

双序列比对

日期：2022-11-2

实验者：生信 2001 张子栋

[MarkdownNotes/软件第2次作业.md at main · Bluuur/MarkdownNotes \(github.com\)](#)

[生物信息学原理/软件第2次作业.md · blur/MarkdownNotes - 码云 - 开源中国 \(gitee.com\)](#)

实验内容与目的

- 序列比对工具 BLAST 在线版的使用
 - Blastn, megablastn, dc megablastn
 - Blastp, PSI-blast, PHI-blast
 - blastx, tblastn, tblastx
 - 两序列比对 bl2seq
- 熟练使用序列相似性检索工具 BLAST 搜寻 NCBI 相关数据库
- 理解常用序列相似性检索工具 BLAST 的参数设置、结果解读与结果过滤
- 了解 BLAST 工具本地版的使用
- 了解 EBI 网站的序列相似性检索工具

实验流程和结果

1. 以大麦 Mlo 基因 (Z83834) 为查询序列

1. 用 Blastn 能在 nr/nt 数据库中检索到多少条与之同源的序列 (E-value<1e-30) ? 有多少条是禾本科中的?

■ 检索到 289 条

NIH National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastn suite » results for RID-PKZFJH3H013

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](#)

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 1e-30 Query Coverage to
[Filter](#) [Reset](#)

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 1000

☒ select all 289 sequences selected

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	H. vulgare mRNA for Mlo protein	Hordeum vulga...	3458	3458	100%	0.0	100.00%	1917	Z83834.1
<input checked="" type="checkbox"/>	PREDICTED: Hordeum vulgare subsp. vulgare protein MLO (LOC123449793). mRNA	Hordeum vulga...	3375	3375	100%	0.0	99.01%	2053	XM_045127125.1
<input checked="" type="checkbox"/>	Hordeum vulgare subsp. vulgare cDNA clone: FLbat5006. mRNA sequence	Hordeum vulga...	3358	3358	99%	0.0	98.95%	1906	AK248332.1
<input checked="" type="checkbox"/>	PREDICTED: Triticum aestivum protein MLO (LOC100037642). transcript variant X2. mRNA	Triticum aestivum	2344	2344	90%	0.0	90.22%	2007	XM_044515560.1
<input checked="" type="checkbox"/>	Triticum aestivum cDNA clone: WT005_M02 cultivar: Chinese Spring	Triticum aestivum	2341	2341	90%	0.0	90.21%	1896	AK333144.1
<input checked="" type="checkbox"/>	PREDICTED: Triticum dicoccoides protein MLO (LOC119303959). mRNA	Triticum dicocc...	2325	2325	87%	0.0	90.95%	1948	XM_037581117.1
<input checked="" type="checkbox"/>	Triticum aestivum seven transmembrane-spanning protein (Mlo2) mRNA, complete cds	Triticum aestivum	2323	2323	89%	0.0	90.15%	1730	AF361932.1
<input checked="" type="checkbox"/>	PREDICTED: Triticum aestivum protein MLO-like (LOC123098845). mRNA	Triticum aestivum	2321	2321	87%	0.0	90.83%	2117	XM_044520925.1

■ 有 133 条是禾本科中的

NIH National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastn suite » results for RID-PKZS1JSU01R

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](#)

Filter Results

Organism only top 20 will appear ☐ exclude
Gramineae (taxid:4479)
[Add organism](#)

Percent Identity to E value to 1e-30 Query Coverage to
[Filter](#) [Reset](#)

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 1000

☒ select all 133 sequences selected

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	H. vulgare mRNA for Mlo protein	Hordeum vulga...	3458	3458	100%	0.0	100.00%	1917	Z83834.1
<input checked="" type="checkbox"/>	PREDICTED: Hordeum vulgare subsp. vulgare protein MLO (LOC123449793). mRNA	Hordeum vulga...	3375	3375	100%	0.0	99.01%	2053	XM_045127125.1
<input checked="" type="checkbox"/>	Hordeum vulgare subsp. vulgare cDNA clone: FLbat5006. mRNA sequence	Hordeum vulga...	3358	3358	99%	0.0	98.95%	1906	AK248332.1
<input checked="" type="checkbox"/>	PREDICTED: Triticum aestivum protein MLO (LOC100037642). transcript variant X2. mRNA	Triticum aestivum	2344	2344	90%	0.0	90.22%	2007	XM_044515560.1
<input checked="" type="checkbox"/>	Triticum aestivum cDNA clone: WT005_M02 cultivar: Chinese Spring	Triticum aestivum	2341	2341	90%	0.0	90.21%	1896	AK333144.1
<input checked="" type="checkbox"/>	PREDICTED: Triticum dicoccoides protein MLO (LOC119303959). mRNA	Triticum dicocc...	2325	2325	87%	0.0	90.95%	1948	XM_037581117.1

2. 换用 megablast 或 discontinuous megablast, 观察检索结果的变化。

■ megablast, 结果为 64 条

NIH National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastn suite » results for RID-PKZY2TSZ01R

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](#)

1 Your results are filtered to match records that include: Gramineae (taxid:4479)
Your results are filtered to match records with expect value between 0 and 1e-30.

Job Title **Z83834:H.vulgare mRNA for Mlo protein**

RID [PKZY2TSZ01R](#) Search expires on 11-09 14:05 pm [Download All](#)

Program BLASTN [Citation](#)

Database nt [See details](#)

Query ID [Z83834.1](#)

Description H.vulgare mRNA for Mlo protein

Molecule type nucleic acid

Query Length 1917

Other reports [Distance tree of results](#) [MSA viewer](#)

Filter Results

Organism only top 20 will appear ☐ exclude
Gramineae (taxid:4479)
[Add organism](#)

Percent Identity to E value to 1e-30 Query Coverage to
[Filter](#) [Reset](#)

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 1000

☒ select all 64 sequences selected

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	H.vulgare mRNA for Mlo protein	Hordeum vulgar...	3541	3541	100%	0.0	100.00%	1917	Z83834.1
<input checked="" type="checkbox"/>	PREDICTED: Hordeum vulgare subsp. vulgare protein MLO (LOC123449793). mRNA	Hordeum vulgar...	3419	3419	100%	0.0	99.01%	2053	XM_045127125.1
<input checked="" type="checkbox"/>	Hordeum vulgare subsp. vulgare cDNA clone: FLbat5006. mRNA sequence	Hordeum vulgar...	3400	3400	99%	0.0	98.95%	1906	AK248332.1
<input checked="" type="checkbox"/>	PREDICTED: Triticum aestivum protein MLO (LOC100037642). transcript variant X2. mRNA	Triticum aestivum	2274	2274	90%	0.0	90.47%	2007	XM_044515560.1
<input checked="" type="checkbox"/>	Triticum aestivum cDNA clone: WT005_M02. cultivar: Chinese Spring	Triticum aestivum	2270	2270	90%	0.0	90.45%	1896	AK333144.1
<input checked="" type="checkbox"/>	PREDICTED: Triticum aestivum protein MLO-like (LOC123098845). mRNA	Triticum aestivum	2255	2255	87%	0.0	91.07%	2117	XM_044520925.1

■ discontinuous megablast, 结果为 129 条

NIH National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastn suite » results for RID-PKZY8V8P01R

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](#)

1 Your results are filtered to match records that include: Gramineae (taxid:4479)
Your results are filtered to match records with expect value between 0 and 1e-30.

Job Title **Z83834:H.vulgare mRNA for Mlo protein**

RID [PKZY8V8P01R](#) Search expires on 11-09 14:05 pm [Download All](#)

Program BLASTN [Citation](#)

Database nt [See details](#)

Query ID [Z83834.1](#)

Description H.vulgare mRNA for Mlo protein

Molecule type nucleic acid

Query Length 1917

Other reports [Distance tree of results](#) [MSA viewer](#)

Filter Results

Organism only top 20 will appear ☐ exclude
Gramineae (taxid:4479)
[Add organism](#)

Percent Identity to E value to 1e-30 Query Coverage to
[Filter](#) [Reset](#)

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 1000

☒ select all 129 sequences selected

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	H.vulgare mRNA for Mlo protein	Hordeum vulga...	3458	3458	100%	0.0	100.00%	1917	Z83834.1
<input checked="" type="checkbox"/>	PREDICTED: Hordeum vulgare subsp. vulgare protein MLO (LOC123449793). mRNA	Hordeum vulga...	3375	3375	100%	0.0	99.01%	2053	XM_045127125.1
<input checked="" type="checkbox"/>	Hordeum vulgare subsp. vulgare cDNA clone: FLbat5006. mRNA sequence	Hordeum vulga...	3358	3358	99%	0.0	98.95%	1906	AK248332.1
<input checked="" type="checkbox"/>	PREDICTED: Triticum aestivum protein MLO (LOC100037642). transcript variant X2. mRNA	Triticum aestivum	2344	2344	90%	0.0	90.22%	2007	XM_044515560.1
<input checked="" type="checkbox"/>	Triticum aestivum cDNA clone: WT005_M02. cultivar: Chinese Spring	Triticum aestivum	2341	2341	90%	0.0	90.21%	1896	AK333144.1
<input checked="" type="checkbox"/>	PREDICTED: Triticum dicoccoides protein MLO (LOC119303959). mRNA	Triticum dicocc...	2325	2325	87%	0.0	90.95%	1948	XM_037581117.1
<input checked="" type="checkbox"/>	Triticum aestivum seven transmembrane-spanning protein (Mlo2). mRNA. complete cds	Triticum aestivum	2323	2323	89%	0.0	90.15%	1730	AF361932.1

3. 尝试修改 Blastn 的参数，观测对检索结果的影响。

- 修改 Expect threshold 为 0.02, Word size 为 15

Program Selection

Optimize for

☐ Highly similar sequences (megablast)

☐ More dissimilar sequences (discontiguous megablast)

☒ Somewhat similar sequences (blastn)

Choose a BLAST algorithm ?

BLAST Search database Nucleotide collection (nr/nt) using Blastn (Optimize for somewhat similar sequences)

☒ Show results in a new window

Note: Parameter values that differ from the default are highlighted in yellow and

Algorithm parameters

General Parameters

Max target sequences 1000

Select the maximum number of aligned sequences to display ?

Short queries ☒ Automatically adjust parameters for short input sequences ?

Expect threshold 0.02

Word size 15

Max matches in a query range 0

Scoring Parameters

Match/Mismatch Scores 2,-3

Gap Costs Existence: 5 Extension: 2

- 得到 129 条结果

NIH National Library of Medicine National Center for Biotechnology Information

BLAST® » blastn suite » results for RID-PM04Z5XZ01R

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](#)

Your results are filtered to match records that include: Gramineae (taxid:4479)

Your results are filtered to match records with expect value between 0 and 1e-30.

Job Title Z83834:H.vulgare mRNA for Mlo protein

RID PM04Z5XZ01R Search expires on 11-09 14:09 pm [Download All](#)

Program BLASTN Citation

Database nt See details

Query ID Z83834.1

Description H.vulgare mRNA for Mlo protein

Molecule type nucleic acid

Query Length 1917

Other reports [Distance tree of results](#) [MSA viewer](#)

Filter Results

Organism only top 20 will appear ☐ exclude

Gramineae (taxid:4479)

[Add organism](#)

Percent Identity E value Query Coverage

to to 1e-30 to

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

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments

Download Select columns Show 1000

☒ select all 129 sequences selected

[GenBank](#) [Graphics](#) [Distance tree of results](#) [MSA Viewer](#)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> H.vulgare mRNA for Mlo protein	Hordeum vulga...	3458	3458	100%	0.0	100.00%	1917	Z83834.1
<input checked="" type="checkbox"/> PREDICTED: Hordeum vulgare subsp. vulgare protein MLO (LOC123449793). mRNA	Hordeum vulga...	3375	3375	100%	0.0	99.01%	2053	XM_045127125.1
<input checked="" type="checkbox"/> Hordeum vulgare subsp. vulgare cDNA clone: FLbat5006. mRNA sequence	Hordeum vulga...	3358	3358	99%	0.0	98.95%	1906	AK248332.1
<input checked="" type="checkbox"/> PREDICTED: Triticum aestivum protein MLO (LOC100037642). transcript variant X2. mRNA	Triticum aestivum	2344	2344	90%	0.0	90.22%	2007	XM_044515560.1
<input checked="" type="checkbox"/> Triticum aestivum cDNA clone: WT005_M02. cultivar: Chinese Spring	Triticum aestivum	2341	2341	90%	0.0	90.21%	1896	AK333144.1
<input checked="" type="checkbox"/> PREDICTED: Triticum dicoccoides protein MLO (LOC119070569). mRNA	Triticum dicocc...	2325	2325	87%	0.0	90.05%	1948	XM_037581117.1

4. 找出 Mlo 基因的编码蛋白序列，用 Blastp 检索到的与 Mlo 蛋白同源的序列与用 PSI-Blast 检索到的同源序列是否有差别？

■ 打开第一个搜索结果

BLAST® » blastn suite » results for RID-PM04Z5XZ01R

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 results are filtered to match records that include: Gramineae (taxid:4479)
Your results are filtered to match records with expect value between 0 and 1e-30.

Job Title **Z83834:H.vulgare mRNA for Mlo protein**

RID [PM04Z5XZ01R](#) Search expires on 11-09 14:09 pm [Download All](#)

Program BLASTN [Citation](#)

Database nt [See details](#)

Query ID [Z83834.1](#)

Description H.vulgare mRNA for Mlo protein

Molecule type nucleic acid

Query Length 1917

Other reports [Distance tree of results](#) [MSA viewer](#)

Filter Results

Organism only top 20 will appear ☐ exclude

Gramineae (taxid:4479)

[+ Add organism](#)

Percent Identity to E value to 1e-30 Query Coverage to

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

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

Alignment view [Pairwise](#) ☐ CDS feature [Restore defaults](#) [Download](#)

129 sequences selected

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

H.vulgare mRNA for Mlo protein

Sequence ID: [Z83834.1](#) Length: 1917 Number of Matches: 1

Range 1: 1 to 1917 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
3458 bits(3834)	0.0	1917/1917(100%)	0/1917(0%)	Plus/Plus

Query 1 GGCTGCTCCGCCAGCAAAACCAGACACACAGCAGCGTACCTGCGTACGTAGCGTGCGCTTT 60

Sbjct 1 GGCTGCTCCGCCAGCAAAACCAGACACACAGCAGCGTACCTGCGTACGTAGCGTGCGCTTT 60

■ 跳转至 GenBank

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

Alignment view [Pairwise](#) ☐ CDS feature [Restore defaults](#)

129 sequences selected

[Download](#) [GenBank](#) [Graphics](#)

H.vulgare mRNA for Mlo protein

Sequence ID: [Z83834.1](#) Length: 1917 Number of Matches: 1

Range 1: 1 to 1917 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
3458 bits(3834)	0.0	1917/1917(100%)	0/1917(0%)	Plus/Plus

Query 1 GGCTGCTCCGCCAGCAAAACCAGACACACAGCAGCGTACCTGCGTACGTAGCGTGCGCTTT 60

Sbjct 1 GGCTGCTCCGCCAGCAAAACCAGACACACAGCAGCGTACCTGCGTACGTAGCGTGCGCTTT 60

Query 61 CttttttttCCTTTCGCTCTCTTGCTTGTCCGGCCGCCACGTGATAGCCGGCCACG 120

Sbjct 61 CTTTTTTTTCCTTTCGCTCTCTTGCTTGTCCGGCCGCCACGTGATAGCCGGCCACG 120

Query 121 GCCAGGCACCTCGCGTTGCGTGCATCTGCGTGTGCGTACCTGGTAGAGCGGCC 180

Sbjct 121 GCCAGGCACCTCGCGTTGCGTGCATCTGCGTGTGCGTACCTGGTAGAGCGGCC 180

Query 181 GTCTGCTGTCTCCGGCAAGGAAGGAGTTGCGCGGTGACCGATGTCGGACAAAAAG 240

■ 复制氨基酸序列至 Blastp 搜索

- 二者结果不一致, PSI-blast 能利用初始 Blastp 得到的搜索结果, 构建 PSSM 进行迭代检索, 找出亲缘关系较远, 序列相似度较低的序列。

■ Blastp 结果

BLAST® » blastp suite » results for RID-PM0SZ1KV013

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](#)

Job Title **Protein Sequence**

RID [PM0SZ1KV013](#) Search expires on 11-09 14:19 pm [Download All](#)

Program Quick BLASTP [Citation](#)

Database nr [See details](#)

Query ID lcl|Query_84290

Description unnamed protein product

Molecule type amino acid

Query Length 533

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 to E value to Query Coverage to

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

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 500

☒ select all 500 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"/>	protein MLO [Hordeum vulgare subsp. vulgare]	Hordeum vulgare subsp. vulgare	1013	1013	100%	0.0	100.00%	533	XP_044983060.1
<input checked="" type="checkbox"/>	MLO [Triticum aestivum]	Triticum aestivum	889	889	100%	0.0	87.87%	534	AGT80151.1
<input checked="" type="checkbox"/>	protein MLO [Triticum dicoccoides]	Triticum dicoccoides	889	889	98%	0.0	89.02%	534	XP_037437014.1
<input checked="" type="checkbox"/>	protein MLO [Aegilops tauschii subsp. strangulata]	Aegilops tauschii subsp. strangulata	887	887	99%	0.0	88.74%	534	XP_020198001.1
<input checked="" type="checkbox"/>	MLO protein [Triticum aestivum]	Triticum aestivum	885	885	98%	0.0	88.83%	534	AAK60566.1
<input checked="" type="checkbox"/>	protein MLO-like isoform X2 [Triticum dicoccoides]	Triticum dicoccoides	883	883	99%	0.0	88.18%	534	XP_037426549.1

■ PIS-blast 结果

BLAST® » blastp suite » results for RID-PM0DR3X101R

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](#)

Job Title **Protein Sequence**

RID [PM0DR3X101R](#) Search expires on 11-09 14:13 pm [Download All](#)

Program PSI-BLAST Iteration 1 [Citation](#)

Database nr [See details](#)

Query ID lcl|Query_54545

Description unnamed protein product

Molecule type amino acid

Query Length 533

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 to E value to Query Coverage to

PSI-BLAST incl. threshold 0.005 [Filter](#) [Reset](#)

Run PSI-Blast iteration 2

Number of sequences 500 [Run](#)

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 500

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

Sequences with E-value BETTER than threshold

☒ select all 500 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	Newly added
<input checked="" type="checkbox"/>	protein MLO [Hordeum vulgare subsp. vulgare]	Hordeum vulgare...	1106	1106	100%	0.0	100.00%	533	XP_044983060.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	hypothetical protein ZWY2020_035246 [Hordeum vulgare]	Hordeum vulgare	1093	1093	100%	0.0	98.15%	541	KAI4995343.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	protein MLO [Triticum dicoccoides]	Triticum dicoccoides	969	969	98%	0.0	89.02%	534	XP_037437014.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	protein MLO [Aegilops tauschii subsp. strangulata]	Aegilops tauschii...	969	969	99%	0.0	88.74%	534	XP_020198001.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	MLO [Triticum aestivum]	Triticum aestivum	968	968	100%	0.0	87.87%	534	AGT80151.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

5. 使用 BlastX 预测 Mlo 基因的编码蛋白。

■ 下载 FASTA 格式文件

Job Title

Z83834:H.vulgare mRNA for Mlo protein

RID

PM04Z5XZ01R Search expires on 11-09 14:09 pm [Download All](#) ▼

Program

BLASTN [?](#) [Citation](#) ▼

Database

nt [See details](#) ▼

Query ID

[Z83834.1](#)

Description

H.vulgare mRNA for Mlo protein

Molecule type

nucleic acid

Query Length

1917

Other reports

[Distance tree of results](#) [MSA viewer](#) [?](#)

Filter Results

Organism only top 20 will appear

Gramineae (taxid:4479)

[+ Add organism](#)

Percent Identity

to

E value

to

Descriptions

Graphic Summary

Alignments

Taxonomy

Alignment view

Pairwise

☐ CDS feature [?](#) [Restore defaults](#)

129 sequences selected [?](#)

Download ▼

GenBank

Graphics

☒ FASTA (complete sequence)
 ☐ FASTA (aligned sequences)
 ☐ GenBank (complete sequence)
 ☐ Text (aligned sequences)

Continue

Cancel

Number of Matches: 1

▼ Next Match

▲ Previous Match

Accession	Length	Gaps	Strand
Z83834.1	1917(100%)	0/1917(0%)	Plus/Plus

Query 61

CTTTTTCCTTCGCTCTCTGCTGCTCCGCCGCCACGTCGATAGCCGCCACG

Sbjct 61

CTTTTTCCTTCGCTCTCTGCTGCTCCGCCGCCACGTCGATAGCCGCCACG

■ 在 blastx 中输入序列

blastn

blastp

blastx

tblastn

tblastx

Translated BLAST: blastx

BLASTX search protein databases using a translated nucleotide query. [more...](#)

Enter Query Sequence

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

CCTTCAGCCAGGGATGAGACAAGTTTCTGTA

TTCATGTTAGTCCCAATGTATAGCCAACATAGGATGTGATGATTCGTACAAT

AAGAAATACAATTTTACTGAGTC

Or, upload file

选择文件 未选择文件 [?](#)

Genetic code

Standard (1) ▼

Job Title

Z83834.1 H.vulgare mRNA for Mlo protein

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

☐ Align two or more sequences [?](#)

Query subrange [?](#)

From

To

Choose Search Set

Database

Non-redundant protein sequences (nr) ▼ [?](#)

Organism

Optional

Enter organism name or id—completions will be suggested

☐ 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

BLAST

Search database nr using Blastx (search protein databases using a translated nucleotide query)

☐ Show results in a new window

+ Algorithm parameters

■ 第一个结果预测编码蛋白相同

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

Job Title

Z83834.1 H.vulgare mRNA for Mlo protein

RID

PM102BV3013

Search expires on 11-09 14:23 pm

Download All

Program

BLASTX

Citation

Database

nr

See details

Query ID

lcl|Query_35300

Description

Z83834.1 H.vulgare mRNA for Mlo protein

Molecule type

dna

Query Length

1917

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

100 sequences selected

GenPept

Graphics

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	protein.MLO [Hordeum vulgare subsp. vulgare]	Hordeum vulgare subsp. vulgare	1011	1011	83%	0.0	100.00%	533	XP_044983060.1
<input checked="" type="checkbox"/>	hypothetical protein ZWY2020_035246 [Hordeum vulgare]	Hordeum vulgare	999	999	83%	0.0	98.15%	541	KAH4995343.1
<input checked="" type="checkbox"/>	MLO [Triticum aestivum]	Triticum aestivum	887	887	83%	0.0	87.87%	534	AGT80151.1
<input checked="" type="checkbox"/>	protein.MLO [Triticum dicoccoides]	Triticum dicoccoides	887	887	82%	0.0	89.02%	534	XP_037437014.1
<input checked="" type="checkbox"/>	protein.MLO-like [Triticum aestivum]	Triticum aestivum	886	886	83%	0.0	88.56%	534	XP_044376860.1
<input checked="" type="checkbox"/>	protein.MLO [Aegilops tauschii subsp. strangulata]	Aegilops tauschii subsp. strangulata	885	885	83%	0.0	88.74%	534	XP_020198001.1
<input checked="" type="checkbox"/>	MLO protein [Triticum aestivum]	Triticum aestivum	884	884	82%	0.0	88.83%	534	AAK60566.1
<input checked="" type="checkbox"/>	protein.MLO-like isoform X2 [Triticum dicoccoides]	Triticum dicoccoides	882	882	83%	0.0	88.18%	534	XP_037426549.1

2. 用 bl2seq 分析大麦和小麦 Mlo 基因 mRNA 序列编码区和蛋白质产物的同源性

○ 大麦氨基酸序列

gene

CDS

ORIGIN

```

/db_xref= taxon:112509
/chromosome="4H"
<1..>1899
/gene="LOC123449793"
/note="Derived by automated computational analysis using
gene prediction method: Gnomon. Supporting evidence
includes similarity to: 2 mRNAs, 35 ESTs, 26 Proteins, and
100% coverage of the annotated genomic feature by RNAseq
alignments, including 41 samples with support for all
annotated introns"
/db_xref="GeneID:123449793"
208..1809
/gene="LOC123449793"
/codon_start=1
/product="protein MLO"
/protein_id="XP_044983060.1"
/db_xref="GeneID:123449793"
/translation="MSDKKGVPARELPETPSWAVVFAAMVLVSVLMEHGLHKLGHW
FQHRHKALWEALEKMAELMLVGFISLLIVTQDPIIAKICISEDADVMWPCKRGT
EGRKPSKYVDYCEGKVALMSTGSLHQLHVFIFVLAVFHTVYSVITIALSRMKMRTWK
KWETETTSLEYQFANDPARFRFTHQTSFVKRHLGLSSTPGIRWVVAFFRQFFRSVTKV
DYLTLRAGFINAHLNQSKFDFHKYIKRSMEDDFKVVVGISLPLWGVAILTLFLDING
VGTLIWIWISFIPLVILLCVGKLEMIIMEMALEIQDRASVIKGAAPVVEPSNKKFFWFHRP
DWVLFIFIHLTLFQNAFQMAHFVWTVATPGLKKCYHTQIGLSIMKVVVGLALQFLCSYM
TFPLYALVTQMGSNMKRSIFDEQTSKALTNRWNTAKEKKKVRDMDLMAQMIQDATPS
RGSSPMPSRGSSPVHLLHKGMRSDDPQSAPTSPTQQEARDMPVVVAHPVHRLNPN
DRRRSASSALEADIPSADFSFSQG"

```


○ 小麦氨基酸序列

CDS

100% coverage of the annotated genomic feature by RNAseq alignments, including 45 samples with support for all annotated introns”
/db_xref=“GeneID:100037642”
36..1640
/gene=“LOC100037642”
/codon_start=1
/product=“protein MLO isoform X2”
/protein_id=“XP_044371495.1”
/db_xref=“GeneID:100037642”
/translation=“MADDDEYPPARTLPETPSWAVALVFAVMIIVSVLLEHALHKLGH
WFHKKRHNALAELEKIKAEMLVGFISLLAVTQDPISGICISEKAASIMRPCKLPP
GSVKSKYKDYYCAKQGVSLMSTGSLHQLHIFIVLAVFHVTVSVIIMALSRLKMRTW
KKWETETASLEYQFANDPARFRFTHQTSFVKRHLGLSSTPGVRWVVAFFRQFFRSVT
VDYLTLAGFINAHLSHNSKDFHFKYIKRSMEDDFKVVVGISLPLWCVAITLFLDID
GIGTLTWISFIPLVILLCVGTKLEMIEMALEIQDRASVIKAPVVEPSNKKFFWFHR
PDWVLF IHLTLFQNAFQMAHFVWTVATPGLKKCFMHIGLSIMKVVLGLALQFLCSY
ITFPLYALVTQMSNMKRSIFDEQTAKALTNWRNTAKEKKKVRDMDLMAQMIGDATP
SRGASPMPSRGSSPVHLLHKGMRSDDPQSTPTSPRAMEEARDMYPVVVAHPVHRLNP
ADRRRSVSSSALDVDIPSADFSFSQG”

ORIGIN

○ 在 Blastp 中比对

Alian Sequences Protein BLAST

blastnblastpblastxtblastntblastx

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

Reset pageBookmark

Enter Query Sequence

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

MSDKKGVPAELPETPSWAVAVVFAAMVLVSVLMEHGLHKLGHW
FQHRHKALWEALEKMKAEMLVGFISLLIVTQDPPIAKICISEDAADVWPC
KRG

Query subrange ?

From

To

Or, upload file

选择文件未选择文件 ?

Job Title

Enter a descriptive title for your BLAST search ?

☒ Align two or more sequences ?

Enter Subject Sequence

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

SRGASPMPSRGSSPVHLLHKGMRSDDPQSTPTSPRAMEEARDMYPVVV
AHPVHRLNP
ADRRRSVSSSALDVDIPSADFSFSQG

Subject subrange ?

From

To

Or, upload file

选择文件未选择文件 ?

Program Selection

Algorithm

☒ blastp (protein-protein BLAST)
Choose a BLAST algorithm ?

BLAST

Search protein sequence using Blastp (protein-protein BLAST)
☐ Show results in a new window

- 相似性为 88.18%

[Filter Search](#)
[Save Search](#)
[Search Summary](#)

[How to read this report](#)
[BLAST Help Videos](#)
[Back to Traditional Results Page](#)

Job Title

Protein Sequence

RID

[PM4APB4N114](#)
Search expires on 11-09 15:20 pm
[Download All](#)

Program

Blast 2 sequences [Citation](#)

Query ID

Ic|Query_38523 (amino acid)

Query Descr

unnamed protein product

Query Length

533

Subject ID

Ic|Query_38525 (amino acid)

Subject Descr

None

Subject Length

534

Other reports

[Multiple alignment](#)
[MSA viewer](#)

Filter Results

Percent Identity

to

E value

to

Query Coverage

to

Filter

Reset

Descriptions

Graphic Summary

Alignments

Dot Plot

Sequences producing significant alignments

Download

Select columns

Show

100

☒ select all
 1 sequences selected

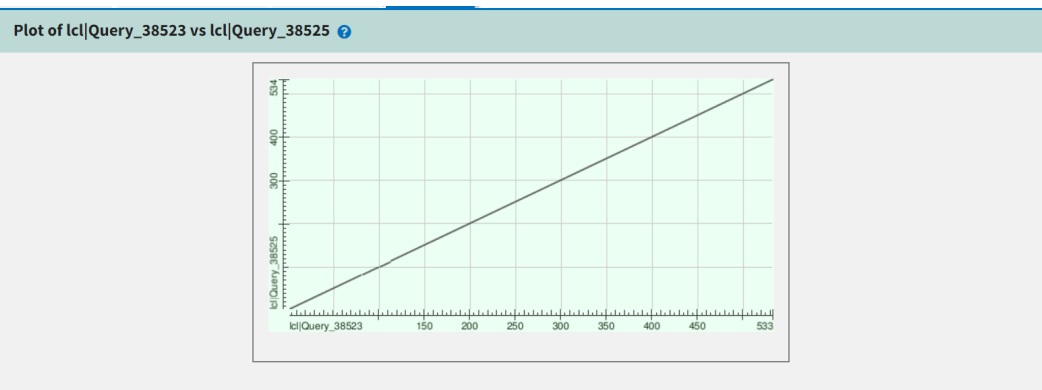
[Graphics](#)
[Multiple alignment](#)
[MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	unnamed protein product		965	965	99%	0.0	88.18%	534	Query_38525

Range 1: 4 to 534 [Graphics](#)

▼ [Next Match](#) ▲ [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
965 bits(2495)	0.0	Compositional matrix adjust.	470/533(88%)	492/533(92%)	4/533(0%)
Query	3	DKKGVPARELPETPSWAVAVFVFAAMVLVSVLMEIHLIKLGHWFQRIKKAALWEALEKMK			62
Sbjct	4	D + PAR LPETPSWAVA+VFA M++VSVL+EII LIKLGHWF RIK AL EALEK+KA DDEYPPTARTLPETPSWAVAVFVAVMI VSVLLEIALHLKLGHWFIKRIKNALEALEK IKA			63
Query	63	ELMLVGFISLLLTIVTQDPIIAKICISEDAADVMPCKRGTEGRKPSKYDY—CPEGKVA			120
Sbjct	64	ELMLVGFISLLLTIVTQDPI + ICISE A+M PCK G SKY DY +GKV+ ELMLVGFISLLLAVTQDPI-SGICISEKAASIMRPCKL-PPGSVKSKYDYCAKQGVKS			121
Query	121	LMSTGSLIHLHVFIFVLAVFHVITYSVITIALSRCLKMRTWKKWETETTSLEYQFANDPARF			180
Sbjct	122	LMSTGSLIHLHVFIFVLAVFHVITYSVI +ALSRCLKMRTWKKWETET SLEYQFANDPARF LMSTGSLIHLHVFIFVLAVFHVITYSVIIMALSRCLKMRTWKKWETETASLEYQFANDPARF			181
Query	181	RFTIHQTSFVKRIHLGLSSTPGIRWVVAFFRQFFRSVTKVDYLTLAGFINAHLNSKSFDF			240
Sbjct	182	RFTIHQTSFVKRIHLGLSSTPG+RWVVAFFRQFFRSVTKVDYLTLAGFINAHL NSKIFDF RFTIHQTSFVKRIHLGLSSTPGVRWVVAFFRQFFRSVTKVDYLTLAGFINAHLNSKSFDF			241
Query	241	IKYIKRSMEDDFKVVVGISLPLWGVAILTLFLDINGVGTLIWISFIPLVILLCVGKTLEM			300
Sbjct	242	IKYIKRSMEDDFKVVVGISLPLW VAILTLFLD1+G+GTL WISFIPLVILLCVGKTLEM IKYIKRSMEDDFKVVVGISLPLWCVAAILTLFLD1DGLTGLTWISFIPLVILLCVGKTLEM			301
Query	301	IIMEMALEIQDRASVIGKAPVVEPSNKFVFWHRPDWVLFIIILTLFQNAFQMAHFVWTVA			360
Sbjct	302	IIMEMALEIQDRASVIGKAPVVEPSNKFVFWHRPDWVLFIIILTLFQNAFQMAHFVWTVA IIMEMALEIQDRASVIGKAPVVEPSNKFVFWHRPDWVLFIIILTLFQNAFQMAHFVWTVA			361
Query	361	TPGLKKCYHTIGLSIMKVVVGLALQFLCSYMTPLVALVTQMGSNMKRSIFDEQTSKAL			420
Sbjct	362	TPGLKKC+II IGLSIMKV+GLALQFLCSY+TPPLVALVTQMGSNMKRSIFDEQT+KAL TPGLKKCFIMIHLGLSIMKVVVGLALQFLCSYITPLVALVTQMGSNMKRSIFDEQTAKAL			421
Query	421	TNWRNTAKEKKKVRDMDMLMAQIMGATPSRGSSPMPSRGSSPVHLLIKMGSRSDDPQSA			480
Sbjct	422	TNWRNTAKEKKKVRDMDMLMAQIMGATPSRG+SPMPSRGSSPVHLLIKMGSRSDDPQS TNRNTAKEKKKVRDMDMLMAQIMGATPSRGSSPMPSRGSSPVHLLIKMGSRSDDPQST			481
Query	481	PTSPRTQEQEARDMPVVAIVPVIHRLNPDRRRSSASSALEADIPSADFSFSQG			533
Sbjct	482	PTSPR +EARDMPVVAIVPVIHRLNP DRRRS SSSAL+ DIPSADFSFSQG PTSPRAMEARDMPVVAIVPVIHRLNPDRRRSSSALDVIDPSADFSFSQG			534



- mRNA 序列编码区的同源性

◦ 下载 FASTA 文件

FASTA

PREDICTED: Hordeum vulgare subsp. vulgare protein MLO (LOC123449793), mRNA

NCBI Reference Sequence: XM_045127125.1

Send to: ☒ Selected region

Customize view:

Articles about the LOC123449793 gene

Development of 500k full-length cDNAs in barley: a tool for screening. [Zhu, Hui, 2009]

The barley Mlo gene: a novel control element of plant pathogen resistance. [Cull, 1997]

See all...

Reference sequence information

RefSeq protein product

See the reference protein sequence for protein MLO (XP_044893080.1)

More about the gene LOC123449793

J. LOC123449793.jvarkit

FASTA

PREDICTED: Triticum aestivum protein MLO (LOC100037642), transcript variant X2, mRNA

NCBI Reference Sequence: XM_044515560.1

Send to: ☐ Whole sequence ☒ Selected region

Customize view:

Articles about the LOC100037642 gene

Comparative Temporal Transcription Profiling of Wheat and Rice (Plant Sci. 2014)

Assessment of adaptive evolution between wheat and rice as de [BMC Genomics. 2009]

Molecular phylogeny and evolution of the plant-specific gene [J. Mol. Evol. 2003]

See all...

Reference sequence information

RefSeq alternative splicing

See the other reference mRNA sequence entries for the LOC100037642 gene (XM_044515560.1)

RefSeq protein product

◦ 在 blastn 中比对

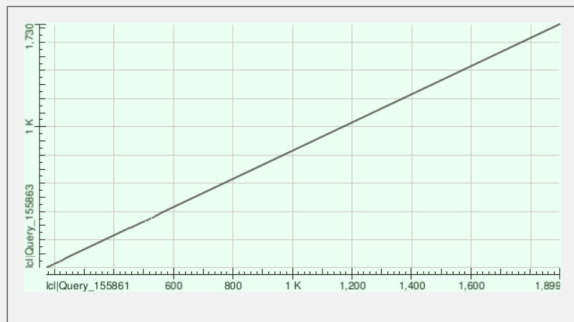
mRNA 序列编码区的同源性为 91%

XM_044515560.1:263-1992 PREDICTED: Triticum aestivum protein MLO (LOC100037642), transcript variant X2, mRNA
Sequence ID: Query_155863 Length: 1730 Number of Matches: 1

Range 1: 1 to 1730 [Graphics](#) [▼ Next Match](#) [▲ Previous Match](#)

Score	2283 bits(1236)	Expect	0.0	Identities	1576/1740(91%)	Gaps	23/1740(1%)	Strand	Plus/Plus
Query	173	CTCCGGGCAAGGAAGGAGTTGCG-GCGGTCGACGATGTCGGACAAAAAGGGT-GCC	230						
Sbjct	1	CTCCGGGCAAGGAAGGAGTTGCGCTGG-GGACCGATGGCGGAC-GACGACGATACCC	58						
Query	231	GGC-GCG-GGA-GCTGCCGAGAGCGCGCTCGTGGCGGTGGCGGTGGTCTTCGCCGCAT	287						
Sbjct	59	CCCAGCGAGGAGCGTCCCGGAGAGCGCGCTCGTGGCGGTGGCCCTCGTCTCGCGGTAT	118						
Query	288	GGTGTCTGTTCCGTCCTCATGGAGCAGCGCTCCAAAGTCGGCCATTGGTTCACGA	347						
Sbjct	119	GATCATCGTTCGCTCCTCGGAGCAGCGCTCCAAAGTCGGCCATTGGTTCACAA	178						
Query	348	CCGGCACAAGAAGCCCTGTGGGAGGCGCTGGAGAAGTGAAGCGGAGCTCATGCTGT	407						
Sbjct	179	GCGGCACAAGAAGCGCTGTGGGAGGCGCTGGAGAAGTGAAGCGGAGCTCATGCTGT	238						
Query	408	GGGCTTCATATCCTGCTCCTCATGTCACGAGGACCCCATCATGCGCAAGATATGCAT	467						
Sbjct	239	GGGCTTCATCTGCTGCTGCTCGCGTGACGAGGACCCATC-T-CCGGGATATGCAT	295						
Query	468	CTCCGAGGATGCCCGGACGCTATGTGGCCCTGCAAGCGCGGACCGAGG-GCCG-CAAG	525						
Sbjct	296	CTCCGAGAAGGCCGCCAGCATCATGCGGCCCTGCAAGG-TGCCCCCTGGCTCGTCAAG	353						
Query	526	CCCAGCAAGTAC---GTTGACTACTGC-CC-GGAGGCAAGTGGCGCTCATGTCCACG	579						
Sbjct	354	---AGCAAGTACAAAGACTACTGCGCCAAACAGGGCAAGGTGTGCTCATGTCCACG	410						
Query	580	GGCAGCTTGCACCAAGTGCAGGCTTTCATCTTCGTGCTCGCGGTCTTCATGTACACCTAC	639						
Sbjct	411	GGCAGCTTGCACCAAGTGCACATATTCATCTTCGTGCTCGCGGTCTTCATGTACACCTAC	470						
Query	640	AGCGTCATCACCATAGCTCTAAGCCGCTCTCAAAATGAGAACATGGAAGAAATGGGAGACA	699						
Sbjct	471	AGCGTCATCATCATGGCTCTAAGCCGCTCTCAAAATGAGAACCTGGAAGAAATGGGAGACA	530						

Plot of lcl|Query_155861 vs lcl|Query_155863 ?



3. 总结有哪些方法可以加快 Blast 检索速度？有哪些方法可以降低 Blast 结果的假阳性？

◦ 加快 Blast 检索速度的方法有：

1. 多线程进行 Blast
2. 增大 word 的值
3. 设置并限制比对方向

◦ 降低 Blast 结果假阳性的方法有：

1. 屏蔽重复性的低复杂度区域
2. 选取 E-value 更小的结果
3. 更严苛的打分和罚分规则

讨论

BLAST 是一套在蛋白质数据库或 DNA 数据库中进行相似性比较的分析工具。BLAST 程序能迅速与公开数据库进行相似性序列比较。