microRNA的搜索

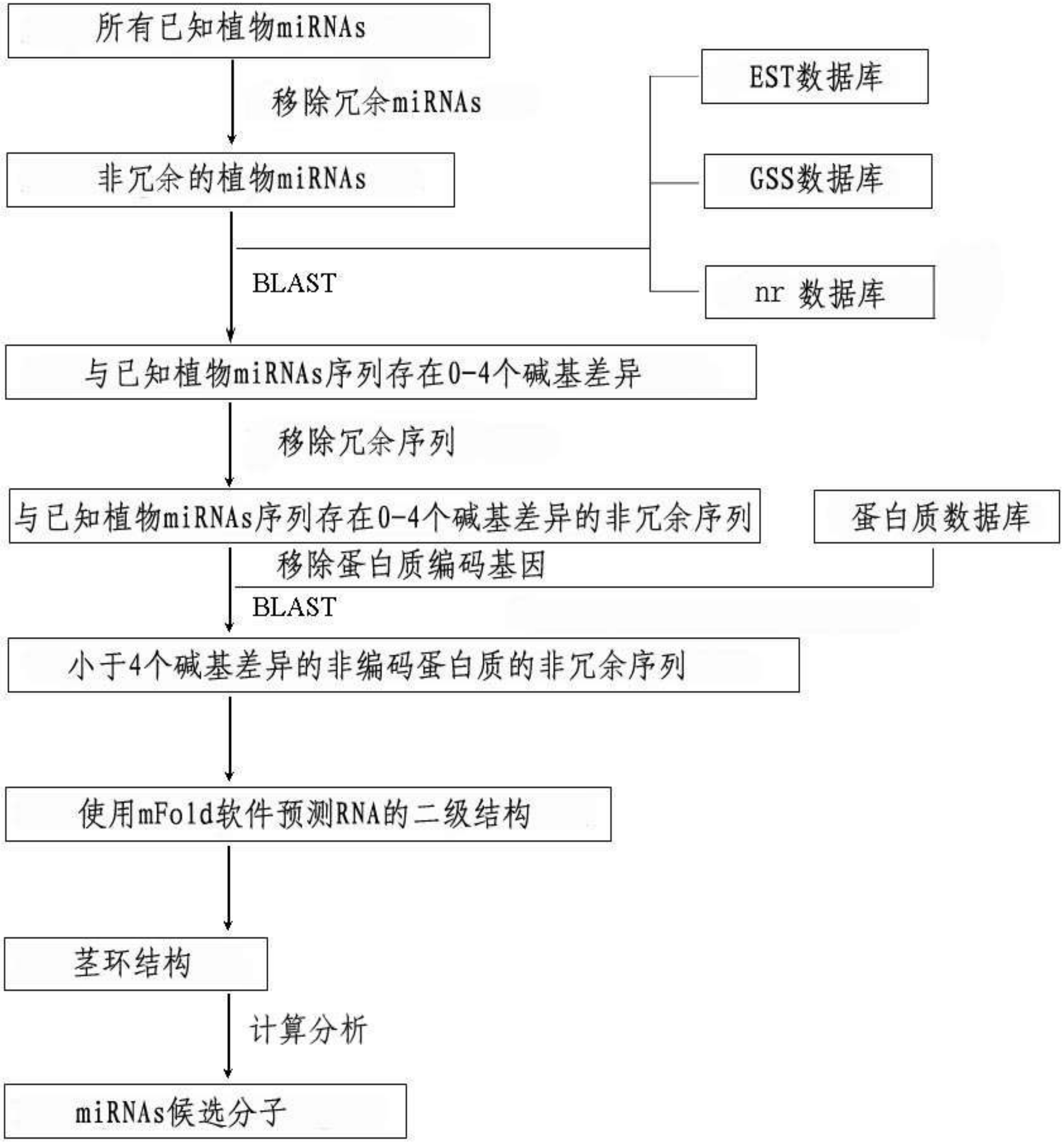
南昌大学生物学实验教学中心 汪艳璐

miRNA的预测

1、对已知植物miRNA进行分析与筛选，将非冗余的miRNA数据作为“种子序列”，应用BLAST N程序在小麦nr数据库、EST数据库及GSS数据库进行搜寻，获得可能的miRNA序列。允许结果序列与“种子序列”存在小于4个碱基差异，但不允许存在缺口。

2、将筛选出来的小于4个碱基差异的候选miRNA序列，再次进行BLAST分析(如果两条序列相似度>98%，我们认为它们为同一序列)，移除一些重复的序列。然后将非冗余的候选miRNA序列与蛋白质数据库进行BLAST X分析，以排除一些编码蛋白基因序列。

3、将筛选出来含有小于4个碱基差异的非编码蛋白质的序列，采用MFold3.2折叠程序检测其最小自由能以及判断其是否具有miRNA的典型的茎环结构特征，即选取候选miRNA所在的基因组位点两侧60~300 nt的序列，利用MFold软件进行二级结构的折叠，MFold的执行参数为默认设置。最后计算其MFEI值，将所得的结果数据保存，以便后续的数据分析


所有已知植物miRNAs
移除冗余miRNAs
非冗余的植物miRNAs
EST数据库
GSS数据库
nr 数据库
BLAST
与已知植物miRNAs序列存在0-4个碱基差异
移除冗余序列
与已知植物miRNAs序列存在0-4个碱基差异的非冗余序列
蛋白质数据库
移除蛋白质编码基因
BLAST
小于4个碱基差异的非编码蛋白质的非冗余序列
使用mFold软件预测RNA的二级结构
茎环结构
计算分析
miRNAs候选分子

miRNA靶基因的预测

对于筛选出来的马铃薯miRNA候选分子，我们使用miRU软件对其靶基因进行预测。另外，我们还使用BLASTN在GenBank数据库中搜索其可能的靶分子。将筛选出来的靶分子，利用BLAST X程序在相关蛋白质数据库中进行同源搜索，预测其靶基因可能的功能。

常用的数据库及分析软件

1、GenBank的非冗余数据库、表达序列标签数据库及基因鉴定序列数据库 (http://www.ncbi.nlm.nih.gov)；

2、miRNA数据库 (http://www.mirbase.org/)；

3、JCVI的植物基因组数据库 (http://www.tigr.org/plantProjects.shtml)；

4、比对软件BLAST N和BLAST X (http://www.ncbi.nlm.nih.gov/BLAST/)；

5、RNA折叠软件MFold3.2 (http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi)；

6、NCBI 蛋白质数据库 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein)；

7、植物miRNA预测软件MicroHARVESTER (http://www-ab.informatik.uni-tuebingen.de/brisbane/tb/index.php?view=index)；

8、拟南芥与水稻功能蛋白数据库 (http://ppdb.gene.nagoya-u.ac.jp/)；

9、靶基因预测软件miRU (http://bioinfo3.noble.org/miRNA/miRU.htm)；

10、多序列比对分析软件Clustal W (ftp://ftp.ebi.ac.uk/pub/software/dos/clustalw/)；

11、其他分析软件：BioEdit、Oligo 6.0、Primer Premier 5、Lasergene 6、Paup软件等。

miRBase

Latest miRBase blog posts

Bug fixes to release 20 MySQL database dumps By sam (July 17, 2013)

Read no further unless you care about the MySQL database dumps in the database_files directory on the FTP site. A couple of people (many thanks Jeff and Jakob) found errors in the the release 20 MySQL database dumps: a small number of new mature sequences were not linked to their hairpin precursors, and the ends [...]

miRBase 20 released By sam (June 24, 2013)

Phew. After considerably more pain and tears than usual, miRBase 20 is finally available on the website and for download on the FTP site (see also the README file). The gap between releases has also been longer than usual, which means that the increase in data is greater than usual (probably explaining the increase in [...]

miRNA count: 24521 entries

Release 20: June 2013

Search by miRNA name or keyword

Download published miRNA data

Download page | FTP site

This site is featured in:

NetWatch - *Science* 303:1741 (2004)

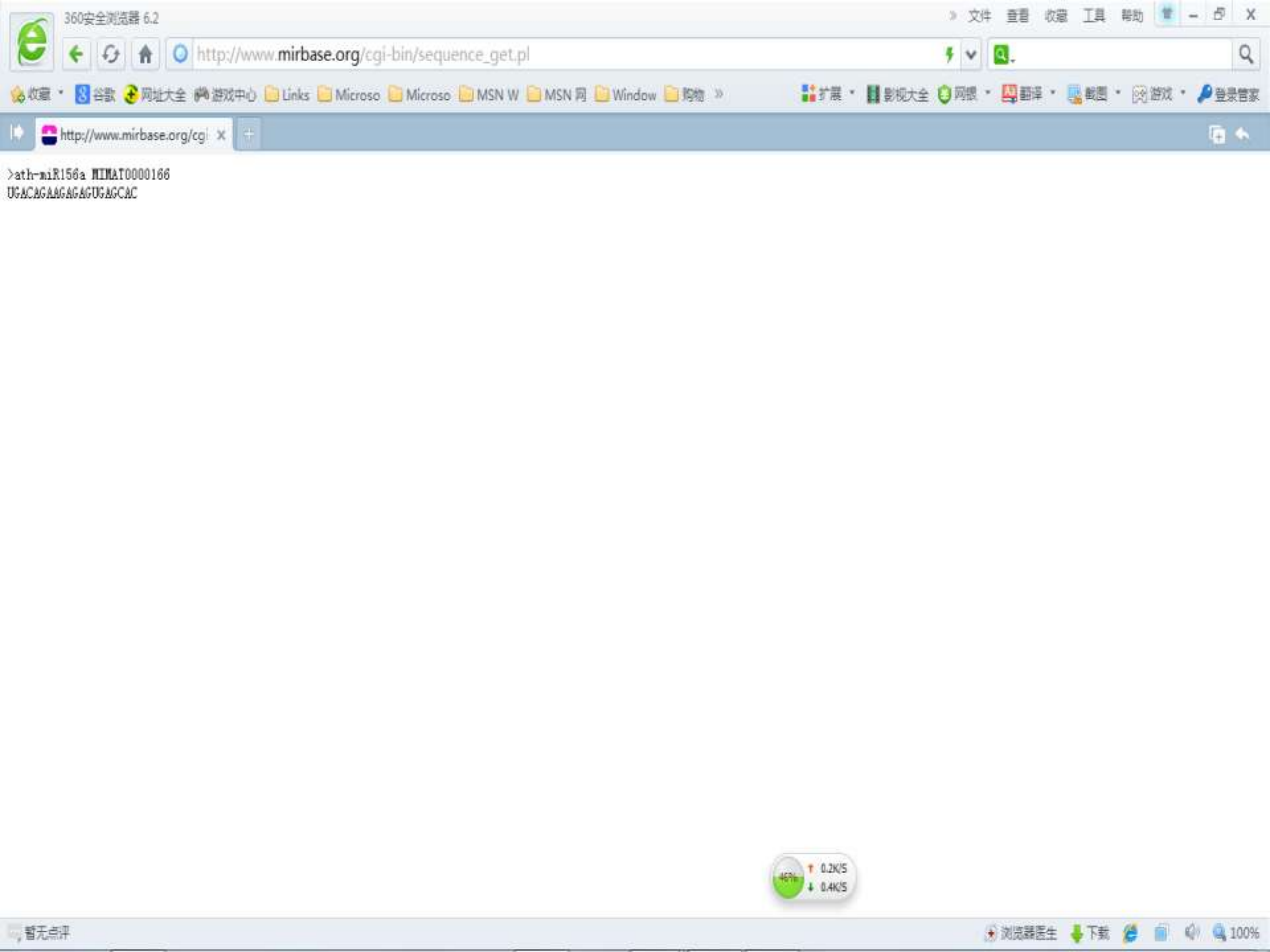
Highlights, Web watch - *Nature Reviews Genetics* 5:244 (2004)

miRBase: the microRNA database

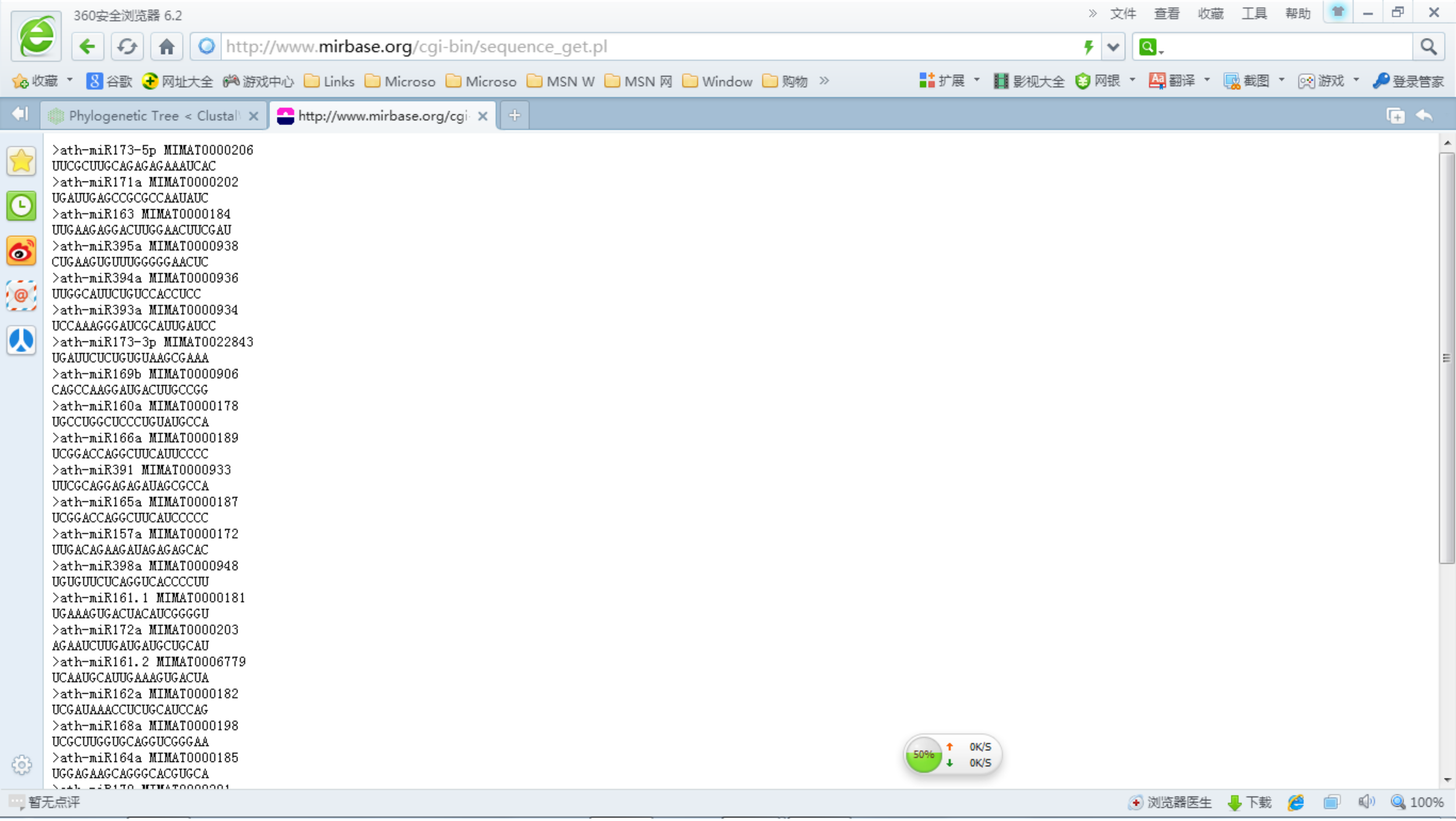
miRBase provides the following services:

- The miRBase database is a searchable database of published miRNA sequences and annotation. Each entry in the miRBase Sequence database represents a predicted hairpin portion of a miRNA transcript (termed mir in the database), with information on the location and sequence of the mature miRNA sequence (termed miR). Both hairpin and mature sequences are available for searching and browsing, and entries can also be retrieved by name, keyword, references and annotation. All sequence and annotation data are also available for download.
- The miRBase Registry provides miRNA gene hunters with unique names for novel miRNA genes prior to publication of results. Visit the help pages for more information about the naming service.

To receive email notification of data updates and feature changes please subscribe to the miRBase announcements mailing list. Any queries about the website or naming service should be directed at mirbase@manchester.ac.uk.



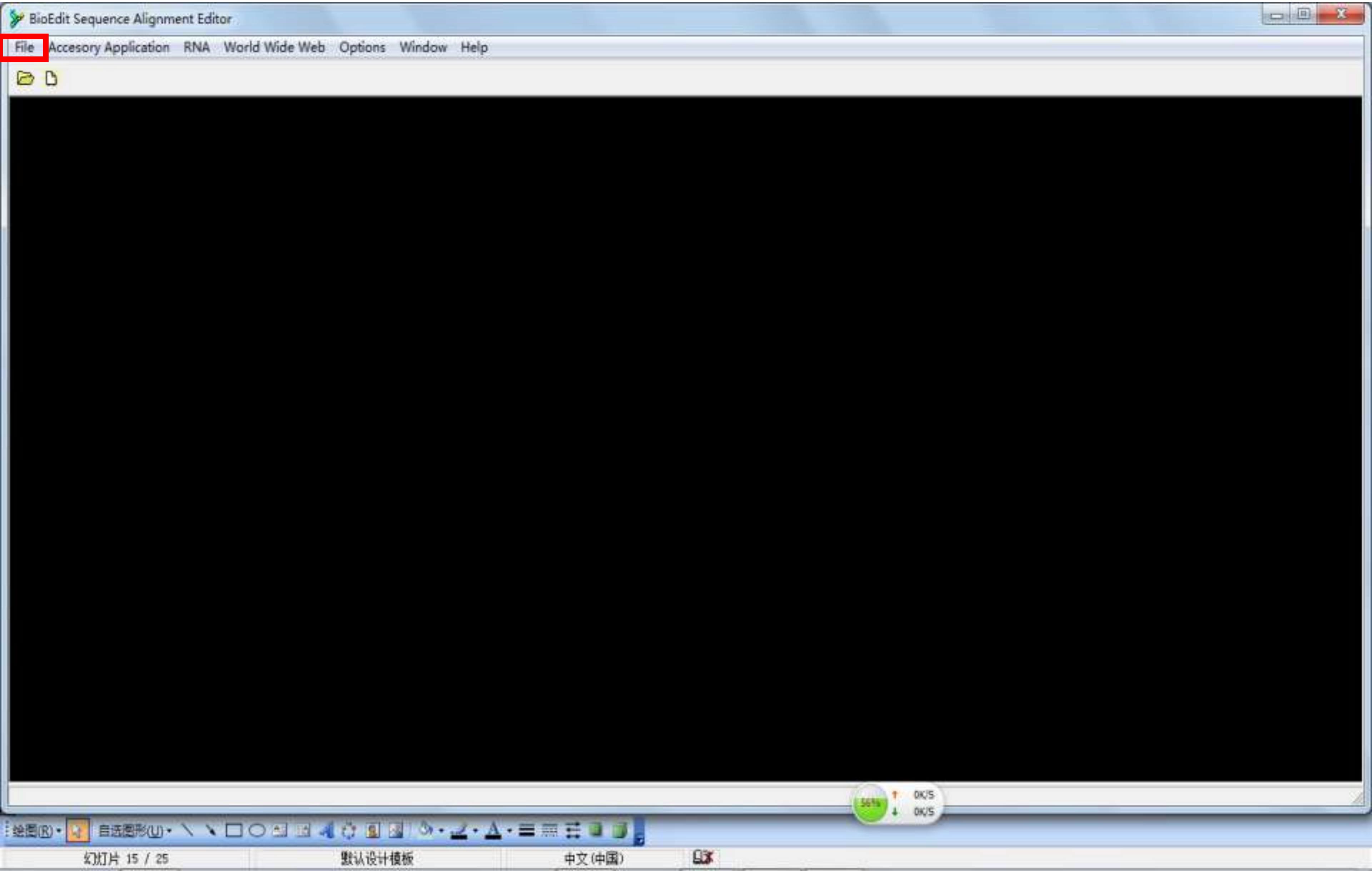
```
>ath-miR156a MIMAT0000166
UGACAGAAGAGAGUGAGCAC
```

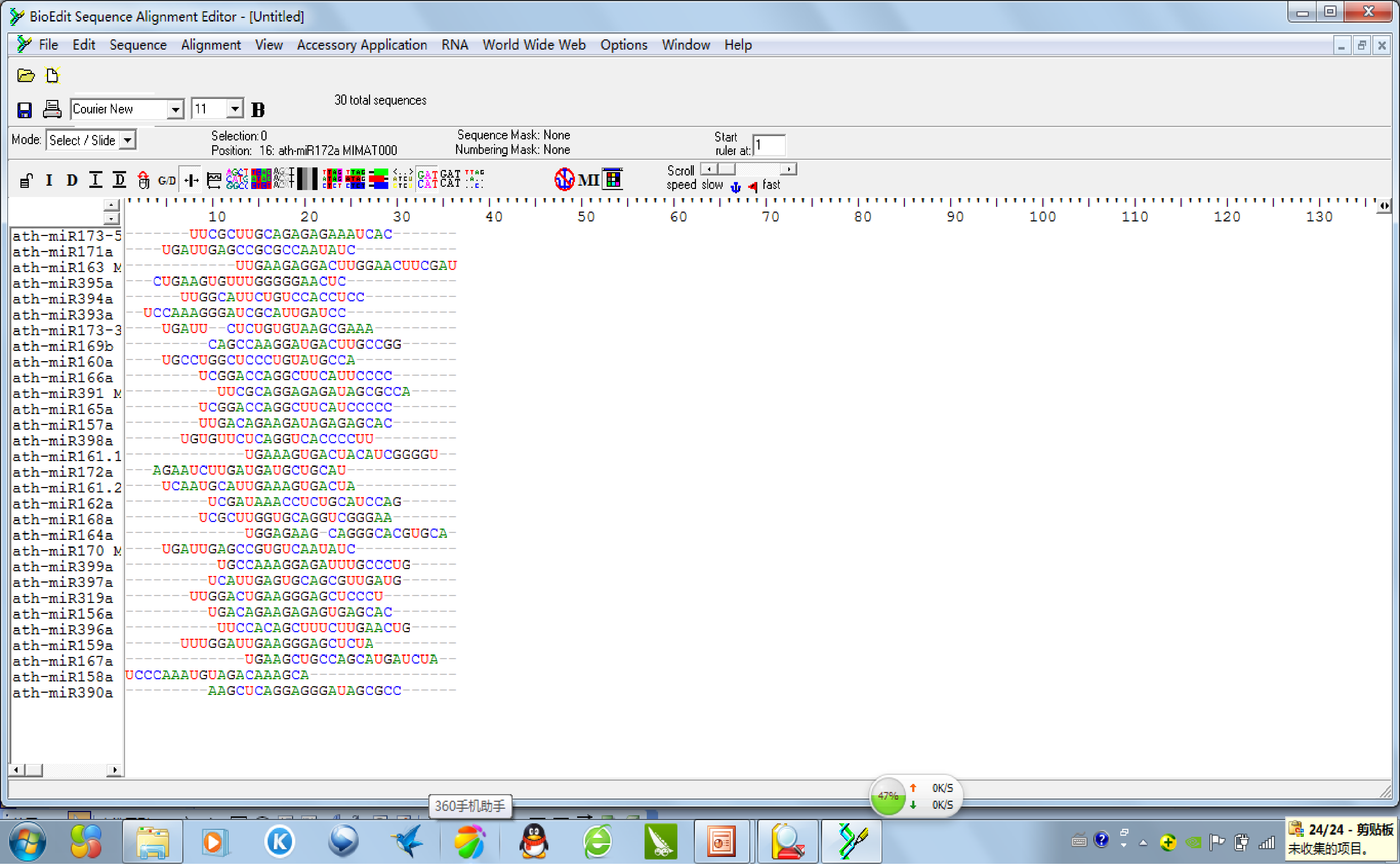


```
>ath-miR173-5p MIMAT0000206
UUCGCUUGCAGAGAGAAAUCAC
>ath-miR171a MIMAT0000202
UGAUUGAGCCGCGCCAAUAUC
>ath-miR163 MIMAT0000184
UUGAAGAGGACUUGGAACUUCGAU
>ath-miR395a MIMAT0000938
CUGAAGUGUUUGGGGGAACUC
>ath-miR394a MIMAT0000936
UUGGCAUUCUGUCCACCUCC
>ath-miR393a MIMAT0000934
UCCAAAGGGAUCGCAUUGAUCC
>ath-miR173-3p MIMAT0022843
UGAUUCUCUGUGUAAGCGAAA
>ath-miR169b MIMAT0000906
CAGCCAAGGAUGACUUGCCGG
>ath-miR160a MIMAT0000178
UGCCUGGCUCCCUGUAUGCCA
>ath-miR166a MIMAT0000189
UCGGACCAGGCUUCAUUCCCC
>ath-miR391 MIMAT0000933
UUCGCAGGAGAGAUAGCGCCA
>ath-miR165a MIMAT0000187
UCGGACCAGGCUUCAUCCCCC
>ath-miR157a MIMAT0000172
UUGACAGAAGAUAGAGAGCAC
>ath-miR398a MIMAT0000948
UGUGUUCUCAGGUCACCCCUU
>ath-miR161.1 MIMAT0000181
UGAAAGUGACUACAUCGGGGU
>ath-miR172a MIMAT0000203
AGAAUCUUGAUGAUGCUGCAU
>ath-miR161.2 MIMAT0006779
UCAAUGCAUUGAAAGUGACUA
>ath-miR162a MIMAT0000182
UCGAUAAACCUCUGCAUCCAG
>ath-miR168a MIMAT0000198
UCGCUUGGUGCAGGUCGGGAA
>ath-miR164a MIMAT0000185
UGGAGAAGCAGGGCACGUGCA
```

注：可新建一个文本文档，将序列粘贴进去，保存。

比对：bioedit中的clustal w





30 total sequences

Mode: Select / Slide

Selection: 0
Position: 16: ath-miR172a MIMAT000

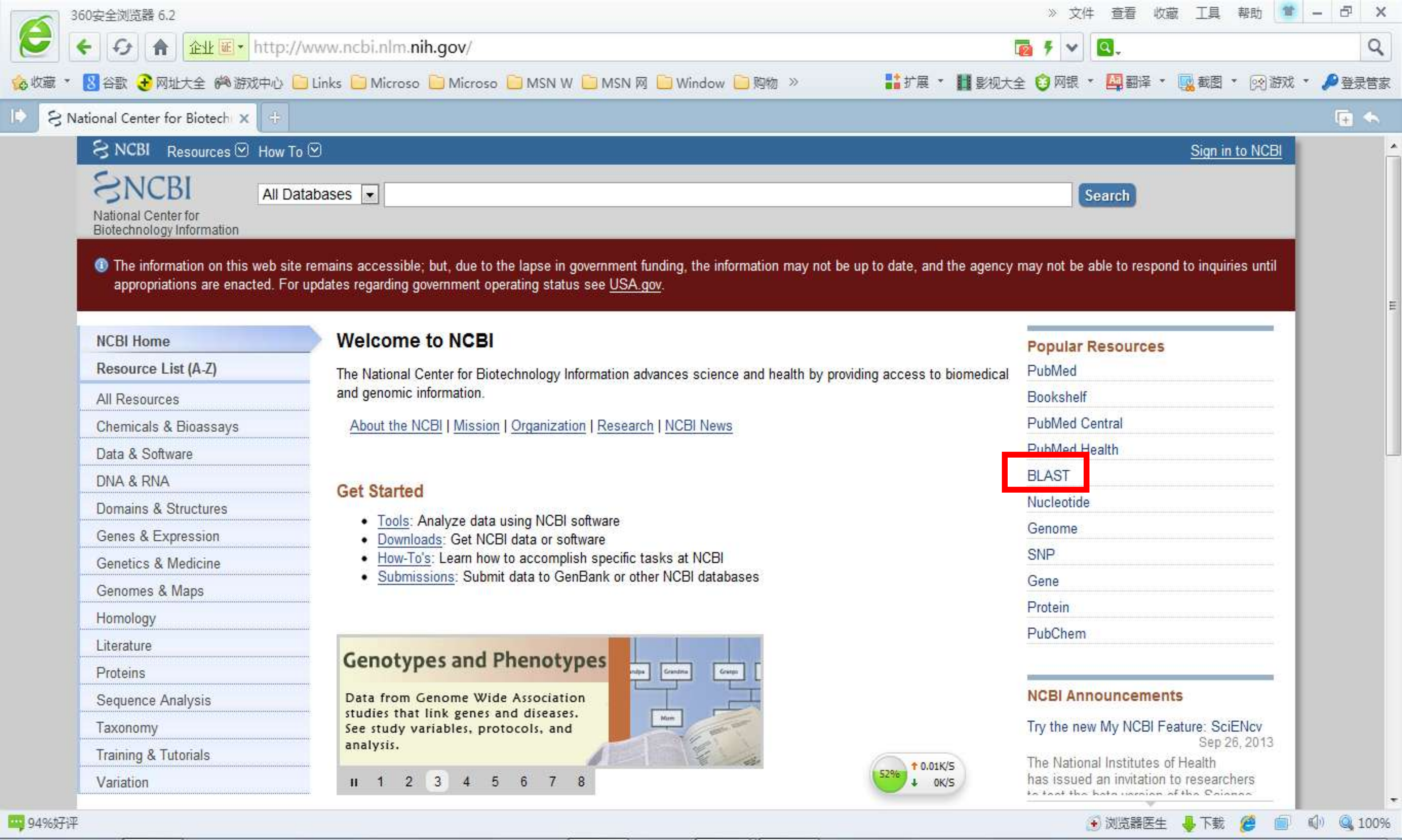
Sequence Mask: None
Numbering Mask: None

Start ruler at: 1

```
                      10        20        30        40        50        60        70        80        90       100       110       120       130
ath-miR173-5 -------UUCGCUUGCAGAGAGAAAUCAC-------
ath-miR171a  ----UGAUUGAGCCGCGCCAAUAUC-----------
ath-miR163 M ------------UUGAAGAGGACUUGGAACUUCGAU
ath-miR395a  ---CUGAAGUGUUUGGGGGAACUC------------
ath-miR394a  ------UUGGCAUUCUGUCCACCUCC----------
ath-miR393a  --UCCAAAGGAUCGCAUUGAUCC-------------
ath-miR173-3 ----UGAUU--CUCUGUGUAAGCGAAA---------
ath-miR169b  ---------CAGCCAAGGAUGACUUGCCGG------
ath-miR160a  ----UGCCUGGCUCCCUGUAUGCCA-----------
ath-miR166a  --------UCGGACCAGGCUUCAUUCCCC-------
ath-miR391 M ----------UUCGCAGGAGAGAUAGCGCCA-----
ath-miR165a  --------UCGGACCAGGCUUCAUCCCCC-------
ath-miR157a  --------UUGACAGAAGAUAGAGAGCAC-------
ath-miR398a  ------UGUGUUCUCAGGUCACCCCUU---------
ath-miR161.1 -------------UGAAAGUGACUACAUCGGGGU--
ath-miR172a  ---AGAAUCUUGAUGAUGCUGCAU------------
ath-miR161.2 ----UCAAUGCAUUGAAAGUGACUA-----------
ath-miR162a  ---------UCGAUAAACCUCUGCAUCCAG------
ath-miR168a  --------UCGCUUGGUGCAGGUCGGGAA-------
ath-miR164a  -------------UGGAGAAG-CAGGGCACGUGCA-
ath-miR170 M ----UGAUUGAGCCGUGUCAAUAUC-----------
ath-miR399a  ----------UGCCAAAGGAGAUUUGCCCUG-----
ath-miR397a  ---------UCAUUGAGUGCAGCGUUGAUG------
ath-miR319a  -------UUGGACUGAAGGGAGCUCCCU--------
ath-miR156a  ---------UGACAGAAGAGAGUGAGCAC-------
ath-miR396a  ----------UUCCACAGCUUUCUUGAACUG-----
ath-miR159a  ------UUUGGAUUGAAGGGAGCUCUA---------
ath-miR167a  -------------UGAAGCUGCCAGCAUGAUCUA--
ath-miR158a  UCCCAAAUGUAGACAAAGCA----------------
ath-miR390a  ---------AAGCUCAGGAGGGAUAGCGCC------
```

注：去除冗余miRNA，剩下的序列用blastn 软件在EST、GSS和核酸数据库中进行筛选

比对：blastn（http://www.ncbi.nlm.nih.gov）



The information on this web site remains accessible; but, due to the lapse in government funding, the information may not be up to date, and the agency may not be able to respond to inquiries until appropriations are enacted. For updates regarding government operating status see USA.gov.

- NCBI Home
- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | Research | NCBI News

Get Started

- Tools: Analyze data using NCBI software
- Downloads: Get NCBI data or software
- How-To's: Learn how to accomplish specific tasks at NCBI
- Submissions: Submit data to GenBank or other NCBI databases

Genotypes and Phenotypes

Data from Genome Wide Association studies that link genes and diseases. See study variables, protocols, and analysis.

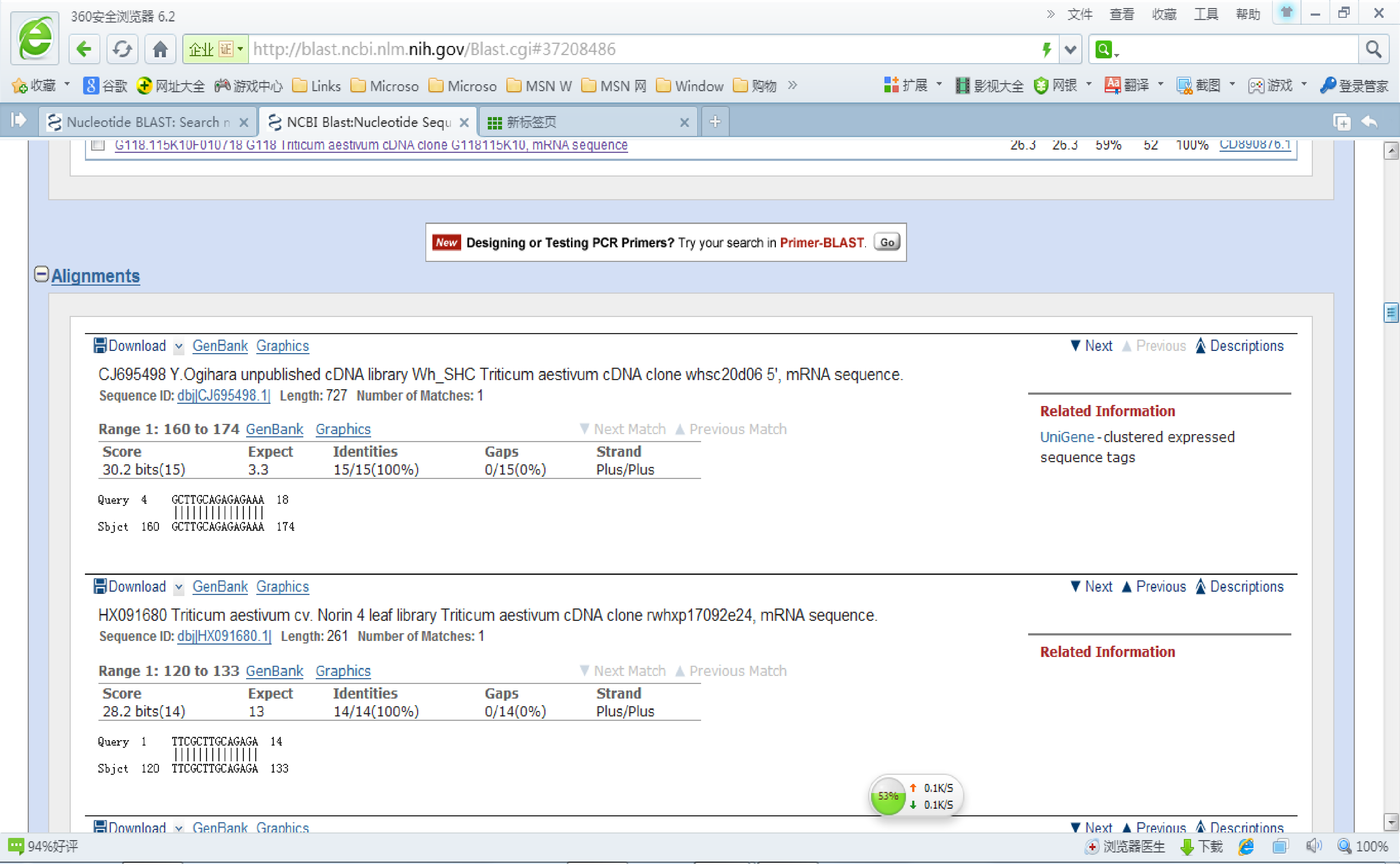
Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- PubMed Health
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI Announcements

Try the new My NCBI Feature: SciENcv
Sep 26, 2013

The National Institutes of Health has issued an invitation to researchers



Designing or Testing PCR Primers? Try your search in Primer-BLAST.

Alignments

Download GenBank Graphics | Next Previous Descriptions

CJ695498 Y.Ogihara unpublished cDNA library Wh_SHC Triticum aestivum cDNA clone whsc20d06 5', mRNA sequence.

Sequence ID: dbj|CJ695498.1| **Length:** 727 **Number of Matches:** 1

Related Information

UniGene - clustered expressed sequence tags

Range 1: 160 to 174 GenBank Graphics | Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	3.3	15/15(100%)	0/15(0%)	Plus/Plus

```
Query  4    GCTTGCAGAGAGAAA  18
            |||||||||||||||
Sbjct  160  GCTTGCAGAGAGAAA  174
```

Download GenBank Graphics | Next Previous Descriptions

HX091680 Triticum aestivum cv. Norin 4 leaf library Triticum aestivum cDNA clone rwhxp17092e24, mRNA sequence.

Sequence ID: dbj|HX091680.1| **Length:** 261 **Number of Matches:** 1

Related Information

Range 1: 120 to 133 GenBank Graphics | Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
28.2 bits(14)	13	14/14(100%)	0/14(0%)	Plus/Plus

```
Query  1    TTCGCTTGCAGAGA  14
            ||||||||||||||
Sbjct  120  TTCGCTTGCAGAGA  133
```

Download GenBank Graphics | Next Previous Descriptions

注：选择匹配度达到90%以上的序列保留

Basic BLAST

Choose a BLAST program to run.

nucleotide blast	Search a **nucleotide** database using a **nucleotide** query *Algorithms:* blastn, megablast, discontiguous megablast
protein blast	Search **protein** database using a **protein** query *Algorithms:* blastp, psi-blast, phi-blast, delta-blast
blastx	Search **protein** database using a **translated nucleotide** query
tblastn	Search **translated nucleotide** database using a **protein** query
tblastx	Search **translated nucleotide** database using a **translated nucleotide** query

Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

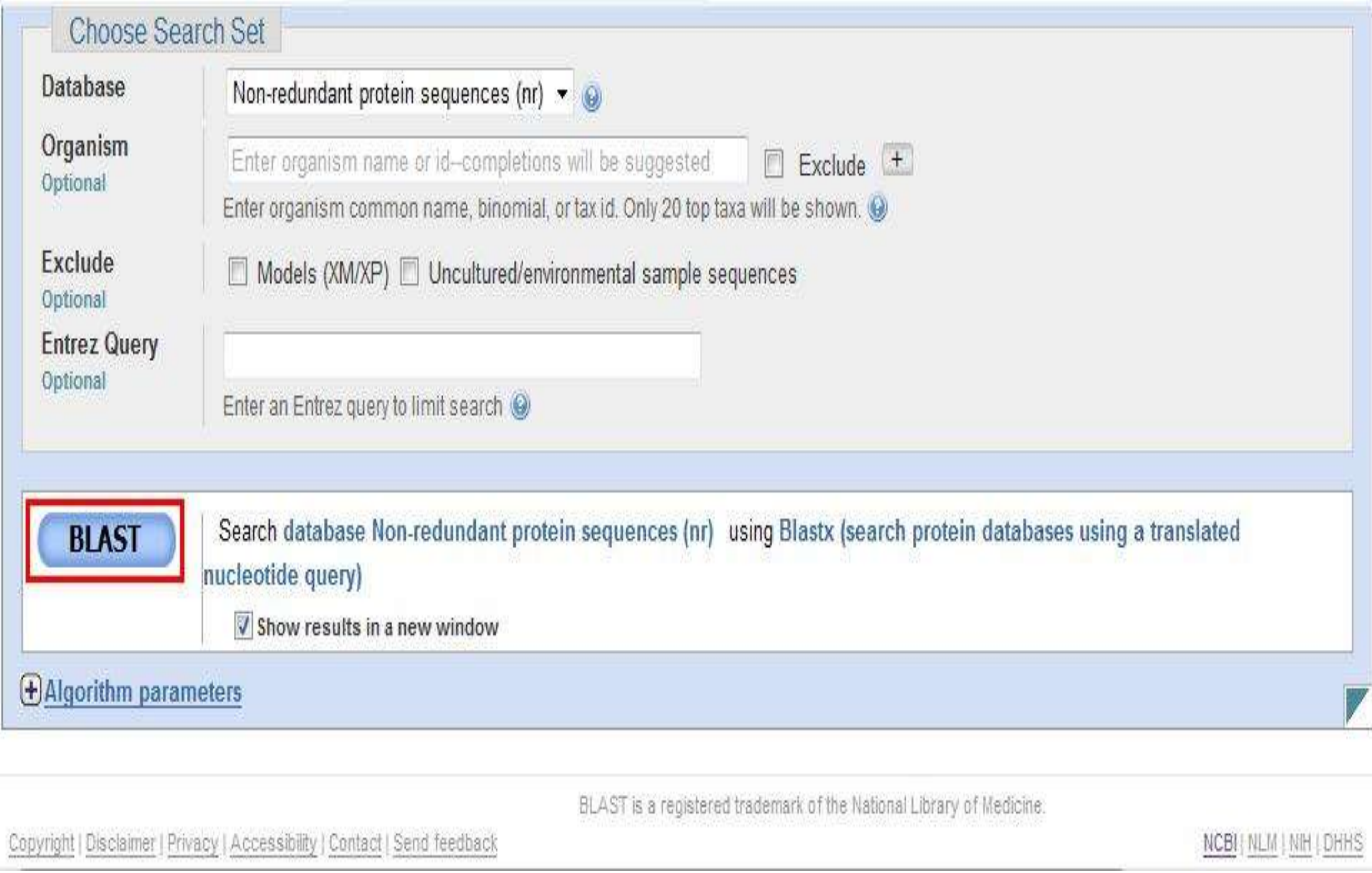
A new version of the stand-alone BLAST applications has been released.
Mon, 10 Sep 2012
14:00:00 EST

More BLAST news...

Tip of the Day

How to save custom search pages.

So you have made a few BLAST searches and after adjusting the database, organism limits and maybe a few



Choose Search Set

Database: Non-redundant protein sequences (nr)

Organism Optional: Enter organism name or id--completions will be suggested ☐ Exclude +

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

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

Entrez Query Optional:

Enter an Entrez query to limit search

BLAST

Search database Non-redundant protein sequences (nr) using Blastx (search protein databases using a translated nucleotide query)

☑ Show results in a new window

+ Algorithm parameters

BLAST is a registered trademark of the National Library of Medicine.

Copyright | Disclaimer | Privacy | Accessibility | Contact | Send feedback

NCBI | NLM | NIH | DHHS

注：blastx的目的是剔除编码序列