in your basket

Tools Download (5) Remove View: Cards Table Customize columns

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
<input type="checkbox"/> H6WNF2	H6WNF2-APHGO	Cytochrome P450	CYP6CY3-2	Aphis gossypii (Cotton aphid)	393 AA
<input type="checkbox"/> H6WJF4	H6WJF4-APHGO	Cytochrome P450 CYP6CY3-1	CYP6CY3-1	Aphis gossypii (Cotton aphid)	355 AA
<input type="checkbox"/> V5SQ25	V5SQ25-MYZPE	Cytochrome P450 CYP6CY3	CYP6CY3	Myzus persicae (Green peach aphid) (Aphis persicae)	511 AA
<input type="checkbox"/> A0A7T3R134	A0A7T3R134-RHOPD	Cytochrome P450 CYP6CY3	CYP6CY3	Rhopalosiphum padi (Bird cherry-oat aphid) (Aphis padi)	511 AA
<input type="checkbox"/> E7CGB1	E7CGB1-MYZPE	Cytochrome P450	CYP6CY3	Myzus persicae (Green peach aphid) (Aphis persicae)	511 AA

2) 点击 align 5 sequence

Align

Find a protein sequence by UniProt ID (e.g. P05067 or A4_HUMAN or UPI000000001) to align with the Clustal Omega program. You can also paste a list of IDs.

UniProt IDs

OR

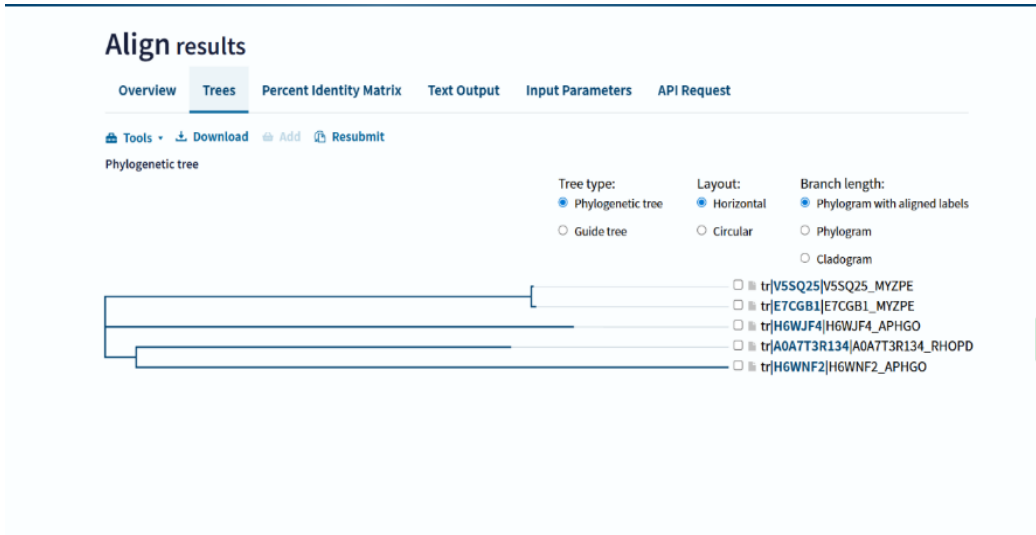
Enter multiple protein or nucleotide sequences (50 max), separated by a FASTA header. You may also load from a text file.

```
>tr|H6WNF2|H6WNF2-APHGO Cytochrome P450 (Fragment) OS=Aphis gossypii OX=80765 GN=CYP6CY3-2 PE=2 SV=1
MIRDPEIINNVLDKDFSYFHRGIYDFSVNLLSNQLFFHENPQKLIKALSPAFTSGK
LKLHYDQIKESDDELKNIHKNFMIIDKIEVRDMLGKYSTDVIGTCIFGLKLNVAUSDON
STFRKYGKSLFVPSLRTHRELSLMSIPTLLNIZLKRDFPADATEFFHSAFHETIYREK
NNIVRNDFVQTLIQARNDLVNLKSIQGERFSESQIZANAFVHFAAGFETVSTAMSFCLY
ELALKKHIZQDRVRQEIINLKLKSNGLINNELLIEIENLMDVLAETLRKYPPTFALFRKAS
QTYHVPNDSLTIEKDKIIPYISLHYDPKVFYDPEVDFPERFSPPEEKAKRISGTYLPPG
DGPRTICIKRFAELERKALVLEILTKFEVLPCPE
>tr|H6WJF4|H6WJF4-APHGO Cytochrome P450 CYP6CY3-1 (Fragment) OS=Aphis gossypii OX=80765 GN=CYP6CY3-1 PE=2 SV=1
```

Your input contains 5 sequences.

Name your Align job

3) 查看进化树；

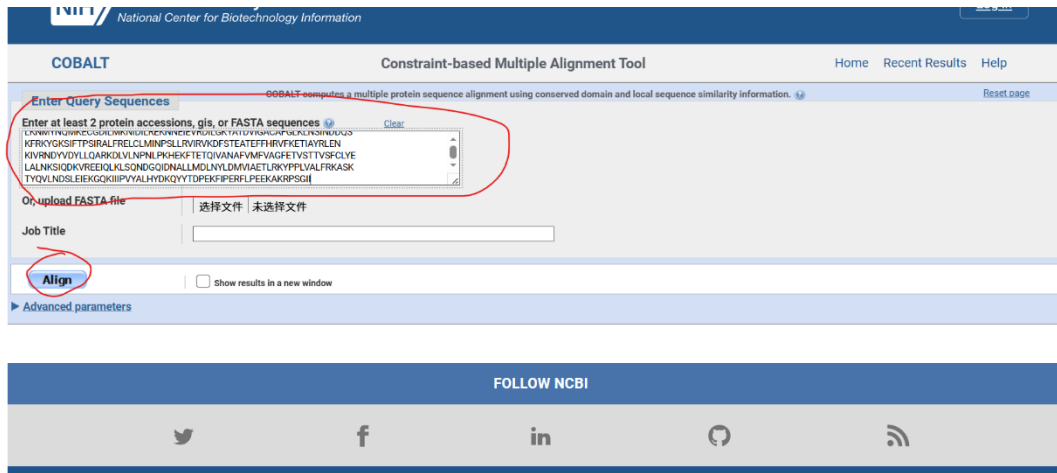


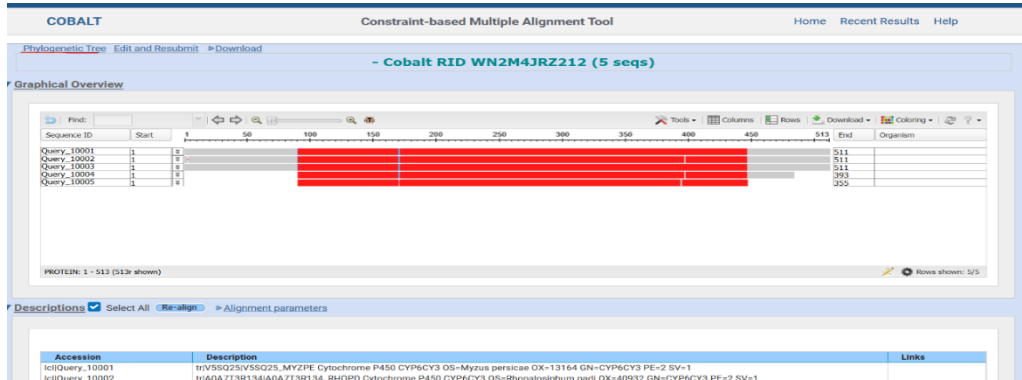
4) 查看每两条序列的比对结果

	tr H6WNF2 H6WNF2_APHGO	tr H6WJF4 H6WJF4_APHGO	tr A0A7T3R134 A0A7T3R134_RHOPD	tr V5SQ25 V5SQ25_MYZPE	tr E7CGB1 E7CGB1_MYZPE
tr H6WNF2 H6WNF2_APHGO	100.00%	70.14%	73.54%	71.17%	71.17%
tr H6WJF4 H6WJF4_APHGO	70.14%	100.00%	76.06%	75.42%	75.42%
tr A0A7T3R134 A0A7T3R134_RHOPD	73.54%	76.06%	100.00%	77.25%	77.06%
tr V5SQ25 V5SQ25_MYZPE	71.17%	75.42%	77.25%	100.00%	99.80%
tr E7CGB1 E7CGB1_MYZPE	71.17%	75.42%	77.06%	99.80%	100.00%

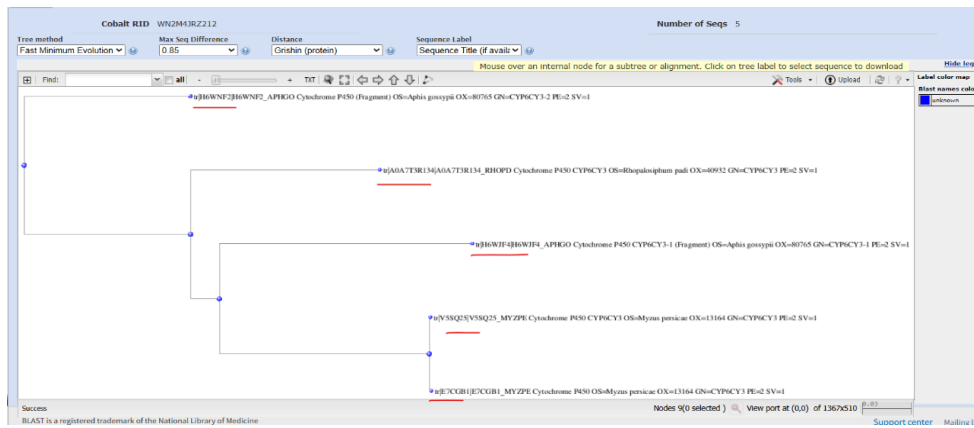
A2.2 利用 NCBI 进行多序列对比

1) 进入 COBALT, 之后将序列粘贴至序列框;





2) 点击结果上面的系统进化树，也可以查看下面的比对信息。



A2.3 利用 EBI 的 T-Coffee 进行多序列对比。

1) 首先将序列输入输入框，点击提交 submit;

2) 点击进化树，便可查看这几个基因形成的系统进化树，同时还有文件形式的结果可以通过 result files 进行查看，其中也包含每两条的相似性比对结果；

T-Coffee

Multiple Sequence Alignment (MSA)

Home Help & Privacy Recent Jobs **Input Form**

Feedback

Welcome to the Job Dispatcher website! If you need assistance or have feedback, please contact us.

Results for Job ID: tcoffee-I20260331-133535-0341-25266024-p1m

Copy

Resubmission

1 current label

Tool Output Alignments Guide Tree **Phylogenetic Tree** Results Viewers Result Files Submission Details

Phylogenetic Tree

```
(
  (
    tr|A0A7T3R134|A0A7T3R134_RHOPD:0.10223,
    tr|H6WJF4|H6WJF4_APHGO:0.16241)
  :0.06828,
  (
    tr|E7CGB1|E7CGB1_MYZPE:0.00131,
    tr|V5SQ25|V5SQ25_MYZPE:0.00665)
  :0.00000
```

Tool Output	Alignments	Guide Tree	Phylogenetic Tree	Results Viewers	Result Files	Submission Details
Tool Output			tcoffee-I20260331-133535-0341-25266024-p1m.tco		Download	
Input Sequences			tcoffee-I20260331-133535-0341-25266024-p1m-sequences		Download	
Alignment in CLUSTAL format			tcoffee-I20260331-133535-0341-25266024-p1m-align		Download	
Guide Tree			tcoffee-I20260331-133535-0341-25266024-p1m-gtree		Download	
Phylogenetic Tree			tcoffee-I20260331-133535-0341-25266024-p1m-phytree		Download	
Percent Identity Matrix			tcoffee-I20260331-133535-0341-25266024-p1m-pidmat		Download	

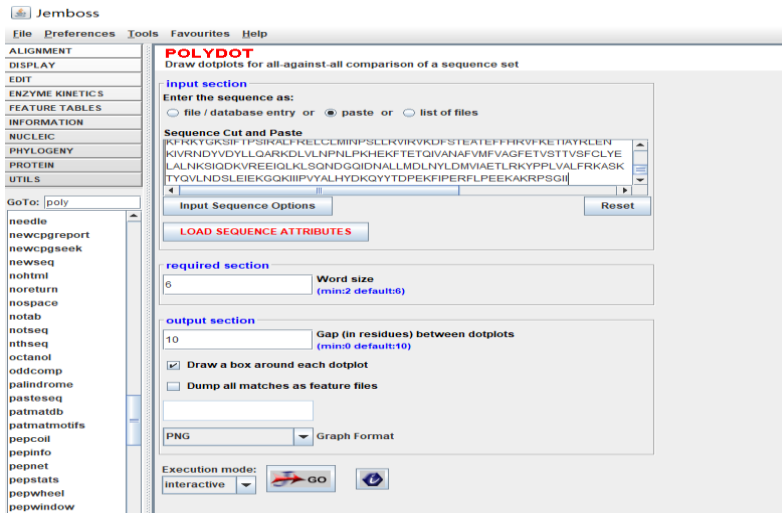
⌄
⌄
⌄
⌄
⌄

Percent Identity Matrix - created by Clustal2.1

1:	tr A0A7T3R134 A0A7T3R134_RHOPD	100.00	77.45	76.06	73.54	77.65
2:	tr E7CGB1 E7CGB1_MYZPE	77.45	100.00	75.71	71.43	99.80
3:	tr H6WJF4 H6WJF4_APHGO	76.06	75.71	100.00	70.14	75.71
4:	tr H6WJF2 H6WJF2_APHGO	73.54	71.43	70.14	100.00	71.43
5:	tr V5SQ25 V5SQ25_MYZPE	77.65	99.80	75.71	71.43	100.00

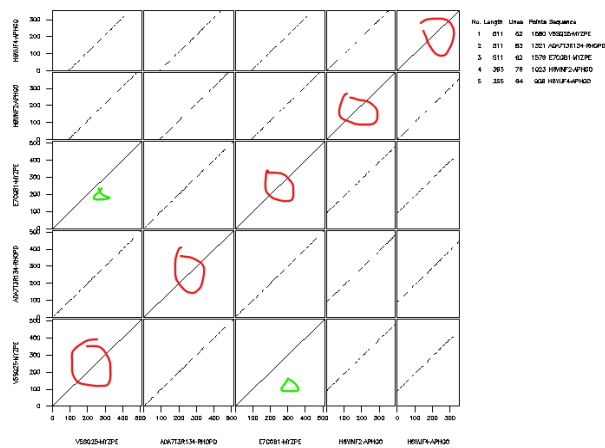
A2.4 利用 polydot 生成多序列比对点阵图

1) 将序列输入 jemboss 软件的 polydot 程序的序列框中;



2) 结果显示, 红色圈出的斜对角线为自身比对, 绿色标记的为相似性比较高的, 即桃蚜中的 CYP6CYP3, 登记号不同, 但种属相同 V5SQ25_MYZPE 和 E7CGB1_MYZPE, 而棉蚜中的 CYP6CYP3-1 和 CYP6CYP3-2 的相似性是最底的。

Poly dotplot of C:\Users\...AppData\Local\Temp\polydot07...
Wed 8 Apr 2026 19:25:50



B BLAST

B1 利用 CNCB 中的 blastp 进行相似性比较

1) 将 CYP6CYP3-1 基因的蛋白质输入第一框中, 将在 uniprot 中找到的五个相似蛋白序列 (V5SQ25、A0A7T3R134、E7CGB1、H6WJF2 和 H6WJF4) 输入参考序列框;

BLASTN **BLASTP** BLASTX TBLASTN TBLASTX

输入搜索序列

请输入序列: [示例](#) [清空](#)

```
>tr|H6WJF4|H6WJF4_APHGO Cytochrome P450 CYP6CY3-1 (Fragment) OS=Aphis gossypii OX=80765
GN=CYP6CY3-1 PE=2 SV=1
MIRDP EIINDVLIKEFASFPDRGIYSDFSVNP LSNLFFMENPQWKTRSKLSPAFTSGK LKNMYNQMK E
CGDILMKNDILREKNNEIEVRDILGKYATDVIGACAFGLK LNSINDQSKFRKYGKSI FTSPSIRALFRE
```

或上传文件: [选择文件](#)

任务名字: 30 / 30

选择目标: 与数据库比对 与输入序列比对

查询子区间: 到:

输入参考序列

请输入序列: [示例](#) [清空](#)

```
>tr|V5SQ25|V5SQ25_MYZPE Cytochrome P450 CYP6CY3 OS=Myzus persicae OX=13164 GN=CYP6CY3 PE=2
SV=1
MNLSSSTTDWWIYIASACLVGVTIYYFCISTF SKWEKLNVPYIRPIPLGFN FVRVALSK
DHPLEFYNKIYYKAGLKYGGFLQMRTPYLMIRDPEI INNVLIKDFSSFPDRGIYDLAA
```

或者, 上传文件: [选择文件](#)

查询子区间: 到:

2) 结果显示, 第一个为自身对比, 其余对比结果与 Uniprot 多序列对比结果相同, CYP6CYP3-1 与 CYP6CY3 (uniprot 编号: A0A7T3R134) 高度相似。

为进一步深入了解该蛋白, 建议开展以下分析:

- 保守结构域与活性位点分析:** 使用CDD、Pfam或InterPro数据库确认P450的保守结构域, 并精确标注血红素结合残基 (如Cys)。
- 多序列比对与系统发育分析:** 将查询序列与更多昆虫的CYP6家族及已知功能的P450进行比对, 构建系统发育树, 明确其在进化树中的具体位置, 并推断其与其他抗性相关P450 (如CYP6M1) 的关系。
- 三维结构建模:** 以已知的P450晶体结构 (如P450BM3) 为模板, 进行同源建模。重点分析底物结合口袋的氨基酸组成, 以预测其可能的底物结合特性。
- 分子对接研究:** 将模型与已知的杀虫剂 (如吡虫啉) 进行分子对接, 从结构层面探索其代谢活性的分子基础。
- 实验验证:** 建议通过体外表达该蛋白, 进行酶活测定, 直接验证其对特定杀虫剂的代谢能力; 或通过RNA干扰技术在活体蚜虫中敲低该基因, 观察其对杀虫剂敏感性的变化。

表格描述 图形概览 比对详情 Taxonomy

以下表格展示的是比对显著的结果序列 [下载](#) [选择列](#) [显示条目](#)

5 条序列已选中 [距离树](#) [多序列比对](#)

序列描述	物种名称	最高得分	总得分	覆盖率	期望值	相似度	序列长	序列号
tr_H6WJF4_H6WJF4_APHGO_Cytochrome_P450_CYP6CY3_1...	N/A	807	807	100.00%	0	100.00%	355	BL_ORD_ID
tr_A0A7T3R134_A0A7T3R134_RHOPD_Cytochrome_P450_CYP...	N/A	630	630	99.00%	0	76.27%	511	BL_ORD_ID
tr_E7CGB1_E7CGB1_MYZPE_Cytochrome_P450_OS_Myzus_pe...	N/A	628	628	99.00%	0	75.78%	511	BL_ORD_ID
tr_V5SQ25_V5SQ25_MYZPE_Cytochrome_P450_CYP6CY3_OS...	N/A	628	628	99.00%	0	75.78%	511	BL_ORD_ID
tr_H6WJF2_H6WJF2_APHGO_Cytochrome_P450_Fragment...	N/A	573	573	99.00%	0	70.34%	393	BL_ORD_ID

B2 NCBI BLAST 实例练习与拓展学习

B2.1 NCBI BLAST 简要介绍

The screenshot shows the NCBI BLAST website. At the top, there is a navigation bar with the NIH logo and the text "National Library of Medicine National Center for Biotechnology Information 美国国立医学图书馆". Below this, the word "BLAST" is prominently displayed. The main content area features a "Basic Local Alignment Search Tool" section with a description in English and Chinese. A "Web BLAST" section offers three options: "Nucleotide BLAST" (nucleotide to nucleotide), "blastx" (translated nucleotide to protein), and "Protein BLAST" (protein to protein). A "BLAST Genomes" section at the bottom allows users to select an organism for searching, with options for Human, Mouse, Rat, and Microbes. A search bar is located at the bottom right of the BLAST Genomes section.

基础局部比对搜索工具
Basic Local Alignment Search Tool
BLAST 用于寻找生物序列间的相似性区域。
BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.
该程序将核苷酸或蛋白序列与序列数据库进行比对，并计算匹配结果的统计显著性。
Learn more
The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families. BLAST 可用于推断序列间的功能与进化关系，并辅助鉴定基因家族成员。

Web BLAST

- Nucleotide BLAST**
nucleotide to nucleotide
核酸 BLAST
(核酸 → 核酸)
- blastx**
(翻译后的核酸 → 蛋白)
blastx translated nucleotide to protein
- tblastn**
protein to translated nucleotide
tblastn (蛋白 → 翻译后的核酸)
- Protein BLAST**
protein to protein
蛋白 BLAST
(蛋白 → 蛋白)

BLAST Genomes
Enter organism common name, scientific name, or tax id
Human Mouse Rat Microbes Search

因此，BLAST 的功能主要有以下四点

序列鉴定 (ID)：拿到一段未知序列，通过 BLAST 确定它属于哪个基因/物种；

功能注释：预测新测序基因的生物学功能；

同源寻找：在不同物种中寻找同源基因；

进化分析：获取同源序列，构建进化树。

Standalone and API BLAST 独立部署与应用程序接口 BLAST：指非网页版的本地运行与程序化调用方式。

- Download BLAST**
Get BLAST databases and executables
获取 BLAST 数据库及可执行程序
- Use BLAST API**
Call BLAST from your application
从应用程序中调用 BLAST
- Use BLAST in the cloud**
Start an instance at a cloud provider
在云服务商处启动计算实例

Specialized searches

- SmartBLAST**
Find proteins highly similar to your query
寻找与查询序列高度相似的蛋白序列。
- 引物 BLAST Primer-BLAST**
Design primers specific to your PCR template
针对 PCR 模板设计特异性引物
- Global Align**
比较两条序列的全长度相似性
Compare two sequences across their entire span (Needleman-Wunsch)
- 保守结构域搜索 CD-search**
在序列中查找保守的结构域。
Find conserved domains in your sequence
- 免疫球蛋白 BLAST IgBLAST**
Search immunoglobulins and T cell receptor sequences
搜索免疫球蛋白 (Ig) 及 T 细胞受体 (TCR) 序列。
- 载体屏蔽 VecScreen**
Search sequences for vector contamination
搜索序列中的载体污染序列。
- 保守结构域架构搜索 CDART**
Find sequences with similar conserved domain architecture
寻找具有相似保守结构域架构的序列。利用结构域和蛋白约束条件对序列进行比对。
- 多重序列比对 Multiple Alignment**
Align sequences using domain and protein constraints

B2.2 灵长目 alpha 血红蛋白搜索

以人血红蛋白 alpha 亚基 HBA_HUMAN 为搜索序列，采用默认参数利用中国国家生物信息中心 CNCB 蛋白质数据库搜索程序 BlastP，搜索 SwissProt 数据库，找出灵长目动物 Primates（分类学数据库登录号 taxid 9443）中 alpha 血红蛋白；

Standard Protein BLAST

blastn **blastp** blastx tblastn tblastx

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

Reset page Bookmark

Enter Query Sequence

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

Query subrange

Or, upload file

Job Title
Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database

Organism exclude
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

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

Search database swissprot using Blastp (protein-protein BLAST) Show results in a new window

结果显示有 100 条序列；

[← Edit Search](#) [Save Search](#) [Search Summary](#) [How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

! Your search is limited to records include: **Primates (taxid:9443)**

Job Title HBA_HUMAN - P69905, Hemoglobin subunit alpha, ...
RID YEBF4D19014 Search expires on 04-22 15:57 pm [Download All](#) [Filter Results](#)
Program BLASTP [Citation](#) [Filter Results](#)
Database swissprot [See details](#)
Query ID lcl|Query_6163958
Description HBA_HUMAN - P69905, Hemoglobin subunit alpha, HBA1; ...
Molecule type amino acid
Query Length 142
Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#)

Filter Results
Organism *only top 20 will appear* exclude
 Type common name, binomial, taxid or group name
[+ Add organism](#)
Percent Identity **E value** **Query Coverage**
 to to to

Descriptions [Graphic Summary](#) [Alignments](#) [Taxonomy](#)

Sequences producing significant alignments [Download](#) [Select columns](#) Show 100

select all 100 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"/>	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain; C...	Homo sapiens	286	286	100%	1e-101	100.00%	142	P69905.2
<input checked="" type="checkbox"/>	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain; C...	Gorilla gorilla gorilla	282	282	99%	8e-100	99.29%	141	P01923.1
<input checked="" type="checkbox"/>	RecName: Full=Hemoglobin subunit alpha-1; AltName: Full=Alpha-1-globin; AltName: Full=Hemoglobin alpha-1 ch...	Hylobates lar	281	281	100%	2e-99	98.59%	142	Q9T3S3.2
<input checked="" type="checkbox"/>	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain; C...	Pongo pygmaeus	281	281	100%	3e-99	97.89%	142	P06635.2

配置算法参数，依次改变常规参数：最大目标序列、期望阈值、字长，分析它们对输出结果的影响。

Max target sequences（最大目标序列）：返回排名靠前的指定数量结果，比如100，就返回排名靠前的100条结果，不足100，就会返回全部结果。

Expect threshold（期望阈值，E值）：E值越小，越严格，随机匹配的可能性越低，结果更可信但可能漏掉较弱的真实同源关系。

Word size（字长）：字长越大，搜索速度越快，但敏感度降低，可能漏掉远源同源序列。

General Parameters

Max target sequences [?](#)
 Select the maximum number of aligned sequences to display

Short queries Automatically adjust parameters for short input sequences [?](#)

Expect threshold [?](#)

Word size [?](#)

Max matches in a query range [?](#)

配置算法参数，依次改变计分参数：计分矩阵、空位罚分，分析它们对输出结果的影响。

Matrix（替换矩阵）：PAM 适合远源比对，BLOSUM 适合通用比对；PAM 数字越大，越“宽容”，可能找到远源关系，数字越小越保守；BLOSUM 数字越大

则越严格，搜索的结果可能更少，找高度相似序列，数字越小则越宽容，能找到关系更远的序列。

Gap Costs(空位罚分): 比如图中存在罚分 11 就是每新建一个空位，扣 11 分；延伸罚分 1 就是把已有的空位再拉长 1 个字母，每个只扣 1 分。空位罚分（存在 11/延伸 1）让算法倾向少开空位、但允许空位很长，能避免把一次插入缺失错拆成多段，比对结果更符合真实进化关系，大多时候默认用高存在罚分低延伸罚分，只有在明确知道两个序列非常相似、只有零星单氨基酸差异时，才考虑调低存在罚分。

例如：假设两条序列比对时，需要对齐一个 5 个氨基酸的差异区域：

(1) 一个长空位（1），长 5）

扣分 = 存在 11 + 延伸 1×4 = 15 分

(2) 5 个短空位（每个长 1）

扣分 = (11 + 0)×5 = 55 分

Scoring Parameters	
Matrix	PAM250 ?
Gap Costs	Existence: 13 Extension: 2 ?
Compositional adjustments	Conditional compositional score matrix adjustment ?

利用筛选结果功能，找出黑猩猩（Chimpanzee, Pan troglodytes）中同源蛋白；结果显示有七条来自黑猩猩的同源蛋白；

Filter Results

Organism *only top 20 will appear* exclude
 Pan troglodytes (taxid:9598)
 + Add organism

Percent Identity to E value to Query Coverage to
 Filter Reset

Sequences producing significant alignments Download Select columns Show 100

Your results are filtered to match records that include: **Pan troglodytes (taxid:9598)** Reset

select all 7 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
RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain; Con...	Pan troglodytes	286	286	100%	1e-101	100.00%	142	P69907.2
RecName: Full=Hemoglobin subunit alpha-3; AltName: Full=Alpha-3-globin; AltName: Full=Hemoglobin alpha-2 chaj...	Pan troglodytes	266	266	99%	3e-93	88.65%	141	P01935.1
RecName: Full=Hemoglobin subunit zeta; AltName: Full=Hemoglobin zeta chain; AltName: Full=Zeta-globin [Pan tro...	Pan troglodytes	172	172	100%	2e-56	59.15%	142	P06347.2
RecName: Full=Hemoglobin subunit delta; AltName: Full=Delta-globin; AltName: Full=Hemoglobin delta chain [Pan tr...	Pan troglodytes	116	116	98%	4e-34	44.14%	147	P61772.2
RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-globin; AltName: Full=Hemoglobin beta chain [Pan tro...	Pan troglodytes	114	114	98%	2e-33	43.45%	147	P68873.2
RecName: Full=Hemoglobin subunit gamma-1; AltName: Full=Gamma-1-globin; AltName: Full=Hemoglobin gamma-...	Pan troglodytes	113	113	98%	5e-33	41.38%	147	P61920.2

利用筛选结果功能，找出与 HBA_HUMAN 相同位点高于 95%的物种，下载 FASTA 格式序列；结果显示有 21 条相似度>95%的序列；

Filter Results

Organism *only top 20 will appear* exclude
 Type common name, binomial, taxid or group name
 + Add organism

Percent Identity 95 to E value to Query Coverage to
 Filter Reset

Sequences producing significant alignments Download Select columns Show 100

Your results are filtered to match records with **percent identity** between 95 and 100. Reset

select all 21 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
RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain; Con...	Homo sapiens	286	286	100%	1e-101	100.00%	142	P69905.2
RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain; Con...	Gorilla gorilla gorilla	282	282	99%	8e-100	99.29%	141	P01923.1
RecName: Full=Hemoglobin subunit alpha-1; AltName: Full=Alpha-1-globin; AltName: Full=Hemoglobin alpha-1 chaj...	Hylobates lar	281	281	100%	2e-99	98.59%	142	Q9TS35.2
RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain; Con...	Pongo pygmaeus	281	281	100%	3e-99	97.89%	142	P06635.2
RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain; Con...	Semnopithecus e...	278	278	99%	2e-98	97.87%	141	P01924.1

B2.3 人的脑红蛋白序列搜索

以人的血红蛋白 alpha 亚基 HBA_HUMAN 为搜索序列，利用中国国家生物信息中心 CNCB 蛋白质数据库搜索程序 BlastP，搜索 SwissProt 数据库，设置字

长、计分矩阵和期望阈值，找出包括脑红蛋白 Neuroglobin 在内的 12 个珠蛋白，说明上述参数对搜索结果的影响。

以人的血红蛋白 alpha 亚基 HBA_HUMAN 为搜索序列，用 NCBI 位点特异迭代法 PSI-Blast 搜索 SwissProt 数据库，选择物种 Homo Sapiens, taxid 9606，选择默认字长 Word size，默认计分矩阵 Matrix 和默认空位罚分 Gap costs，分析运行结果。

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 [?](#)

GSAQVKGHG
KKVADALTNVAHVDDMPNALSALSADLHAHKLRVDPVNFKLLSHCLLVTLAAH
LPAEFTP
AVHASLQKFLASVSTVLTISKYR

From
To

Or, upload file 未选择文件 [?](#)

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

Align two or more sequences [?](#)

Choose Search Set

Database ?

Organism Optional exclude
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

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 [?](#)

Algorithm parameters

General Parameters

Max target sequences
Select the maximum number of aligned sequences to display ?

Short queries Automatically adjust parameters for short input sequences ?

Expect threshold ?

Word size ?

Max matches in a query range ?

Scoring Parameters

Matrix ?

Gap Costs Existence: 11 Extension: 1 ?

Compositional adjustments ?

Filters and Masking

Filter Low complexity regions ?

Mask Mask for lookup table only ?
 Mask lower case letters ?

PSI/PHI/DELTA BLAST

Upload PSSM ?
Optional

PSI-BLAST Threshold ?

Pseudocount ?

BLAST

Search database swissprot using **PSI-BLAST (Position-Specific Iterated BLAST)**

Show results in a new window

Number of sequences Run

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Select columns Show ?

11 sequences selected GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

Sequences with E-value BETTER than threshold

select all 11 sequences selected PSI-BLAST iteration 1

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	New added
<input checked="" type="checkbox"/>	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglo...	Homo sapiens	286	286	100%	1e-101	100.00%	142	P69905.2	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>	RecName: Full=Hemoglobin subunit theta-1; AltName: Full=Hemoglobin theta-1 chain; AltNam...	Homo sapiens	182	182	100%	4e-60	61.97%	142	P09105.2	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>	RecName: Full=Hemoglobin subunit zeta; AltName: Full=HBAZ; AltName: Full=Hemoglobin zet...	Homo sapiens	176	176	100%	7e-58	59.86%	142	P02008.2	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>	RecName: Full=Hemoglobin subunit mu; AltName: Full=Hemoglobin mu chain; AltName: Full=...	Homo sapiens	135	135	99%	9e-42	45.39%	141	Q6B0K9.1	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>	RecName: Full=Hemoglobin subunit delta; AltName: Full=Delta-globin; AltName: Full=Hemoglo...	Homo sapiens	114	114	98%	1e-33	43.45%	147	P02042.2	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>	RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-globin; AltName: Full=Hemoglob...	Homo sapiens	114	114	98%	1e-33	43.45%	147	P68871.2	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>	RecName: Full=Hemoglobin subunit gamma-2; AltName: Full=Gamma-2-globin; AltName: Full=...	Homo sapiens	113	113	98%	5e-33	41.38%	147	P69892.2	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>	RecName: Full=Hemoglobin subunit gamma-1; AltName: Full=Gamma-1-globin; AltName: Full=...	Homo sapiens	112	112	98%	2e-32	41.38%	147	P69891.3	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>	RecName: Full=Hemoglobin subunit epsilon; AltName: Full=Epsilon-globin; AltName: Full=Hem...	Homo sapiens	101	101	95%	2e-28	39.01%	147	P02100.2	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>	RecName: Full=Cytooglobin; AltName: Full=Histoglobin; Short=HGb; AltName: Full=Nitric oxyge...	Homo sapiens	68.9	68.9	96%	2e-15	28.08%	190	Q8WWM9.1	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>	RecName: Full=Myoglobin; AltName: Full=Nitrite reductase MR; AltName: Full=Pseudoperoxid...	Homo sapiens	51.2	51.2	100%	6e-09	27.52%	154	P02144.2	<input checked="" type="checkbox"/>		

点击 Alignments 按钮，将双序列输出 Pairwise 格式改为多序列输出格式 Flat query-anchored with letters for identities，找出完全一致的保守位点，如色氨酸 W，组氨酸 H；性质相似的位点，如酸性氨基酸谷氨酸 E 或天冬氨酸 D。

- (1) 完全一致：一整列完全相同。
- (2) 相似：虽然氨基酸不同，但是一列的氨基酸化学性质相同。

Alignment view Flat query-anchored with letters for identities Line length: 60 Restore defaults

Download

Query range 1: 1 to 60

```

Query 1  MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFSLFPTTKTYFPHF-----D----- 48
P69905.2 1  MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFSLFPTTKTYFPHF-----D----- 48
P09105.2 1  MALSAEDRALVRLWKKLGSNVGYTTEALERTFLAPPATKTYFPHL-----D----- 48
P02008.2 1  MSLTKTERTIIIVSMWAKISTQADTIGTETLERLFLSHPQTKTYFPHF-----D----- 48
Q6B0K9.1 1  MLSAQERAQIAQVWDLIAGHEAQFGAELLRLRFTVYPSTKYVPHL-----S----- 47
P02042.2 4  LTPEEKTAVALWGVNVDALVGEALGRLLVVYPWTRQFFESFG-----D----- 48
P68871.2 4  LTPEEKSAVTALWGVNVDALVGEALGRLLVVYPWTRQFFESFG-----D----- 48
P69892.2 4  FTEEDKATITSLWGVNVE--DAGGETLGRLLVVYPWTRQFFDSFG-----N----- 48
P69891.3 4  FTEEDKATITSLWGVNVE--DAGGETLGRLLVVYPWTRQFFDSFG-----N----- 48
P02100.2 8  EKAAVTSLWKMVNE--EAGGEALGRLLVVYPWTRQFFDSFG-----N----- 48
Q8WWM9.1 22  AERKAVQAMWARLYANCEDVGVAILVRFVNFPSAKQYFSQF-----KHMEDPL 70
P02144.2 1  MGLSDGEWQLVLNVWGVKVEADIPGHGQEVILRLFKGHPETLEKDFKFKHLKSED----- 54
    
```

Download

Query range 2: 61 to 120

```

Query 49  LS---H---GSA---QVKGHGKVVADALTNVAHV---DDMPNALSALSDLHAH 90
P69905.2 49  LS---H---GSA---QVKGHGKVVADALTNVAHV---DDMPNALSALSDLHAH 90
P09105.2 49  LS---P---GSS---QVRAHGKVVADALSLAVERL---DDLPHALSALSHLHAC 90
P02008.2 49  LH---P---GSA---QLRAHGSKVVAAVGDAVKSI---DDIGGALSSELHAY 90
Q6B0K9.1 48  AC---Q---DAT---QLLSHGQRMLAAGVAQVHY---DNLRAALSPLADLHAL 89
P02042.2 49  LS---SPDAVMGNP---KVKAHGKVVLAGFSDGLAHL---DNLKGTFSQLSELHCD 95
P68871.2 49  LSTPDVAVM---GNP---KVKAHGKVVLAGFSDGLAHL---DNLKGTFAQLSELHCD 95
P69892.2 49  LS---S---ASAIMGNPKVKAHGKVVLSLGDATKHL---DDLKGTFAQLSELHCD 95
P69891.3 49  LS---S---ASAIMGNPKVKAHGKVVLSLGDATKHL---DDLKGTFAQLSELHCD 95
P02100.2 49  LS---SPSAILGNP---KVKAHGKVVLSFGDAIKNM---DNLKPAFAKLSSELHCD 95
Q8WWM9.1 71  EM---E---RSP---QLRKHACRVMGALNTVVENLHDPDKVSSVVALVKGKHAL 115
P02144.2 55  EM---K---ASE---DLKKHGATVLTALGGILKKK---GHHEAEIKPLAQSHAT 96
    
```

在上述第一轮搜索基础上，PSI-BLAST 根据不同位点保守性不同，构建位点特异计分矩阵，进行第二轮搜索。

Descriptions Graphic Summary Alignments Taxonomy number of sequences 500

Sequences producing significant alignments Download Select columns Show 500

11 sequences selected GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

Sequences with E-value BETTER than threshold

select all 11 sequences selected PSI-BLAST iteration 1

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	New add
RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemogl...	Homo sapiens	286	286	100%	1e-101	100.00%	142	P69905.2	<input checked="" type="checkbox"/>		
RecName: Full=Hemoglobin subunit theta-1; AltName: Full=Hemoglobin theta-1 chain; AltName...	Homo sapiens	182	182	100%	4e-60	61.97%	142	P09105.2	<input checked="" type="checkbox"/>		
RecName: Full=Hemoglobin subunit zeta; AltName: Full=HBAZ; AltName: Full=Hemoglobin zet...	Homo sapiens	176	176	100%	7e-58	59.86%	142	P02008.2	<input checked="" type="checkbox"/>		
RecName: Full=Hemoglobin subunit mu; AltName: Full=Hemoglobin mu chain; AltName: Full=...	Homo sapiens	135	135	99%	9e-42	45.39%	141	Q6B0K9.1	<input checked="" type="checkbox"/>		
RecName: Full=Hemoglobin subunit delta; AltName: Full=Delta-globin; AltName: Full=Hemogl...	Homo sapiens	114	114	98%	1e-33	43.45%	147	P02042.2	<input checked="" type="checkbox"/>		
RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-globin; AltName: Full=Hemoglob...	Homo sapiens	114	114	98%	1e-33	43.45%	147	P68871.2	<input checked="" type="checkbox"/>		
RecName: Full=Hemoglobin subunit gamma-2; AltName: Full=Gamma-2-globin; AltName: Full=...	Homo sapiens	113	113	98%	5e-33	41.38%	147	P69892.2	<input checked="" type="checkbox"/>		
RecName: Full=Hemoglobin subunit gamma-1; AltName: Full=Gamma-1-globin; AltName: Full=...	Homo sapiens	112	112	98%	2e-32	41.38%	147	P69891.3	<input checked="" type="checkbox"/>		
RecName: Full=Hemoglobin subunit epsilon; AltName: Full=Epsilon-globin; AltName: Full=Hem...	Homo sapiens	101	101	95%	2e-28	39.01%	147	P02100.2	<input checked="" type="checkbox"/>		
RecName: Full=Cytooglobin; AltName: Full=Histoglobin; Short=HGb; AltName: Full=Nitric oxide...	Homo sapiens	68.9	68.9	96%	2e-15	28.08%	190	Q8WWM9.1	<input checked="" type="checkbox"/>		
RecName: Full=Myoglobin; AltName: Full=Nitrite reductase MB; AltName: Full=Pseudoperoxid...	Homo sapiens	51.2	51.2	100%	6e-09	27.52%	154	P02144.2	<input checked="" type="checkbox"/>		

Run PSI-BLAST Iteration 2 with max number of sequences 500 Run 再次搜索

二次搜索得到 12 条序列，多了一条。

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	New added
RecName: Full=Hemoglobin subunit zeta; AltName: Full=HBAZ; AltName: Full=Hemoglobin zet... Homo sapiens	Homo sapiens	194	194	100%	3e-65	59.86%	142	P02008.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemogl... Homo sapiens	Homo sapiens	187	187	100%	2e-62	100.00%	142	P69905.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full=Hemoglobin subunit delta; AltName: Full=Delta-globin; AltName: Full=Hemogl... Homo sapiens	Homo sapiens	186	186	98%	5e-62	42.07%	147	P02042.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full=Hemoglobin subunit gamma-2; AltName: Full=Gamma-2-globin; AltName: Full=... Homo sapiens	Homo sapiens	186	186	98%	1e-61	39.31%	147	P69892.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-globin; AltName: Full=Hemoglob... Homo sapiens	Homo sapiens	184	184	98%	3e-61	42.07%	147	P68871.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full=Hemoglobin subunit theta-1; AltName: Full=Hemoglobin theta-1 chain; AltName: Full=... Homo sapiens	Homo sapiens	182	182	100%	3e-60	61.97%	142	P09105.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full=Hemoglobin subunit gamma-1; AltName: Full=Gamma-1-globin; AltName: Full=... Homo sapiens	Homo sapiens	182	182	98%	4e-60	39.31%	147	P69891.3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full=Hemoglobin subunit epsilon; AltName: Full=Epsilon-globin; AltName: Full=Hem... Homo sapiens	Homo sapiens	178	178	98%	8e-59	37.93%	147	P02100.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full=Hemoglobin subunit mu; AltName: Full=Hemoglobin mu chain; AltName: Full=... Homo sapiens	Homo sapiens	176	176	99%	4e-58	45.39%	141	Q6B0K9.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full=Myoglobin; AltName: Full=Nitrite reductase MB; AltName: Full=Pseudoperoxid... Homo sapiens	Homo sapiens	161	161	100%	7e-52	26.35%	154	P02144.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full=Cytoglobin; AltName: Full=Histoglobin; Short=HGb; AltName: Full=Nitric oxide... Homo sapiens	Homo sapiens	161	161	96%	2e-51	28.08%	190	Q8WWM9.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full=Neuroglobin; AltName: Full=Nitrite reductase [Homo sapiens]	Homo sapiens	57.8	57.8	93%	2e-11	22.54%	151	Q9NPG2.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

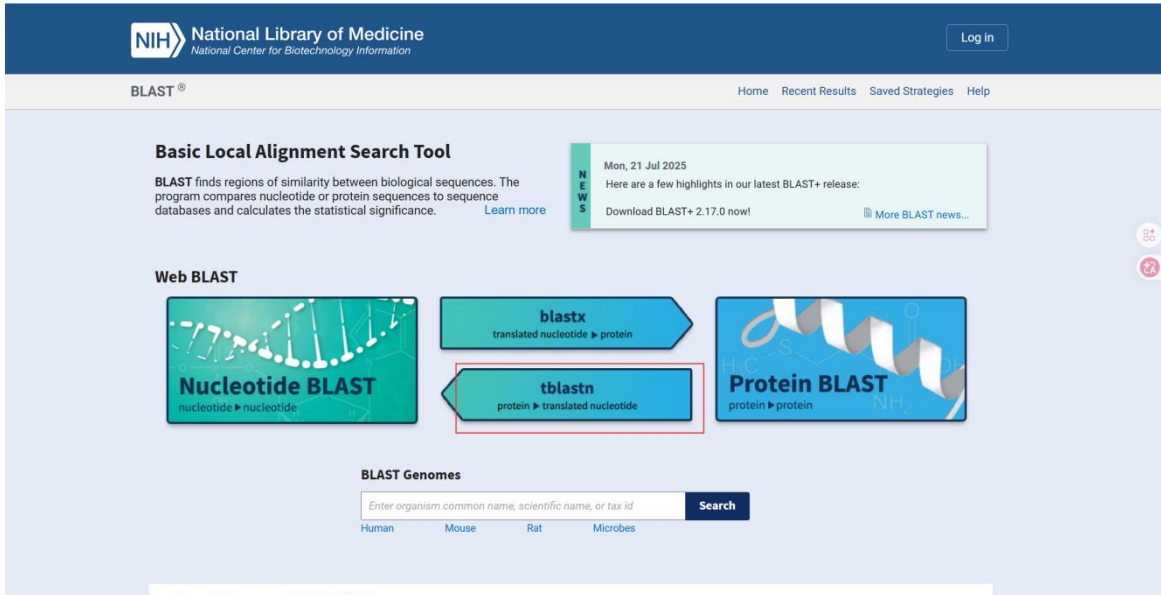
分析搜索结果，说明 PSI-BLAST 的基本原理。

PSI-BLAST 是用来寻找远程同源序列的工具，它会迭代搜索并构建一个位置特异性得分矩阵（PSSM）；PSI-BLAST 迭代：先找相似序列，总结每个位置的氨基酸偏好，再用偏好模式去搜远缘同源序列（比如第一轮搜索结果显示某一位置百分百是 H，那么这个位置只能是 H，不是 H 就打负分，某个位置可能是 L、V、I，那么只要这个位置是三者之一都可以），反复循环直到找不到新序列。

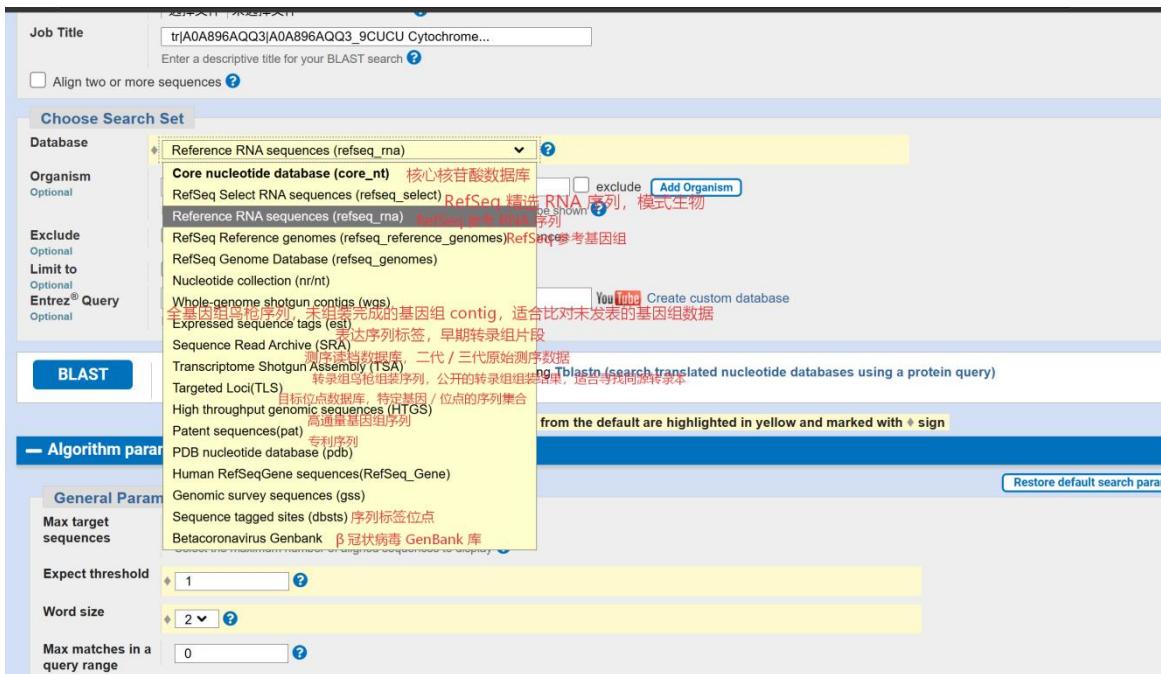
B2.4 mRNA 序列搜索

以双斑长跗萤叶甲的细胞色素 c 氧化酶亚基 I COX1 为搜索序列，用 NCBI tBlastN 搜索参考序列数据库 RefSeq 中的细胞色素 c 氧化酶家族 mRNA 序列，参数如下：期望阈值 E 为 1，字长 Word size 为 2，计分矩阵 Matrix 为 PAM250，空位罚分 Gap cost 为起始空位罚分 13，延伸空位罚分 2，其余参数默认。

首先进入 NCBI 的 BLAST 页面，打开 tblastn;



Enter Query Sequence 框输入双斑长跗萤叶甲的细胞色素 c 氧化酶亚基 I COX1 蛋白质序列，数据库选择 Reference RNA sequences (refseq_rna)



物种选择叶甲科 Chrysomelidae，期望阈值 E 为 1，字长 Word size 为 2，计分矩阵 Matrix 为 PAM250，空位罚分 Gap cost 为起始空位罚分 13，延伸空位罚分 2，其余参数默认。

Enter Query Sequence

Enter accession number(s), gi|id, or FASTA sequence(s) [Show](#) [Query subrange](#)

Align two or more sequences

Choose Search Set

Database: **Reference RNA sequences (refseq_rna)**

Organism: exclude [Add Organism](#)

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

Limit to: Sequences from type material

Entrez[®] Query: [Create custom database](#)

BLAST Search database Reference RNA sequences (refseq_rna) using Tblastn (search translated nucleotide databases using a protein query)

Show results in a new window

Note: Parameter values that differ from the default are highlighted in yellow and marked with a sign

Algorithm parameters

General Parameters

Max target sequences: 100

Expect threshold: 1

Word size: 2

Max matches in a query range: 0

Scoring Parameters

Matrix: PAM250

Gap Costs: Existence: 13 Extension: 2

Compositional adjustments: Conditional compositional score matrix adjustment

Filters and Masking

Filter: Low complexity regions

Mask: Mask for lookup table only Mask lower case letters

结果显示，在限定叶甲科（Chrysomelidae）物种的条件下，通过 tBLASTn 对比 RefSeq RNA 数据库，共获得 2 条显著匹配结果，均为玉米根萤叶甲（*Diabrotica virgifera virgifera*）的 COX1 同源 mRNA 序列；两条序列的 E 值分别为 1e-44 和 2e-28，均远低于设定阈值 1，Query 覆盖率分别为 100% 和 72%，序列一致性约 61.11%~62.02%，说明该序列与近缘叶甲科物种的 COX1 基因存在同源性，符合线粒体 COX1 基因的种间保守特征，可作为后续系统发育分析的参考依据。

NIH National Library of Medicine National Center for Biotechnology Information [Log in](#)

BLAST[®] » tblastn » results for RID-YMDNWU0014 [Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

[< Edit Search](#) [Save Search](#) [Search Summary](#) [How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

Information Your search is limited to records include: Chrysomelidae (taxid:27439)

Job Title: tr|A0A896AQQ3|A0A896AQQ3_9CUCU Cytochrome...

RID: YMDNWU0014 [Search expires on 04-24 23:11 pm](#) [Download All](#)

Program: TBLASTN [Citation](#)

Database: refseq_rna [See details](#)

Query ID: lcl|Query_2024203

Description: tr|A0A896AQQ3|A0A896AQQ3_9CUCU Cytochrome c oxid...

Molecule type: amino acid

Query Length: 179

Other reports: [?](#)

Filter Results

Organism: only top 20 will appear exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity: to E value: to Query Coverage: to

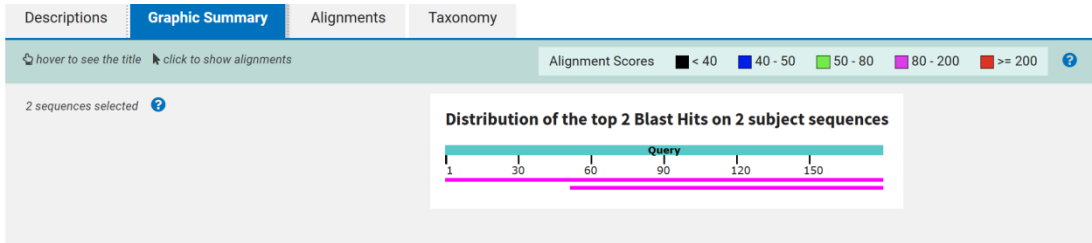
[Filter](#) [Reset](#)

Descriptions [Graphic Summary](#) [Alignments](#) [Taxonomy](#)

Sequences producing significant alignments [Download](#) [Select columns](#) [Show](#) 100 [?](#)

select all 2 sequences selected [GenBank](#) [Graphics](#)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
PREDICTED: Diabrotica virgifera virgifera cytochrome c oxidase subunit 1-like (LOC126893037). mR...	<i>Diabrotica virgifera virgifera</i>	144	144	100%	1e-44	61.11%	1728	XM_050662978.1
PREDICTED: Diabrotica virgifera virgifera cytochrome c oxidase subunit 1-like (LOC126893040). mR...	<i>Diabrotica virgifera virgifera</i>	104	104	72%	2e-28	62.02%	1940	XM_050662980.1



Descriptions **Alignments** Graphic Summary Taxonomy

Alignment view Pairwise Download

2 sequences selected

PREDICTED: Diabrotica virgifera virgifera cytochrome c oxidase subunit 1-like (LOC126893037), mRNA
 Sequence ID: [XM_050662978.1](#) Length: 1728 Number of Matches: 1

Range 1: 202 to 741 Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
144 bits(560)	1e-44	Compositional matrix adjust.	139/180(77%)	165/180(91%)	1/180(0%)	+1
Query 1	ILIRTELGSPGSLIGNDQIYNNVITAHAFimiffvmpimiggfQNLVPLMIGAFDMAF	60				
Sbjct 202	ILVRAELGSPGSLIG+DQIYNNVITAHAFI+ FF+V+PI+IGGFGN LVPL+IGAPD+AF	381				
Query 61	PRMNM+fwllppslfllIMKSIVESGAGTQTVYPP-LSSNIAHSGASVDLAIFSLHLA	119				
Sbjct 382	PR+NN+ F LLPPSL+LL+++ +VE GTG TVYPP LSSNIAH+G+SYDLAIF LHLA	561				
Query 120	GISSLGAINFITTIINRPFKMSMDRPLFWAWAITAIIIIspvLAGAITMLLTD	179				
Sbjct 562	GISSLGAINFITTVINIRPIGITFYRIFLFW*AVVITAILLLSLPVLGAIITILLTD	741				

Related Information
[Gene](#) - associated gene details
[Genome Data Viewer](#) - aligned genomic context

PREDICTED: Diabrotica virgifera virgifera cytochrome c oxidase subunit 1-like (LOC126893040), mRNA
 Sequence ID: [XM_050662980.1](#) Length: 1940 Number of Matches: 1

Range 1: 808 to 1194 Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
104 bits(401)	2e-28	Compositional matrix adjust.	98/129(76%)	116/129(89%)	1/129(0%)	+1
Query 52	MIGAPDAFPRMNM+fwllppslfllIMKSIVESGAGTQTVYPP-LSSNIAHSGASVD	110				
Sbjct 808	+IGAPD+AFPR+NN+ F LLPPSL+LL+++ +VE GTG TVYPP LSSNIAH+G+SVD	987				
Query 111	LAIFSLHLAGISSILGAINFITTIINRPFKMSMDRPLFWAWAITAIIIIspvLAG	170				
Sbjct 988	LAIF LHLAGISSILGAINFITTIIN+RP G++ R+PLFV AV+ITAILLLSLPVLG	1167				
Query 171	AITMLLTD 179					
Sbjct 1168	AIT+LLTD 1194					

Related Information
[Gene](#) - associated gene details
[Genome Data Viewer](#) - aligned genomic context

Reports **Lineage** Organism Taxonomy

2 sequences selected

Organism	Blast Name	Score	Number of Hits	Description
Diabrotica virgifera virgifera	beetles	144	<u>2</u>	Diabrotica virgifera virgifera hits

An official website of the United States government [Here's how you know](#)

NIH National Library of Medicine
National Center for Biotechnology Information

Log In

Nucleotide

GenBank

PREDICTED: Diabrotica virgifera virgifera cytochrome c oxidase subunit 1-like (LOC126893040), mRNA

NCBI Reference Sequence: XM_050662980.1
[FASTA](#) [Graphics](#)

LOCUS XM_050662980 1940 bp mRNA linear INV 19-SEP-2022
DEFINITION PREDICTED: Diabrotica virgifera virgifera cytochrome c oxidase subunit 1-like (LOC126893040), mRNA.
ACCESSION XM_050662980
VERSION XM_050662980.1
DBLINK BioProject: [PRJNA862183](#)
KEYWORDS RefSeq; corrected model; includes ab initio.
SOURCE Diabrotica virgifera virgifera (western corn rootworm)
ORGANISM [Diabrotica virgifera virgifera](#)
Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Altocrustacea; Allotricarida; Hexapoda; Insecta; Pterygota; Neoptera; Eumetabola; Endopterygota; Aparaglossata; Neuropteroides; Coleoptera; Polyphaga; Cucujiformia; Chrysomeloidea; Chrysomelidae; Galerucinae; Diabroticina; Diabroticites; Diabrotica.
COMMENT MODEL [REFSEQ](#): This record is predicted by automated computational analysis. This record is derived from a genomic sequence ([NC_065452](#)) 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 :: [Diabrotica virgifera virgifera](#)
Annotation Release ID1
Annotation Version :: 101
Annotation Pipeline :: NCBI eukaryotic genome annotation pipeline

More about the gene [LOC126893040](#)
LOC126893040 gene

Related information
BioProject
Protein
Taxonomy
Annotated Genomic
Gene
PubMed (RefSeq)

Nucleotide

GenBank

PREDICTED: Diabrotica virgifera virgifera cytochrome c oxidase subunit 1-like (LOC126893037), mRNA

NCBI Reference Sequence: XM_050662978.1
[FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS XM_050662978 1728 bp mRNA linear INV 19-SEP-2022
DEFINITION PREDICTED: Diabrotica virgifera virgifera cytochrome c oxidase subunit 1-like (LOC126893037), mRNA.
ACCESSION XM_050662978
VERSION XM_050662978.1
DBLINK BioProject: [PRJNA862183](#)
KEYWORDS RefSeq; corrected model; includes ab initio.
SOURCE Diabrotica virgifera virgifera (western corn rootworm)
ORGANISM [Diabrotica virgifera virgifera](#)
Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Altocrustacea; Allotricarida; Hexapoda; Insecta; Pterygota; Neoptera; Eumetabola; Endopterygota; Aparaglossata; Neuropteroides; Coleoptera; Polyphaga; Cucujiformia; Chrysomeloidea; Chrysomelidae; Galerucinae; Diabroticina; Diabroticites; Diabrotica.
COMMENT MODEL [REFSEQ](#): This record is predicted by automated computational analysis. This record is derived from a genomic sequence ([NC_065452](#)) 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 :: [Diabrotica virgifera virgifera](#)
Annotation Release ID1

More about the gene [LOC126893037](#)
LOC126893037 gene

Related information
BioProject
Protein
Taxonomy
Annotated Genomic
Gene
PubMed (RefSeq)

Alignment view 可选择不同方式

Alignment view Line length:

Query range 1: 1

Query
MW035260.1 60 XXXNWLVPLMIGAPDMAF
MW32714.1 1670 SFGN*LVPLIIGAPDIAF
NC_057489.1 1670 SFGN*LVPLIIGAPDIAF
MT178289.1 2832 SFGN*LVPLIIGAPDIAF
OJ_504952.1 1669 SFGN*LVPLIIGAPDIAF
MW419951.1 1491 ILIRTELGRPGSLIGNDQIVNVIVTAHAPIIFP IIVIP I IIGRGNWLVPLIIGAPDIAF
FP038016.1 26 NWLVPLIIGAPDIAF
FP038019.1 26 NWLVPLIIGAPDIAF
FP038014.1 26 NWLVPLIIGAPDIAF
FP038011.1 26 NWLVPLIIGAPDIAF
FP038015.1 26 NWLVPLIIGAPDIAF
FP038017.1 26 NWLVPLIIGAPDIAF
FP038018.1 26 NWLVPLIIGAPDIAF
FP038012.1 26 NWLVPLIIGAPDIAF
FP038013.1 26 NWLVPLIIRAPDIAF

B2.5 BLAST 算法部分概念强化

基于《序列数据库搜索系统 BLAST 简介》。

1) 举例说明动态规划算法(Dynamic Programming)和启发式算法 (Heuristic Programming) 的优缺点

这种利用动态规划算法进行双序列比对的方法,用于数据库搜索,找出数据库中所有与查询序列相似的目标序列,则需要将查询序列与数据库中每条序列都进行比对,计算过程严谨规范、结果准确度高,以 Needleman-Wunsch 双序列全局比对为典型应用;但该算法计算量大、运行耗时长,且需要储存大量矩阵信息,内存消耗较高,面对超长序列与海量数据时实用性受限。BLAST 采用启发式算法而非动态规划,搜索速度大为提高。典型代表为 BLAST 序列比对,运算速度快、资源消耗少,适合大规模生物序列快速检索;但该算法不严格追求全局最优,比对结果易受参数设置影响,存在一定假阴性与假阳性风险,结果稳定性和精准度弱于动态规划算法。

2) 简述数据库搜索灵敏度和特异度的含义

所谓灵敏度,是指从数据库中搜索到目标序列的多少;所谓特异度,是指搜索到的目标序列与查询序列是否具有较高相似性。高灵敏度和特异度,用通俗的话来说,就是“该找的都找到,找到的都是要找的”。

3) 期望值 E 和哪些因素有关

期望值 E 可以用公式 $E = kmN / e^{-\lambda S}$ 计算。因此, E 值大小与查询序列长度和数据库大小有关。

4) 简述新版 BLAST 延伸高分对的策略

1977 年发表的新版 BLAST 作了较大改进,高分对延伸方法也与旧版不同。新版 BLAST 在寻找高分三字串时,将阈值从 13 降为 11。降低匹配阈值后,所得三字串更多,数据库扫描时低分匹配 也更多,按理说,高分对延伸耗时也更多。

C 问题

Q1:判断一个基因是否与其他基因具有相似性功能,如何选择比对方法?

Q1 建议——如果是序列长度差异较大,可以选择局部对比(local alignment),更易找出两者之间的保守区域,如果长度相似可以采用全局对比(global alignment),找出匹配高的序列区域。

Q2:点阵图如何理解?

Q2 建议——基于《双序列比对基础和应用实例》中的说明，如果是自身对比，可以判断出重复序列的段数，如果是两条序列对比，对角线的缺口表示空位插入区域，更多解释或者理解需要更详细的结果解释说明参考书和参考文献。

Q3: 在 CNCB 中无法找到蚜虫相似性 CYP6CYP3-1 和 CYP6CYP3-1 蛋白质，uniprot 中显示这两个未经人工检查，像这样的应该如何查询

D 个人总结

G2A 边汉青

近段时间，学习了序列对比和 blast，主要获得了以下收获：

1. 了解认识到对比模型的多样性和其选用算法的基本原理。模型的选择影响输出结果，，例如 BLOSUM62、30、90，他们对于数据库中比对的灵敏度是不同的，90 的灵敏度最低，特异性最强；30 灵敏度最高，特异性最低。通常来说，大部分都适用于 62。核酸序列的比对，一般采用 DNAfull 计分矩阵的选择，而蛋白质序列比对常采用 BLOSUM62 和 PAM250 计分矩阵。

2. 明白了序列对比最重要的功能。可以通过序列对比，进行两个相似序列或多个序列的比对，找出他们共同的保守序列片段或者保守结构域，预测功能，建立系统进化树，确定亲缘关系以及确定先有基因还是先有物种等问题，为探索新基因功能提供方向。

3. 了解了数据库搜索中字长、矩阵和期望值对搜索结果的影响。在数据库中进行序列比对时，字长、矩阵和期望值共同对搜索结果起到影响。字长和打分矩阵主要在算法前期，通过控制“种子”的数量和质量，决定了搜索的灵敏度与广度；而期望值则在算法末端，作为一个统计标尺，最终筛选并输出我们认为具有生物学意义的结果。字长越短，灵敏度越高，找到结果的越多，同时假阳性概率提升；字长越长，灵敏度越低，特异性越强，找到的越少，假阳性概率降低相似度越高。

4. 利用点阵图直观表现序列比对结果。点阵图的结果可以显著直观的体现对比结果，具有较高相似性的同源序列，主要分布在主对角线上，而对角线两侧短线为随机匹配，有时也称为噪声。

5.序列比对的局限性。核酸序列相似度低，但可能其蛋白质序列高度匹配，另一方面可能序列不匹配，但其结构非常相似，例如流感病毒和唾液酶其序列不相似，但结构极其相似。

6.blast 的特殊性。如果有连续重复序列，只能通过 blast 才能查找出来，例如癌胚抗原 CEA5、CEA6,只有通过 blast 才能找出有连续重复序列。

7. 新基因出现方式和不同基因编码同一蛋白。当出现两个基因及相同编码统计蛋白质时，可能是源于在生物进化中基因的复制，而基因复制是创造新基因的一种方式，例如 HBA1 和 HBA2 都是编码 HBA_HUMAN 蛋白质，目前看来这两个基因是没有显著差异的具有高度同源，但在未来有可能会进行分化。

8. 查询序列和数据库序列。查找序列的种类和数据库的种类影响选择的比对方式。

例如，如果查询系列是蛋白质，而数据库是核酸，则应当选择 Tblastn。

G2B 刘奇

我们日常最常用的是 NCBI 的 BLAST，做快速同源性搜索最快，选 blastn（核酸）或 blastp（蛋白），结果里重点看“Identities”（相似度）和“Gaps”（缺口）。如果 NCBI 打不开或太慢，就用 CNCB（中国国家生物信息中心）的 BLAST，界面几乎一样，国内访问更稳定。

要是做蛋白功能注释，使用 Uniprot，输入序列后看“Function”和“Domains”板块，注释信息很直观，还能批量处理。

做双序列比对时，EBI 网站上的 Needle（全局比对，适合长度相近的序列）和 Water（局部比对，适合找保守区）各有用处，两序列整体像就用 Needle，只想知道哪里有相似片段就用 Water。

另外，BIT 这个平台除了常规的多序列比对和 BLAST 功能，还提供了点阵图工具。这个可视化功能很直观，把两条序列分别放在横纵轴上，相似的地方就会显示成一条对角线或点状图案。用它来检测序列内部的重复区域、倒位、缺失或者直接重复特别方便，比如想看某个基因里有没有串联重复序列，跑一张点阵图，一眼就能看出对角线上有没有平行的小短线，比看数字结果快得多。

多序列比对方面，课程做了四个实操：Uniprot 的 Align 工具适合 10 条以内蛋白序列的快速比对，结果用颜色标出保守位点；NCBI 的 COBALT 会同时参考蛋

白保守结构域数据库，适合序列差异较大或需要结构信息辅助的情况；EBI 的 T-Coffee 是把多种算法的结果综合起来，准确度高但速度慢，适合相似度低于 30% 的难比对任务；polydot 则是做多序列点阵图，能直观看出多条序列之间的共线性、保守区域和插入缺失分布。

日常工作可以先用 NCBI 的 BLAST 加 Uniprot，碰到特殊需求再去 EBI 或 CNCB 查漏补缺。

G2C 高倩

过去两周的学习主要围绕双序列比对，多序列比对，NCBI BLAST，CNCB BLAST 的基本概念学习和实例练习。

在序列比对学习章节，学习了双序列比对和多序列比对的多种方法。双序列比对时，用 BIT、PSA、NCBI 等不同平台，对棉蚜的 CYP6CY3-1 和 CYP6CY3-2 基因蛋白序列进行比对，学会用 Dotlet、dottup 生成点阵图，能直观看到两个序列的相似性。多序列比对则尝试了 Uniprot、NCBI 的 COBALT 和 EBI 的 T-Coffee 工具，也用 Polydot 做了多序列点阵图，慢慢摸清了不同工具的操作流程和结果解读方法。

之后重点学习了 BLAST 相关内容。先用 CNCB 的 blastp 对比了 CYP6CY3-1 和相似蛋白序列，验证了比对结果的一致性。在 NCBI BLAST 的实操中，做了实例练习，如搜索灵长目 alpha 血红蛋白、人的脑红蛋白，还用和自身课题相关的双斑长跗萤叶甲的 COX1 序列做了 tBLASTN 搜索。通过调整参数，我慢慢理解了 E 值、字长、计分矩阵等参数对结果的影响。

同时，进一步强化了 BLAST 相关算法概念，BLAST 用的是启发式算法，虽然不是全局最优，但速度更快。不过目前仍然存在不足，对部分参数的设置逻辑理解不够透彻，结果分析也不够深入，后续还需要多练习、多琢磨，把所学的实操方法和理论知识结合起来，提升自己的应用能力。