



**PBEA**  
PLANT BREEDING E-LEARNING IN AFRICA

Published on *Plant Breeding E-Learning in Africa* (<https://pbea.agron.iastate.edu>)

[Home](#) > [Course Materials](#) > [Molecular Plant Breeding](#) > Introduction to Bioinformatics

---

## Introduction to Bioinformatics



By Ursula Frei, Walter Suza, Thomas Lübberstedt, Madan Bhattacharyya (ISU)



Except otherwise noted, this work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

## Introduction

A biological sequence database is a collection of molecular data organized in a manner that allows easy access, management, and update of the data. Biological sequence databases serve an important role of providing access to sequence information to the research community. The databases contain molecular information of multiple organisms and are constantly being updated and re-designed to allow more robust data query and analysis. Examples of biological databases include European Molecular Biology Laboratory (EMBL), GenBank, the National Center for Biotechnology Information (NCBI), and the DNA Databank of Japan (DDBJ). Every sequence submitted to the database has a unique number assigned to it, called the Accession number. Even if the same gene has been submitted several times by different investigators each will have a different accession number.

## Database Types

Databases can be classified into primary (archival), secondary (curated), and composite databases.

- A **primary database** (e.g. EMBL/DDBJ/GenBank for nucleic acids) contains information of the sequence or structure alone, for example, DNA, RNA, or protein sequences.
- A **secondary database** (e.g. eMOTIF at Stanford University, PROSITE of Swiss Institute of Bioinformatics) contains information derived from the primary databases and represent sequences that are consensus of a population, for example, conserved features and motifs of a sequence.
- A **composite database** contains a variety of different primary databases and provides multiple options for database search (e.g. NCBI, MaizeGDB). New tools are continuously developed to make both submission and access to sequence databases more efficient.

The lesson includes practical examples of using database tools. It is recommended that you use “try this” questions to become familiar with sequence databases.

## Objectives

- To introduce some of the most commonly used databases in molecular plant breeding
- To help the student learn the tools for accessing and manipulating biological databases
- To help develop proficiency in the use of biological databases

## Access and Use of Sequence Databases

Once a new sequence has been determined a common step in its analysis is to compare the sequence with related genes that have already been sequenced, often from other organisms. A few things to keep in mind about database searches and sequence databases in general:

1. Do not assume that if a sequence is in the database it must be correct. Databases are full of errors!
2. Similarity with a known protein or gene does not necessarily mean the query is the same gene as the one it has similarity with.
3. Two nucleotide sequences may have low similarity yet code for proteins that are functionally related.
4. Protein sequences may also have low similarity yet still be functionally or structurally related.

# About NCBI

NCBI was created in 1988 as a division of the US National Library of Medicine at the National Institute of Health. The role of NCBI is to create automated systems for storing and analyzing sequence information.

1. To access various resources available through NCBI select **Resources**.
2. We recommend that you set up an account with NCBI to allow you the option of saving your results. Click the **Sign in** link to do so
3. Video tutorials are available under the **Training & Tutorials** link to enhance learning.

The screenshot shows the NCBI homepage with three annotations:

- Annotation 1:** A circle around the "Resources" dropdown menu in the top navigation bar.
- Annotation 2:** A circle around the "Sign in to NCBI" link in the top right corner.
- Annotation 3:** A circle around the "Training & Tutorials" link in the left sidebar menu.

The main content area features a "Welcome to NCBI" section with the following text: "The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information." Below this are links for "About the NCBI", "Mission", "Organization", and "NCBI News & Blog".

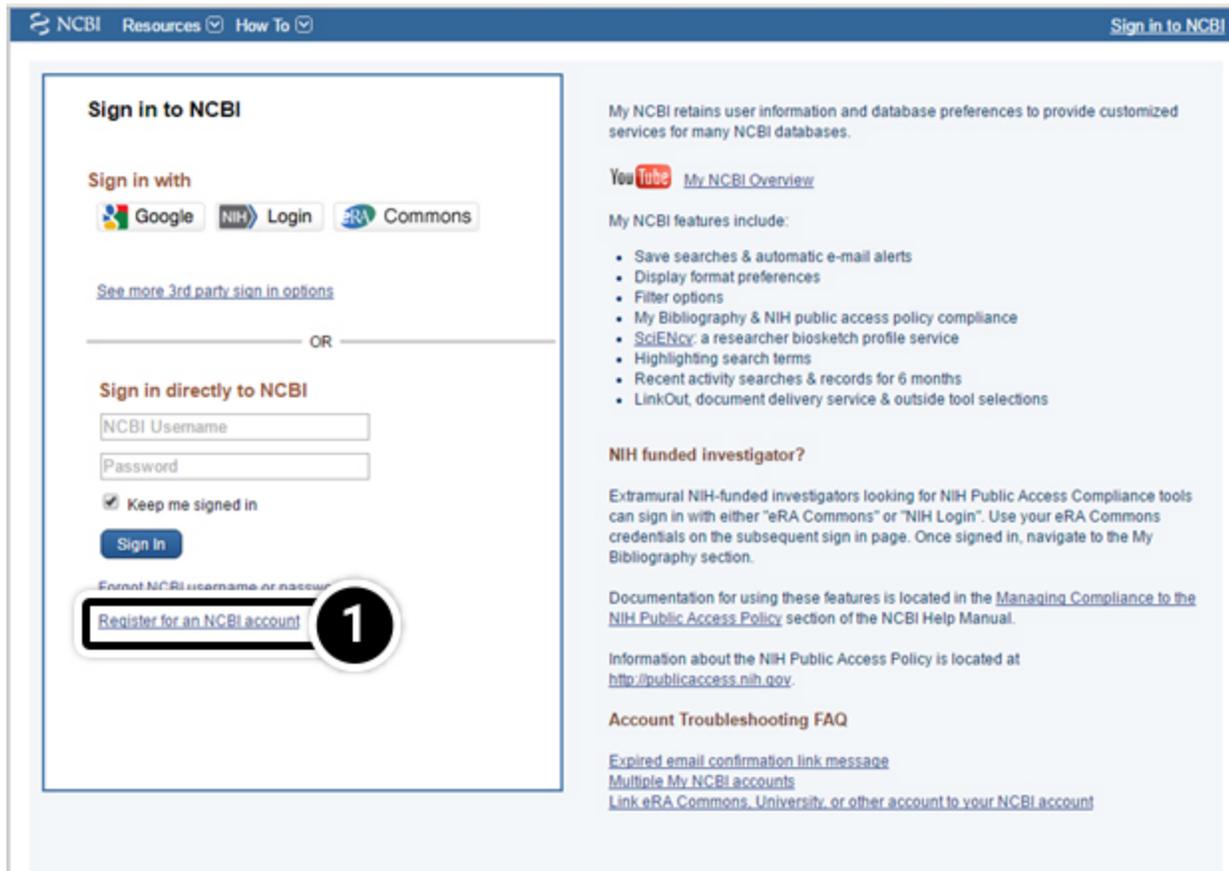
There are six main action cards:

- Submit:** Deposit data or manuscripts into NCBI databases. Icon: Upward arrow.
- Download:** Transfer NCBI data to your computer. Icon: Downward arrow.
- Learn:** Find help documents, attend a class or watch a tutorial. Icon: Open book.
- Develop:** Use NCBI APIs and code libraries to build applications. Icon: Code blocks.
- Analyze:** Identify an NCBI tool for your data analysis task. Icon: Microscope.
- Research:** Explore NCBI research and collaborative projects. Icon: Microscope.

On the right side, there are sections for "Popular Resources" (PubMed, Bookshelf, PubMed Central, PubMed Health, BLAST, Nucleotide, Genome, SNP, Gene, Protein, PubChem) and "NCBI News & Blog" (NYGC NCBI-style bioinformatics hackathon August 6-8, 2018, dated 09 Jul 2018).

# Sign Up for NCBI

1. Click **Register** to set up a new NCBI account.



The screenshot shows the NCBI 'Sign in to NCBI' page. At the top, there are navigation links for 'NCBI', 'Resources', and 'How To', along with a 'Sign in to NCBI' link. The main content area is divided into two columns. The left column contains the sign-in options: 'Sign in with' (Google, NIH Login, eRA Commons), a link to 'See more 3rd party sign in options', an 'OR' separator, 'Sign in directly to NCBI' (with fields for 'NCBI Username' and 'Password', a 'Keep me signed in' checkbox, and a 'Sign in' button), and a link to 'Register for an NCBI account' which is highlighted with a red box and a white circle containing the number '1'. The right column contains information about user information retention, a YouTube link for 'My NCBI Overview', a list of 'My NCBI features include:' (such as saving searches, display preferences, and recent activity), a section for 'NIH funded investigator?' with instructions on using eRA Commons or NIH Login, and links to documentation and a troubleshooting FAQ.

# NCBI Training

**NCBI** was created in 1988 as a division of the US National Library of Medicine at the National Institute of Health. The role of NCBI is to create automated system for storing and analyzing sequence information.

**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

**Training & Tutorials**

All Databases Downloads Tools How To

**Databases**

[NCBI C++ Toolkit Manual](#)  
A comprehensive manual on the NCBI C++ toolkit, including its design and development framework, a C++ library reference, software examples and demos, FAQs and release notes. The manual is searchable online and can be downloaded as a series of PDF documents.

[NCBI Education Page](#)  
Provides links to tutorials and training materials, including PowerPoint slides and print handouts.

[NCBI Glossary](#)  
Part of the NCBI Handbook, this glossary contains descriptions of NCBI tools and acronyms, bioinformatics terms and data representation formats.

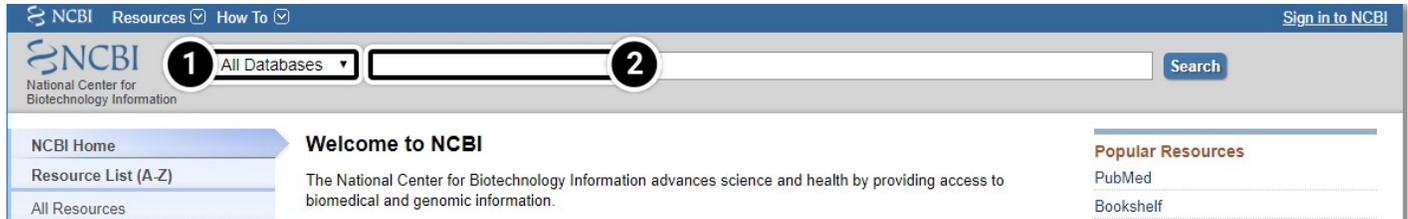
[NCBI Handbook](#)  
An extensive collection of articles about NCBI databases and software. Designed for a novice user, each article presents a general overview of the resource and its design, along with tips for searching and using available analysis tools. All articles can be searched online and downloaded in PDF format, the handbook can be accessed through the NCBI Bookshelf.

[NCBI Help Manual](#)

# Information Retrieval from NCBI

One of the most widely used interfaces for the retrieval of sequence information from biological databases is the [NCBI Entrez system](#). Entrez relies on preexisting, logical relationships between the individual sequences (data points) available in various public databases.

1. Searching all databases is often a good starting point to get an overview of the state of your research field.
2. Searches are based on keywords.



The screenshot shows the top navigation bar of the NCBI website. It includes the NCBI logo, a search bar with a dropdown menu set to "All Databases", and a "Search" button. Below the search bar is a "Welcome to NCBI" message and a "Popular Resources" section with links to PubMed and Bookshelf. The search bar is annotated with a circled "1" pointing to the dropdown menu and a circled "2" pointing to the search input field.

NCBI Resources How To Sign in to NCBI

NCBI National Center for Biotechnology Information

All Databases Search

NCBI Home  
Resource List (A-Z)  
All Resources

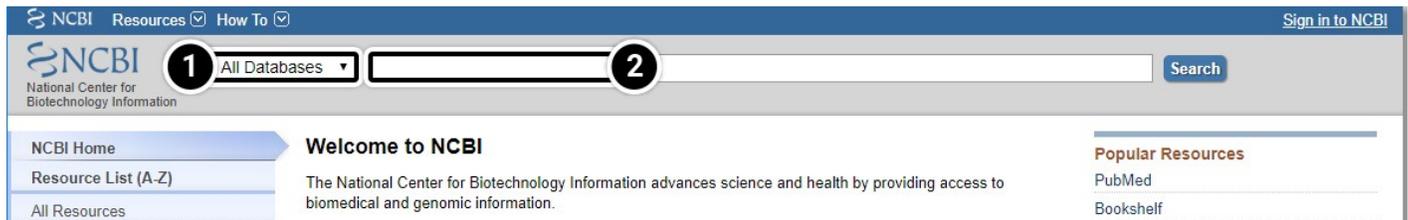
**Welcome to NCBI**  
The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

**Popular Resources**  
PubMed  
Bookshelf

# Searching NCBI by Keywords

Searches can be restricted to a single database or expanded to include all other databases. The simplest way to query is through the use of individual search terms, coupled by Boolean operators such as AND, OR, or NOT. A Boolean operator is a variable that can have only a true or false value.

1. Select individual databases, or search them all.
2. **AND:** To 'AND' two search terms together instructs Entrez to find all documents that contain BOTH terms  
**OR:** To 'OR' two search terms together instructs Entrez to find all documents that contain EITHER term.  
**NOT:** To 'NOT' two search terms together instructs Entrez to find all documents that contain search term 1 BUT NOT search term 2.



The screenshot shows the NCBI search interface. At the top, there is a navigation bar with 'NCBI Resources' and 'How To' menus, and a 'Sign in to NCBI' link. Below this is the search bar area, which includes the NCBI logo, a dropdown menu labeled 'All Databases' (annotated with a circled '1'), a search input field (annotated with a circled '2'), and a 'Search' button. On the left side, there is a sidebar with links for 'NCBI Home', 'Resource List (A-Z)', and 'All Resources'. The main content area features a 'Welcome to NCBI' message and a description of the center's mission. On the right side, there is a 'Popular Resources' section with links to 'PubMed' and 'Bookshelf'.

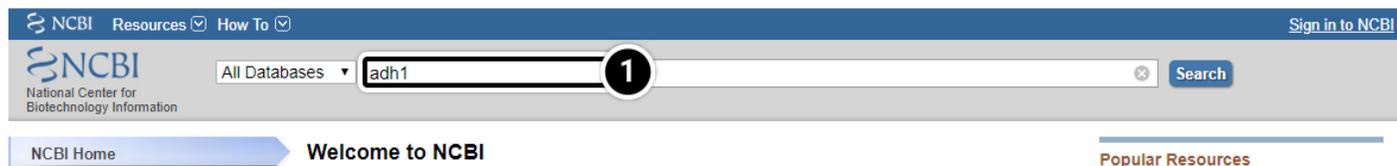
## Try This: Access and Use of Sequence Databases

This activity consists of the following pages:

*Try This: Access and Use of Sequence DBs (Compare the sequences) - 1*

Compare the sequences for the adh1 gene in maize and sorghum. Navigate to the [NCBI site](#).

1. Enter adh1 in the "search across databases" window. How many adh1 candidates did your search find?



The screenshot shows the top navigation bar of the NCBI website. On the left, there is the NCBI logo and the text "National Center for Biotechnology Information". To the right of the logo is a dropdown menu labeled "All Databases" and a search input field containing the text "adh1". A red circle with the number "1" is drawn around the search input field. To the right of the search input field is a "Search" button. In the top right corner of the navigation bar, there is a link that says "Sign in to NCBI". Below the navigation bar, there is a "Welcome to NCBI" message and a "Popular Resources" link.

Try This: Access and Use of Sequence DBs (Results of a search) - 2

Results of a search for "adh1" across all databases:

The screenshot shows the NCBI search interface. At the top, the NIH logo and 'U.S. National Library of Medicine National Center for Biotechnology Information' are visible, along with a 'Log in' button. The search bar contains 'adh1' and a 'Search' button. Below the search bar, it states 'Results found in 32 databases for adh1'. The results are categorized into 'Literature' and 'Genes'. Under 'Literature', there are three items: 'Bookshelf' (23), 'MeSH' (1), and 'NLM Catalog' (1). Under 'Genes', there are three items: 'EST' (107), 'Gene' (393), and 'GEO' (122). The 'Gene' result is highlighted with a red box and a circled '1' next to it.

Literature		Genes	
Bookshelf	23	EST	107
MeSH	1	Gene	393
NLM Catalog	1	GEO	122

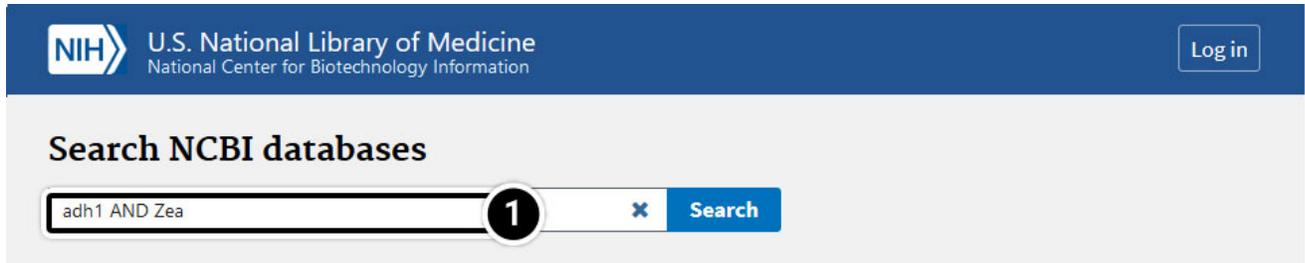
Descriptions for Literature: Books and reports; Ontology used for PubMed indexing; Books, journals and more in the NLM.

Descriptions for Genes: Expressed sequence tag sequences; Collected information about gene loci; Functional genomics studies.

## Try This: Access and Use of Sequence DBs (Compare the results) - 3

Compare the sequences for the *adh1* gene in maize and sorghum:

1. Enter **adh1 AND Zea** in the search window.



NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

Search NCBI databases

adh1 AND Zea **1** Search

Compare the results in the **Gene** category.

2. Boolean operators can be used to restrict a search and allow users to obtain specific information about their organism of interest.

Results found in 15 databases for **adh1 AND Zea**

### Literature

<b>Bookshelf</b>	<b>1</b>	Books and reports
<b>MeSH</b>	0	Ontology used for PubMed indexing
<b>NLM Catalog</b>	0	Books, journals and more in the NLM Collections
<b>PubMed</b>	<b>126</b>	Scientific and medical abstracts/citations
<b>PubMed Central</b>	<b>556</b>	Full-text journal articles
<b>PubMed Health</b>	0	Clinical effectiveness, disease and drug reports

### Genes

<b>EST</b>	<b>5</b>	Expressed sequence tag sequences
<b>Gene</b>	<b>2</b>	Collected information about gene loci
<b>GEO DataSets</b>	0	Functional genomics studies
<b>GEO Profiles</b>	<b>6</b>	Gene expression and molecular abundance profiles
<b>HomoloGene</b>	0	Homologous gene sets for selected organisms
<b>PopSet</b>	<b>17</b>	Sequence sets from phylogenetic and population studies
<b>UniGene</b>	<b>1</b>	Clusters of expressed transcripts

## Try This: Access and Use of Sequence DBs (Operators) - 4

Now try these operators.

1. Enter **adh1 AND Zea[orgn] OR Sorghum[orgn]** in the search window.
2. Results

 U.S. National Library of Medicine  
National Center for Biotechnology Information Log in

---

### Search NCBI databases

1 ✕ Search

Results found in 32 databases for **adh1 AND Zea[orgn] OR Sorghum[orgn]**

### Literature

<b>Bookshelf</b>	<b>265</b>	Books and reports
<b>MeSH</b>	<b>9</b>	Ontology used for PubMed indexing
<b>NLM Catalog</b>	<b>73</b>	Books, journals and more in the NLM Collections
<b>PubMed</b>	<b>5,388</b>	Scientific and medical abstracts/citations
<b>PubMed Central</b>	351	Full-text journal articles
<b>PubMed Health</b>	7	Clinical effectiveness, disease and drug reports

### Genes

<b>EST</b>	<b>232,686</b>	Expressed sequence tag sequences
<b>Gene</b>	<b>33,084</b>	Collected information about gene loci
<b>GEO DataSets</b>	<b>693</b>	Functional genomics studies
<b>GEO Profiles</b>	6	Gene expression and molecular abundance profiles
<b>HomoloGene</b>	0	Homologous gene sets for selected organisms
<b>PopSet</b>	<b>602</b>	Sequence sets from phylogenetic and population studies
<b>UniGene</b>	<b>13,734</b>	Clusters of expressed transcripts

### Genetics

<b>ClinVar</b>	0	Human variations of clinical significance
<b>dbGaP</b>	0	Genotype/phenotype interaction studies
<b>dbVar</b>	0	Genome structural variation studies
<b>---</b>	~	Genetic variation studies

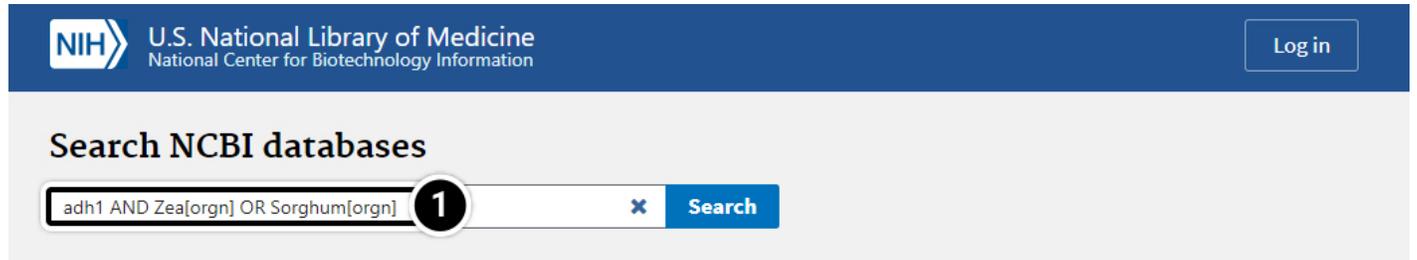
### Proteins

<b>Conserved Domains</b>	0	Conserved protein domains
<b>Identical Protein Groups</b>	<b>54,945</b>	Protein sequences grouped by identity

## Try This: Access and Use of Sequence DBs (Operators) - 5

Now try these operators.

1. Enter **adh1 AND Zea[orgn] OR Sorghum[orgn]** in the search window.



The screenshot shows the top navigation bar of the NCBI website. On the left is the NIH logo and the text "U.S. National Library of Medicine National Center for Biotechnology Information". On the right is a "Log in" button. Below the navigation bar is a search bar with the text "Search NCBI databases". The search input field contains the query "adh1 AND Zea[orgn] OR Sorghum[orgn]" and has a "1" in a circle next to it. To the right of the input field is a blue "Search" button.

What stands out when you compare results when using the search terms "*adh1 AND Zea[orgn] OR Sorghum[orgn]*" and "*adh1 AND (Zea[orgn] OR Sorghum[orgn])*"? Can you identify any differences among the results obtained from the following sets of search terms?

- "*adh1 AND Zea[orgn]*" and "*adh1 AND Zea[orgn] OR Sorghum[orgn]*"
- "*adh1 AND Zea[orgn]*" and "*adh1 AND (Zea[orgn] OR Sorghum[orgn])*"

1. Enter **adh1 AND (Zea[orgn] OR Sorghum[orgn])** in the search window.
2. Compare the results in the **Gene** category.



## Search NCBI databases

adh1 AND (Zea[orgn] OR Sorghum[orgn]) **1** ✕ **Search**

Results found in 15 databases for **adh1 AND (Zea[orgn] OR Sorghum[orgn])**

### Literature

<b>Bookshelf</b>	<b>1</b>	Books and reports
<b>MeSH</b>	0	Ontology used for PubMed indexing
<b>NLM Catalog</b>	0	Books, journals and more in the NLM Collections
<b>PubMed</b>	<b>131</b>	Scientific and medical abstracts/citations
<b>PubMed Central</b>	351	Full-text journal articles
<b>PubMed Health</b>	0	Clinical effectiveness, disease and drug reports

### Genetics

<b>ClinVar</b>	0	Human variations of clinical significance
<b>dbGaP</b>	0	Genotype/phenotype interaction studies

### Genes

<b>EST</b>	<b>5</b>	Expressed sequence tag sequences
<b>Gene</b>	<b>2</b>	Collected information about gene loci
<b>GEO DataSets</b>	0	Functional genomics studies
<b>GEO Profiles</b>	6	Gene expression and molecular abundance profiles
<b>HomoloGene</b>	0	Homologous gene sets for selected organisms
<b>PopSet</b>	<b>19</b>	Sequence sets from phylogenetic and population studies
<b>UniGene</b>	<b>1</b>	Clusters of expressed transcripts

### Proteins

<b>Conserved Domains</b>	0	Conserved protein domains
--------------------------	---	---------------------------

## Try This: Access and Use of Sequence DBs (Gene-centered Info) - 6

1. Click on "Gene" to get gene-centered information on the output in the last screen results (also shown here).
2. Click on the first "adh1".
3. Review the output window.

### Search NCBI databases

Search

Results found in 15 databases for **adh1 AND (Zea[orgn] OR Sorghum[orgn])**

### Literature

- Bookshelf** 1 Books and reports
- MeSH** 0 Ontology used for PubMed indexing
- NLM Catalog** 0 Books, journals and more in the NLM Collections

### Genes

- EST** 5 Expressed sequence tag sequences
- Gene** 2 Collected information about gene loci
- GEO** 0 Functional genomics studies
- DataSets**

NCBI
Sign in to NCBI

Gene
Search

Gene sources  
Genomic

Categories  
Alternatively spliced  
Annotated genes  
Protein-coding

Sequence content  
RefSeq

Status  
Current

Chromosome locations  
more...

[Clear all](#)

[Show additional filters](#)

Tabular ▾ Sort by Relevance ▾

Send to: ▾

[Hide sidebar >>](#)

**Search results**

Items: 2

Showing Current items.

Name/Gene ID	Description	Location	Aliases
<input type="checkbox"/> <a href="#">adh1</a> ID: 542363	alcohol dehydrogenase 1 [ <i>Zea mays</i> ]	Chromosome 1, NC_024459.2 (278821306..278824958, complement)	ZEAMMB73_Zm00001d033931, Adh1-1F, Adh1-1S, GRMZM2G442658, adh1A
<input type="checkbox"/> <a href="#">LOC110436814</a> ID: 110436814	alcohol dehydrogenase 1 [ <i>Sorghum bicolor</i> (sorghum)]	Chromosome 1, NC_012870.2 (7505851..7509144)	SORBI_3001G097600, Adh1

Filters: [Manage Filters](#)

**Results by taxon**

Top Organisms [\[Tree\]](#)

*Sorghum bicolor* (1)

*Zea mays* (1)

**Find related data**

Database: Select

[Find items](#)

**Search details**

[\[adh1 AND \(Zea\[orgn\] OR Sorghum\[orgn\]\)\]](#)

### adh1 alcohol dehydrogenase 1 [*Zea mays*]

Gene ID: 542363, updated on 26-Feb-2018

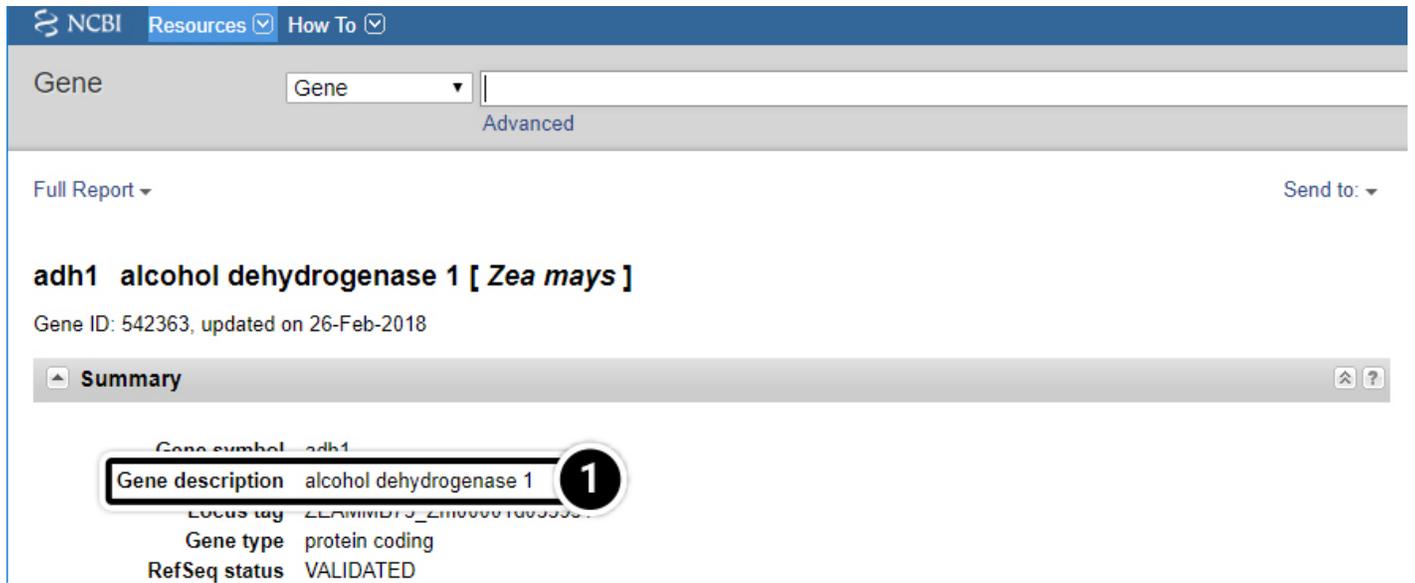
**Summary**

Gene symbol	adh1
Gene description	alcohol dehydrogenase 1
Locus tag	ZEAMMB73_Zm00001d033931
Gene type	protein coding
RefSeq status	VALIDATED
Organism	<a href="#">Zea mays</a>
Lineage	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACMAD clade; Panicoideae; Andropogonodae; Andropogoneae; Tripsacinae; Zea
Also known as	adh1A; Adh1-1F; Adh1-1S; GRMZM2G442658

## Try This: Access and Use of Sequence DBs (adh1) - 7

What is the function of adh1?

1. Answer found below.



The screenshot shows the NCBI Gene database interface. At the top, there is a navigation bar with 'NCBI', 'Resources', and 'How To'. Below this is a search bar with 'Gene' selected in the dropdown and an empty input field. A 'Full Report' dropdown and a 'Send to:' dropdown are visible. The main heading is 'adh1 alcohol dehydrogenase 1 [ Zea mays ]' with 'Gene ID: 542363, updated on 26-Feb-2018' below it. A 'Summary' section is expanded, showing a table of gene properties. The 'Gene description' row is highlighted with a black box and a circled '1'.

Gene symbol	adh1
<b>Gene description</b>	alcohol dehydrogenase 1 <b>1</b>
Locus tag	ZM000001.210000.1
Gene type	protein coding
RefSeq status	VALIDATED

## Try This: Access and Use of Sequence DBs (Nucleotide results) - 8

Let's examine the Nucleotide results.

1. Click this pull-down menu for more information about this gene and select **Nucleotide**.
2. Click **Search**. Your search will result in 100s of hits.

The screenshot displays the NCBI Gene database interface. At the top, there are navigation links for 'NCBI', 'Resources', and 'How To', along with a 'Sign in to NCBI' button. The main search bar contains the text 'adh1 AND (Zea[orgn] OR Sorghum[orgn])' and a 'Search' button with a circled '2' next to it. A dropdown menu is open, showing various database options, with 'Nucleotide' highlighted and a circled '1' next to it. The left sidebar shows the 'Summary' section for the gene 'adh1 alcohol dehydrogenase 1'. The right sidebar contains a 'Table of contents' with links to 'Summary', 'Genomic context', 'Genomic regions, transcripts, and products', 'Bibliography', 'Variation', 'Pathways from BioSystems', 'General gene information', 'General protein information', and 'NCBI Reference Sequences (RefSeq)'. The bottom of the page shows 'Also known as' with the identifiers 'adh1A; Adh1-1F; Adh1-1S; GRMZM2G442658'.

Try This: Access and Use of Sequence DBs (Nucleotide results) - 9

Let's examine the Nucleotide results.

1. After selecting the **nucleotide** option as in previous screen, click on the **adh1 mRNA** as indicated. You may have to scroll down to find it.

[Zea mays Adh1 {intron 3} \[maize, Genomic, 363 nt 2 segments\]](#)

6. 505 bp linear DNA

Accession: AH004391.1 GI: 1683071

[PubMed](#) [Taxonomy](#)

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

[Zea mays alcohol dehydrogenase 1 \(adh1\), mRNA](#)

7. 1,779 bp linear mRNA

Accession: NM\_001111939.2 GI: 1200005210

[Protein](#) [PubMed](#) [Taxonomy](#)

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

[Zea mays isolate adh1-3F1124 transposon Mu3, complete sequence; and alcohol dehydrogenase-like \(adh1\) gene, partial sequence](#)

8. 2,190 bp linear DNA

Accession: U19613.1 GI: 639494

[PubMed](#) [Taxonomy](#)

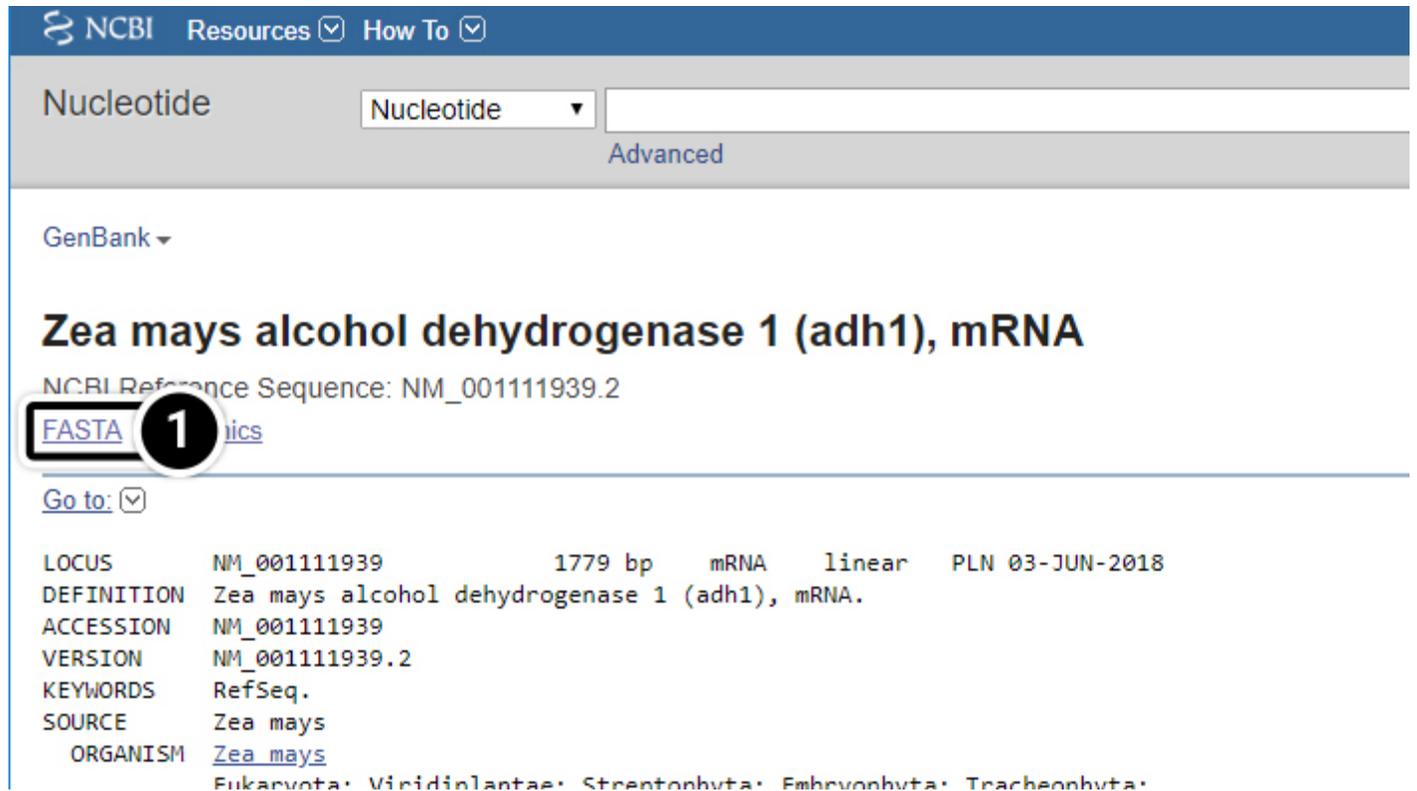
[GenBank](#) [FASTA](#) [Graphics](#)

## Try This: Access and Use of Sequence DBs

Let's examine the Nucleotide results.

1. Click the **FASTA** link
2. Reference sequences are accessed through GenBank to provide non-redundant curated data derived from experimental knowledge of known genes.

Additional information about RefSeq can be found [here](#).



NCBI Resources How To

Nucleotide Nucleotide Advanced

GenBank

### Zea mays alcohol dehydrogenase 1 (adh1), mRNA

NCBI Reference Sequence: NM\_001111939.2

**FASTA** 1 [GenBank](#) [Blast](#) [Align](#) [Map](#) [Download](#) [Help](#)

Go to: [v]

LOCUS	NM_001111939	1779 bp	mRNA	linear	PLN 03-JUN-2018
DEFINITION	Zea mays alcohol dehydrogenase 1 (adh1), mRNA.				
ACCESSION	NM_001111939				
VERSION	NM_001111939.2				
KEYWORDS	RefSeq.				
SOURCE	Zea mays				
ORGANISM	<a href="#">Zea mays</a>				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				

Nucleotide

Nucleotide

Advanced

Learn more about upcoming changes to the Nucleotide, EST, and GSS databases.

FASTA

## Zea mays alcohol dehydrogenase 1 (adh1), mRNA

NCBI Reference Sequence: NM\_001111939.2

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

```
>NM_001111939.2 2 Zea mays alcohol dehydrogenase 1 (adh1), mRNA
GGCTATGTTCCACTCCAGGTGGAGGCTGCAGCCCCGGTTTCGCAAGCCG
CGCCGTGGTTTGCTTGCCACAGGCGGCCAAACCGCACCCCTCCTTCCCGTCGTTTCCCATCTCTTCTCTCC
TTTAGAGCTACCACTATATAAATCAGGGCTCATTTTCTCGCTCCTCACAGGCTCATCTCGCTTTGGATCG
ATTGTTTTCGTAACTGGTGAGGGACTGAGGGTCTCGGAGTGGATTGATTTGGGATTCTGTTCGAAGATT
```

Try This: Access and Use of Sequence DBs

## Zea mays alcohol dehydrogenase 1 (adh1), mRNA

NCBI Reference Sequence: NM\_001111939.2

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

```
>NM_001111939.2 Zea mays alcohol dehydrogenase 1 (adh1), mRNA
AAACCACGGTCCACGGACCACGGCTATGTTCCACTCCAGGTGGAGGCTGCAGCCCCGGTTTCGCAAGCCG
CGCCGTGGTTTGCTTGCCACAGGCGGCCAAACCACACCCTCCTTCCCGTCGTTTCCCATCTCTTCCCTC
TTTAGAGCTACCACTATATAAATCAGGGCTCATTTTCTCGCTCCTCACAGGCTCATCTCGCTTTGGATCG
ATTGGTTTCGTAAGTGGTGAGGGACTGAGGGTCTCGGAGTGGATTGATTGGGATTCTGTTGGAAGATTT
GCGGAGGGGGCAATGGCGACCAGCGGGGAAGGTGATCAAGTGCAAAGCTGCGGTGGCATGGGAGGCCGGC
AAGCCACTGTCGATCGAGGAGGTGGAGGTAGCGCCTCCGAGGCCATGGAGGTGCGCGTCAAGATCCTCT
TCACCTCGCTCTGCCACACCGACGCTACTTCTGGGAGGCCAAGGGGAGACTCCCGTGTCCCTCGGAT
CTTTGGCCACGAGGCTGGAGGTATCATAGAGAGTGTGGAGAGGGTGTGACTGACGTAGCTCCGGGCGAC
CATGTCCTTCTGTGTTCACTGGGAGTGAAGGAGTGTGCCACTGCAAGTCGGCAGAGAGCAACATGT
GTGATCTGCTCAGGATCAACACCGACCGCGGTGTGATGATTGCCGATGGCAAGTCGCGGTTTTCAATCAA
TGGGAAGCCTATCTACCACTTTGTTGGGACTTCCACCTTCAGCGAGTACACCGTCATGCATGTGGGTTGT
GTTGCAAAGATCAACCCTCAGGCTCCCTTGATAAAGTTGCGTCCTTAGCTGTGGTATTTCTACCGGTC
TTGGTGCATCAATTAATGTTGCAAAACCTCCGAAGGGTTCGACAGTGGCTGTTTTCGGTTTAGGAGCCGT
TGGTCTTGCCGCTGCAGAAGGTGCAAGGATTGCTGGAGCGTCAAGGATCATTGGTGTGACCTGAACCCC
AGCAGATTGCAAGAAGCTAGGAAGTTCGGTTCGACTGAATTTGTGAACCCAAAAGACCACAACAAGCCAG
TGCAGGAGGTACTTGTGAGATGACCAACGGAGGGGTCGACCGCAGTGTGGAATGCACTGGCAACATTAA
TGCTATGATCCAAGCTTTGCAATGTGTTGATGATGGTGGGGTGTGCTGTGCTGGTGGGTGTGCCACAT
AAGGACGCTGAGTTCAAGACCCACCCGATGAACTTCTGAACGAAAGGACCCTGAAGGGGACCTTCTTTG
GCAACTATAAGCCACGCACTGATCTGCCAAATGTGGTGGAGCTGTACATGAAAAAGGAGCTGGAGGTGGA
GAAGTTCATCACGCACAGCGTCCCGTTGCGCGAGATCAACAAGGCGTTCGACCTGATGGCCAAGGGGGAG
GGCATCCGCTGCATCATCCGCATGGAGAAGTAGATTTGCTGTCTAGTTTGTGATCTGGCCTGGGCTTGG
GGTTAATAAAAGAGGCAATGCTAGCCTGCCCTTTCGATGAGGAGGTACATACACGCTGGCGATGGACCGC
GCTTGTGTGTCGCGTTTCAAGTTGGCTTTTGCCAAGCAGTAGGGTAGCTTCCCGTGTGCGTAATTATATGG
TATGAACCATCACCTTTGGCTCTACATGGTATGAACGTAAGATACAAATTCAACTACCTCTAGCTCGC
TTGTGTGGTATCTGTATCAGTATTCATGTGTTGTTGCTTATGTGTTGTTGCTTGTATTTGCTGGTG
CTTGTATCGCGGGATGCAATGAGTTGTTG
```

After clicking the FASTA link, what kind of information do you get?

Does the entire mRNA sequence for adh1 you obtained code for a protein product? If not, how would you identify the coding sequence?

Now that you have found the adh1 coding sequence, what is its estimated length?

Enter your answer here.

Show Answer



# BLAST Interface

[Here](#) the user can restrict searches to a specific species and to the assembled reference sequences for that species. For a plant researcher, it may not be necessary to restrict a search except for those working with rice and Arabidopsis. For all other plant species reference sequences are not fully developed.

The screenshot displays the BLAST web interface. At the top, there is a navigation bar with logos for NIH (U.S. National Library of Medicine) and NCBI (National Center for Biotechnology Information), along with a 'Sign in to NCBI' link. Below this is a secondary navigation bar with 'BLAST' and links for 'Home', 'Recent Results', 'Saved Strategies', and 'Help'.

The main content area features a 'Basic Local Alignment Search Tool' section with a brief description and a 'Learn more' link. To the right is a 'NEWS' sidebar with a post titled 'Introducing the BLAST widget - Integrating your BLAST results into NCBI's Genome Data Viewer!' dated June 19, 2018.

The 'Web BLAST' section contains three main options: 'Nucleotide BLAST' (nucleotide to nucleotide), 'blastx' (translated nucleotide to protein), and 'tblastn' (protein to translated nucleotide). A 'Protein BLAST' option (protein to protein) is also visible on the right.

At the bottom, the 'BLAST Genomes' section includes a search input field with the placeholder text 'Enter organism common name, scientific name, or tax id' and a 'Search' button. Below the input field are links for 'Human', 'Mouse', 'Rat', and 'Microbes'.

## BLAST Features

1. Basic BLAST features include blastn, blastp, blastx, tblastn, and tblastx.
2. Specialized features include "**Global Align**" for sequence alignment.

## Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

[Learn more](#)

NEWS

A new version (1.4.0) of the BLAST RNA-seq mapping tool, Magic-BLAST, is now available

Tue, 21 Aug 2018 16:00:00 EST

[More BLAST news...](#)

## Web BLAST

1

The diagram shows two main categories: Nucleotide BLAST and Protein BLAST. Nucleotide BLAST includes 'blastx' (translated nucleotide to protein) and 'tblastn' (protein to translated nucleotide). Protein BLAST includes 'blastp' (protein to protein).

## BLAST Genomes

[Search](#)[Human](#)[Mouse](#)[Rat](#)[Microbes](#)

## Standalone and API BLAST



### Download BLAST

Get BLAST databases and executables



### Use BLAST API

Call BLAST from your application



### Use BLAST in the cloud

Start an instance at a cloud provider

## Specialized searches

### SmartBLAST

Find proteins highly similar to your query

### Primer-BLAST

Design primers specific to your PCR template

### Global Align

Compare two sequences across their entire span (Needleman-Wunsch)

### CD-search

Find conserved domains in your sequence

# Try This: Using NCBI BLAST

## Try This: Using NCBI BLAST

1. Within the **Basic BLAST** window, click on **Nucleotide BLAST**. A new window appears asking you to setup your search options.
2. This is where your query sequence will go.
3. This selects the **Database** you want to search.
4. Other parameters you may want to set different from the standard settings.

### Basic Local Alignment Search Tool

**BLAST** finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

**NEWS**  
Introducing the BLAST widget - integrating your BLAST results into NCBI's Genome Data Viewer!

Analyze your BLAST results in a genome browser and compare those results against other genome assembly annotations. Introducing the Genome Data Viewer (GDV) and the BLAST widget.

Tue, 19 Jun 2018 14:00:00 EST

[More BLAST news...](#)

### Web BLAST



## BLAST® &gt;&gt; blastn suite

Standard Nucleotide

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide data

## Enter Query Sequence

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

2

clear

Query subrange

From

To

Or, upload file

Choose File

No file chosen

Job Title

Enter a descriptive title for your BLAST search

 Align two or more sequences

## Choose Search Set

Database

 Human genomic + transcript  Genomic + transcript  Others (nr etc.):

Nucleotide collection (nr/nt)

3

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

4

Limit to  
Optional Sequences from type materialEntrez Query  
Optional

Enter an Entrez query to limit search

You  Create custom database

## Try This: Using NCBI BLAST

You have various options of entering your query sequence: copy and paste or uploading a saved sequence from your computer.

Your query sequence has to be annotated in FASTA format. FASTA is a text-based format consisting of a definition line followed by the sequence data in single letter code. The definition line starts with the character ">", followed by a sequence name, and ends with a return or newline. Everything that follows until the next ">" will be considered as the sequence data. It is possible to save multiple sequences in one FASTA file.

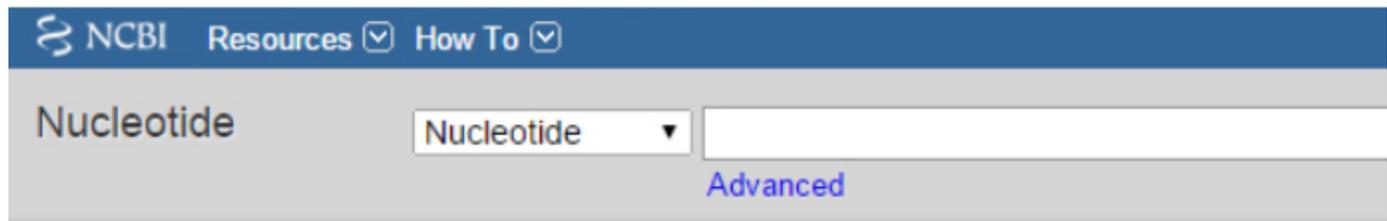
1. In the screenshot below,

Definition line starts with ">" character,

**gi** stands for GenBank identification, followed by GenBank ID number,

**ref** stands for reference sequence, followed by the accession number.

Both GenBank ID and reference sequence numbers can be used to enter a query sequence into BLAST.



[Display Settings:](#)  FASTA

## Zea mays alcohol dehydrogenase1 (adh1), mRNA

NCBI Reference Sequence: NM\_001111939.1

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

```
>gi|162460396|ref|NM_001111939.1|1 Zea mays alcohol dehydrogenase1 (adh1), mRNA
ATCTCGCTTTGGATCGATTGGTTTCGTAACCTGGTCCGGACTGAGGGTCTCGGAGTGGATCGATTTGGGA
TTCTGTTTGAAGATTGCGGAGGGGGCAATGGCGACCGCGGGGAAGGTGATCAAGTGCAAAGCTGCGGT
GGCATGGGAGGCCGGCAAGCCACTGTCGATCGAGGAGGTGGAGGTAGCGCCTCCGCAGGCCATGGAGGTG
CGCGTCAAGATCCTCTTACCTCGCTCTGCCACACCGACGTCTACTTCTGGGAGGCCAAGGGGCAGACTC
CCGTGTTCCCTCGGATCTTTGGCCACGAGGCTGGAGGTATCATAGAGAGTGTGGAGAGGGTGTGACTGA
CGTAGCTCCGGGCGACCATGTCCTTCTGTGTTCACTGGGGAGTGCAAGGAGTGTGCCCACTGCAAGTCCG
GCAGAGAGCAACATGTGTGATCTGCTCAGGATCAACACCGACCGCGGTGTGATGATTGCCGATGGCAAGT
CGCGGTTTTCAATCAATGGGAAGCCTATCTACCACTTTGTTGGGACTTCCACCTTCAAGGAGTACACCGT
CATGCATGTGGGTTGTGTTGCAAAGATCAACCCTCAGGCTCCCCTTGATAAAGTTTGCCTCTTAGCTGT
GGTTATTCTACCGGTCTTGGTGCATCAATTAATGTTGCAAAACCTCCGAAGGGTTCGACAGTGGCTGTTT
TCGGTTTAGGAGCCGTTGGTCTTGCCGCTGCAGAAGGTGCAAGGATTGCTGGAGCGTCAAGGATCATTGG
TGTCGACCTGAACCCAGCAGATTGCAAGAAGCTAGGAAGTTCGGTTGCACTGAATTTGTGAACCCAAAA
GACCAACACCAAGCCAGTGCAGGAGGTACTTGTGAGATGACCAACGGAGGGTTCGACCGCAGTGTGGAAT
```

## Try This: Using NCBI BLAST

You may enter your query (adh1) as a sequence in FASTA format.

1. To do that, copy the entire adh1 sequence
2. Paste it in the **Enter accession numbers(s), gi(s), or FASTA sequence(s)** window.
3. Note that the **Job Title** filled automatically.

### Zea mays alcohol dehydrogenase 1 (adh1), mRNA

NCBI Reference Sequence: NM\_001111939.2

GenBank Graphics

```
>NM_001111939.2 Zea mays alcohol dehydrogenase 1 (adh1), mRNA
AAACCACGGTCCACGGACCACGGCTATGTTCCACTCCAGGTGGAGGCTGCAGCCCCGGTTTCGCAAGCCG
CGCCGTGGTTTGCTTGCCACAGGCGGCCAAACCGCACCCCTCTCCCGTCGTTTCCCATCTTTCCTCC
TTTAGAGTACCACATATAAATCAGGGCTCATTTTCTCGCTCCTCACAGGCTCATCTCGCTTGGATCG
ATTGGTTTCGTAAC
GCGGAGGGGGGCAAT
AAGCCACTGTCGATC
TCACCTCGCTCGCC
CTTTGGCCACGAGGC
CATGTCCTTCTGTG
GTGATCTGCTCAGGA
TGGGAAGCCTATCTA
GTTGCAAAGATCAAC
TTGGTGCATCAATTA
TGGTCTTGCCGCTGC
AGCAGATTCGAAGAA
TGCAGGAGGTACTTG
TGCTATGATCCAAGC
AAGGACGCTGAGTTC
GCAACTATAAGCCAC
GAAGTTCATCACGCA
GGCATCCGCTGCATC
GGTTAATAAAAAGGG
GCTTGTGTGTCGCGT
TATGAACCATCACCT
TTGTGTGGTATCTGT
CTTGTATCGCGGGAT
```

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information

**BLAST** >> blastn suite

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

Enter Query Sequence

BLASTN programs search nucleotide databases using a nucleotide query sequence.

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

Clear Query subrange

From

To

Or, upload file Choose File No file chosen

Job Title NM\_001111939.2 Zea mays alcohol dehydrogenase...

Enter a descriptive title for your BLAST search

Align two or more sequences

## Try This: Using NCBI BLAST

1. Alternatively, you can query your sequence using the **Run BLAST** command. Click "Run BLAST" to query the sequence from the FASTA display screen.

FASTA ▾

### Zea mays alcohol dehydrogenase 1 (adh1), mRNA

NCBI Reference Sequence: NM\_001111939.2

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

```
>NM_001111939.2 Zea mays alcohol dehydrogenase 1 (adh1), mRNA
AAACCACGGTCCACGGACCACGGCTATGTTCCACTCCAGGTGGAGGCTGCAGCCCCGGTTTCGCAAGCCG
CGCCGTGGTTTGCTTGCCACAGGCGGCCAAACCGCACCCCTCTCCCGTCGTTTCCCATCTTCTCTCC
TTTAGAGCTACCACTATATAAATCAGGGCTCATTTTCTCGCTCCTCACAGGCTCATCTCGCTTTGGATCG
ATTGGTTTCGTAAGTGGTGGGACTGAGGGTCTCGGAGTGGATTGATTTGGGATTCTGTTTGAAGATT
GCGGAGGGGGCAATGGCGACCGGGGAAGGTGATCAAGTGCAAAGCTGCGGTGGCATGGGAGGCCGGC
AAGCCACTGTCGATCGAGGAGGTGGAGGTAGCGCTCCGCGAGCCATGGAGGTGCGCGTCAAGATCCTCT
TCACCTCGCTCTGCCACCCGACGTCTACTTCTGGGAGGCCAAGGGGCAGACTCCCCTGTTCCCTCGGAT
CTTTGGCCACGAGGCTGGAGGTATCATAGAGAGTGTGGAGAGGGTGTGACTGACGTAGCTCCGGGCGAC
CATGTCCTTCTGTGTTCACTGGGGAGTGCAAGGAGTGTGCCCACTGCAAGTCCGCGAGAGCAACATGT
```

Send to: ▾

Change region shown ▾

Customize view ▾

Analyze this sequence ▴

1

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Show in Genome Data Viewer

## Try This: Using NCBI BLAST

Clicking on the **Run BLAST** command will lead you to this window.

1. **Accession number** of adh1 will automatically fill in
2. **Job Title** should automatically fill in, if it does not you can click in the **Job Title** field and it should appear automatically.
3. Optimize your search to **megablast** to identify highly similar sequences.
4. Finally, select the **BLAST** button.

The screenshot shows the NCBI BLAST Standard Nucleotide BLAST interface. The page header includes NIH, U.S. National Library of Medicine, NCBI National Center for Biotechnology Information, and a sign-in link. The main navigation bar contains 'BLAST' and 'blastn suite', along with links for Home, Recent Results, Saved Strategies, and Help.

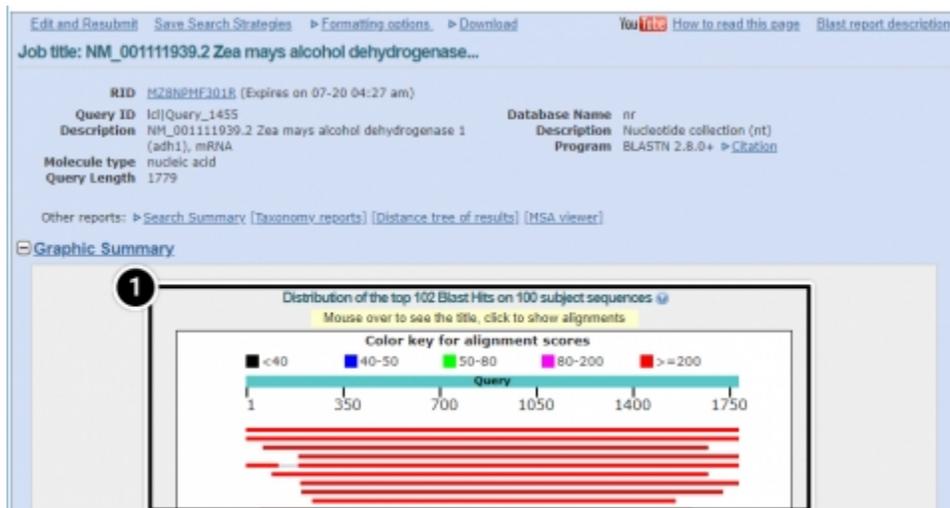
The interface is titled 'Standard Nucleotide BLAST' and features several sections:

- Enter Query Sequence:** A text input field containing 'NM\_001111939.2' is highlighted with a red box and a '1' callout. To its right is a 'Clear' button and a 'Query subrange' section with 'From' and 'To' input fields.
- Or, upload file:** A 'Choose File' button and 'No file chosen' text are present.
- Job Title:** A text input field containing 'NM\_001111939:Zea mays alcohol dehydrogenase...' is highlighted with a red box and a '2' callout. Below it is a small instruction: 'Enter a descriptive title for your BLAST search'.
- Align two or more sequences:** An unchecked checkbox.
- Choose Search Set:** A section with radio buttons for 'Human genomic + transcript', 'Mouse genomic + transcript', and 'Others (nr etc.)'. A dropdown menu is set to 'Nucleotide collection (nr/nt)'. There are also checkboxes for 'Exclude', 'Models (XM/XP)', and 'Uncultured/environmental sample sequences'. A 'Limit to' section has a checkbox for 'Sequences from type material'. An 'Entrez Query' field is also present.
- Program Selection:** A section with radio buttons for 'Highly similar sequences (megablast)', 'More dissimilar sequences (discontiguous megablast)', and 'Somewhat similar sequences (blastn)'. The 'Highly similar sequences (megablast)' option is selected and highlighted with a red box and a '3' callout. Below it is a 'Choose a BLAST algorithm' link.
- BLAST Button:** A large 'BLAST' button is highlighted with a red box and a '4' callout. To its right is the text 'Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)'. Below the button is an unchecked checkbox for 'Show results in a new window'.

At the bottom left, there is a '+ Algorithm parameters' link.

## Try This: Using NCBI BLAST

### 1. **Graphic Summary:** BLAST results that are summarized in a graphic form.



## Try This: Using NCBI BLAST

### 1. **Alignments:** BLAST results that contain sequence alignment information.



The screenshot shows the NCBI BLAST interface for a query sequence. The query is identified as "Zea mays alcohol dehydrogenase 1 (adh1), mRNA" with sequence ID "NM\_001111939.2" and a length of 1779. The results show a single match with a score of 3286 bits (1779), an expect value of 0.0, and 100% identity. The alignment is displayed in a table format with columns for Score, Expect, Identities, Gaps, and Strand. The query and subject sequences are shown with vertical bars indicating the alignment. A "Related Information" link is also visible.

Alignments

Download GenBank Graphics Next Previous Descriptions

Zea mays alcohol dehydrogenase 1 (adh1), mRNA  
Sequence ID: [NM\\_001111939.2](#) Length: 1779 Number of Matches: 1

Range 1: 1 to 1779 [GenBank](#) [Graphics](#) Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
3286 bits(1779)	0.0	1779/1779(100%)	0/1779(0%)	Plus/Plus

Query 1 AAACCCAGBGTCCACGGACACGGCTATGTTCCACTCCAGGTGGAAGCTGCAAGCCCCGTT 60  
Sbjct 1 AAACCCAGBGTCCACGGACACGGCTATGTTCCACTCCAGGTGGAAGCTGCAAGCCCCGTT 60

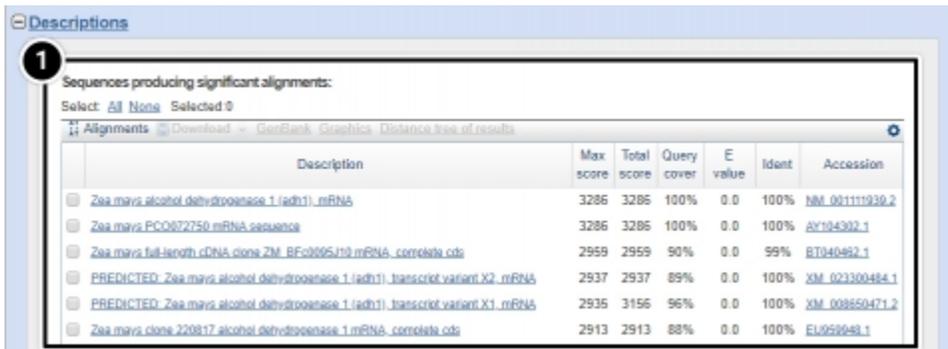
Query 61 TCSCAAGCCGCGCCGTTGGTTTGGCTTGGCCACAGGCGGCCAAACCGCACCTTCTTCCGCT 120  
Sbjct 61 TCSCAAGCCGCGCCGTTGGTTTGGCTTGGCCACAGGCGGCCAAACCGCACCTTCTTCCGCT 120

Query 121 CGTTTCCCATCTCTTCTCTCTTTAGAGCTACCACTATATAAATCAGGGCTCATTTTCTCG 180  
Sbjct 121 CGTTTCCCATCTCTTCTCTCTTTAGAGCTACCACTATATAAATCAGGGCTCATTTTCTCG 180

**Related Information**  
[Gene](#) - associated gene details

## Try This: Using NCBI BLAST

1. **Descriptions:** Accession number and source organism information is provided for sequences producing high alignment scores.



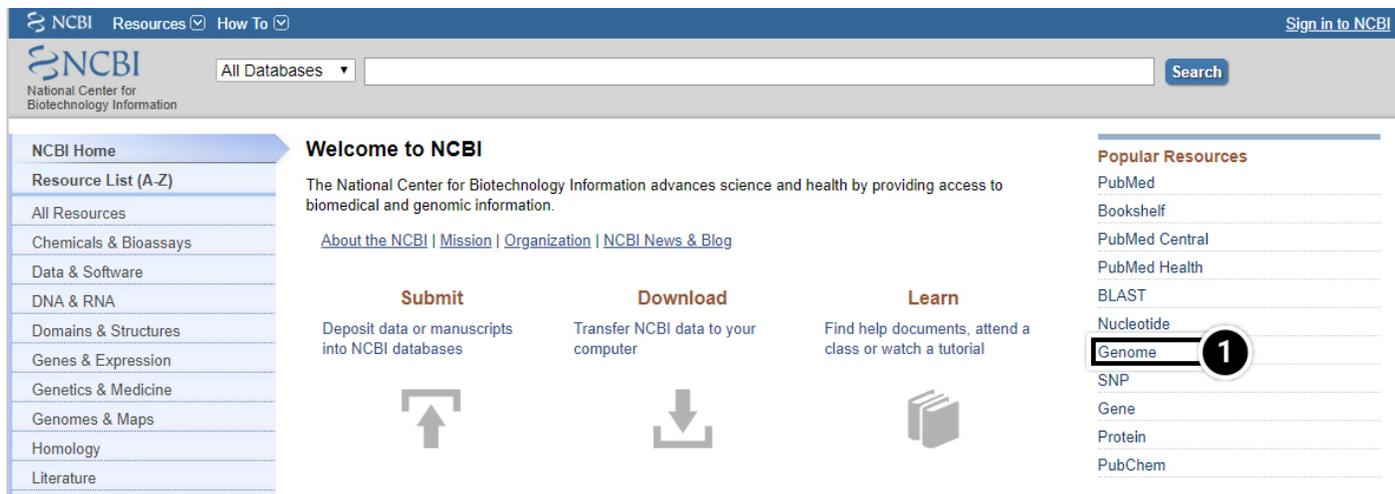
The screenshot shows the 'Descriptions' tab in a BLAST search interface. A circled '1' is in the top left corner. The page title is 'Descriptions'. Below the title, it says 'Sequences producing significant alignments:'. There are two links: 'All None' and 'Selected 0'. Below that, there are several links: 'Alignments', 'Download', 'GenBank', 'Graphics', and 'Distance tree of results'. A table follows with the following columns: 'Description', 'Max score', 'Total score', 'Query cover', 'E value', 'Ident', and 'Accession'. The table contains six rows of data, each with a checkbox on the left.

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Zea mays alcohol dehydrogenase 1 (adh1) mRNA</a>	3286	3286	100%	0.0	100%	<a href="#">NM_001111930.2</a>
<input type="checkbox"/>	<a href="#">Zea mays PC0072750 mRNA sequence</a>	3286	3286	100%	0.0	100%	<a href="#">AY104302.1</a>
<input type="checkbox"/>	<a href="#">Zea mays full-length cDNA clone ZM_BF009510 mRNA - complete cds</a>	2959	2959	90%	0.0	99%	<a href="#">BT040462.1</a>
<input type="checkbox"/>	<a href="#">PREDICTED: Zea mays alcohol dehydrogenase 1 (adh1) transcript variant X2 mRNA</a>	2937	2937	89%	0.0	100%	<a href="#">XM_023300484.1</a>
<input type="checkbox"/>	<a href="#">PREDICTED: Zea mays alcohol dehydrogenase 1 (adh1) transcript variant X1 mRNA</a>	2935	3156	96%	0.0	100%	<a href="#">XM_008850471.2</a>
<input type="checkbox"/>	<a href="#">Zea mays clone 220817 alcohol dehydrogenase 1 mRNA - complete cds</a>	2913	2913	88%	0.0	100%	<a href="#">EU050648.1</a>

## Try This: Using NCBI BLAST

### Step 4: Locating adh1 on a chromosome

1. From the NCBI home page, select **Genome**.

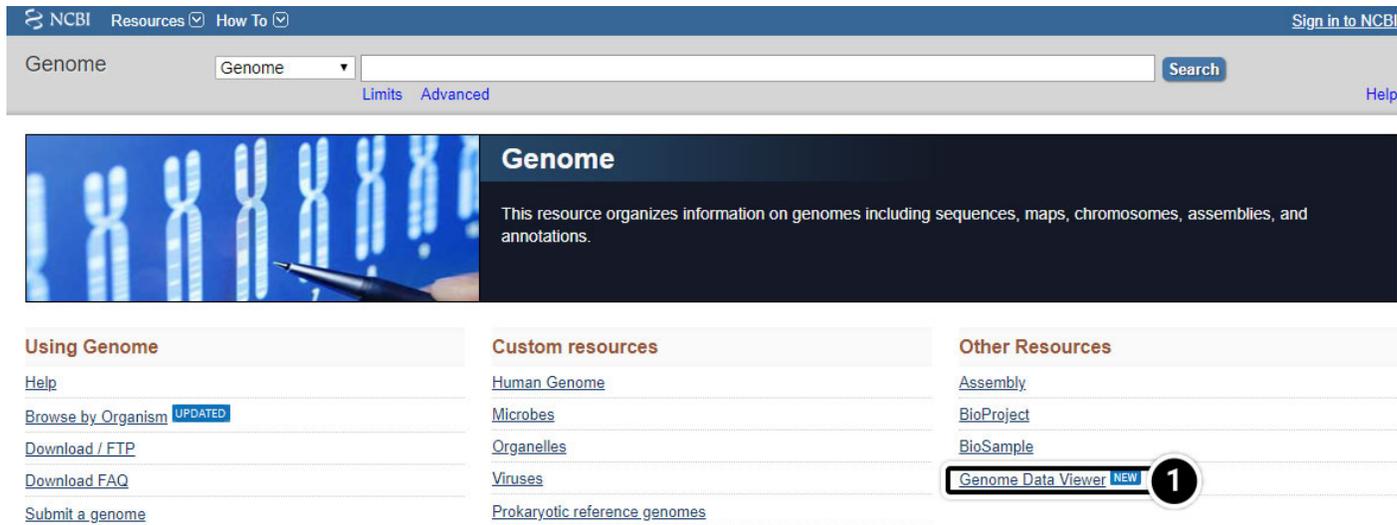


The screenshot shows the NCBI homepage. At the top, there is a navigation bar with 'NCBI Resources' and 'How To' menus, and a 'Sign in to NCBI' link. Below this is a search bar with a dropdown menu set to 'All Databases' and a 'Search' button. On the left side, there is a vertical menu with various categories: NCBI Home, Resource List (A-Z), All Resources, Chemicals & Bioassays, Data & Software, DNA & RNA, Domains & Structures, Genes & Expression, Genetics & Medicine, Genomes & Maps, Homology, and Literature. The 'NCBI Home' item is highlighted with a blue arrow. In the center, there is a 'Welcome to NCBI' section with a brief description and links for 'About the NCBI', 'Mission', 'Organization', and 'NCBI News & Blog'. Below this are three main action buttons: 'Submit' (Deposit data or manuscripts into NCBI databases), 'Download' (Transfer NCBI data to your computer), and 'Learn' (Find help documents, attend a class or watch a tutorial). On the right side, there is a 'Popular Resources' section with a list of links: PubMed, Bookshelf, PubMed Central, PubMed Health, BLAST, Nucleotide, Genome (highlighted with a black box and a circled '1'), SNP, Gene, Protein, and PubChem.

## Try This: Using NCBI BLAST

### Locating adh1 on a chromosome

1. From the genome page, select **Genome Data Viewer (previously known as Map Viewer)**.



NCBI Resources How To Sign in to NCBI

Genome Genome Search Limits Advanced Help

### Genome

This resource organizes information on genomes including sequences, maps, chromosomes, assemblies, and annotations.

#### Using Genome

- [Help](#)
- [Browse by Organism](#) **UPDATED**
- [Download / FTP](#)
- [Download FAQ](#)
- [Submit a genome](#)

#### Custom resources

- [Human Genome](#)
- [Microbes](#)
- [Organelles](#)
- [Viruses](#)
- [Prokaryotic reference genomes](#)

#### Other Resources

- [Assembly](#)
- [BioProject](#)
- [BioSample](#)
- Genome Data Viewer** **NEW** **1**

## Try This: Using NCBI BLAST

### Locating adh1 on a chromosome

1. Within **Genome Data Viewer** home you can select your organism or species.

The screenshot displays the NCBI Genome Data Viewer (GDV) interface. At the top, the navigation bar includes the NIH logo, "U.S. National Library of Medicine", "NCBI National Center for Biotechnology Information", and a "Log in" link. The main header features the "Genome Data Viewer" title and a description: "GDV is a genome browser supporting the exploration and analysis of more than 600 eukaryotic RefSeq genome assemblies." Below this, a "Select organism" dropdown menu is highlighted with a black box and a circled "1". The dropdown list contains: "Zea mays (maize)", "Zea mays (maize)", "Oryzae", "Zea", and "Acanthisittidae (New Zealand wrens)". To the right of the dropdown, a phylogenetic tree shows "maize" (blue circle), "sorghum" (green circle), and "foxtail millet" (green circle) connected by green lines. On the right side of the interface, the "Zea mays (maize) genome" panel is visible. It includes a search bar labeled "Search in genome" with the placeholder text "Location, gene or phenotype" and a search icon. Below the search bar, it provides examples: "Examples: adh1, chr1:278820000-278826000, DNA repair". Under the "Assembly" section, "B73 RefGen\_v4" is selected. At the bottom of the panel are two buttons: "Browse genome" and "BLAST genome".

## Try This: Using NCBI BLAST

### Locating adh1 on a chromosome

1. Try searching the Zea mays genome for the adh1 gene.

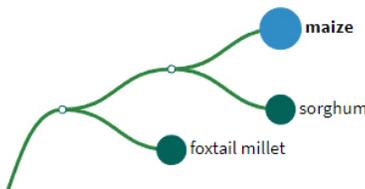
NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Log in

# Genome Data Viewer

GDV is a genome browser supporting the exploration and analysis of more than 600 eukaryotic RefSeq genome assemblies. ⓘ

Select organism

Zea mays (maize)



maize  
sorghum  
foxtail millet

**1**

### Zea mays (maize) genome

Search in genome

adh1

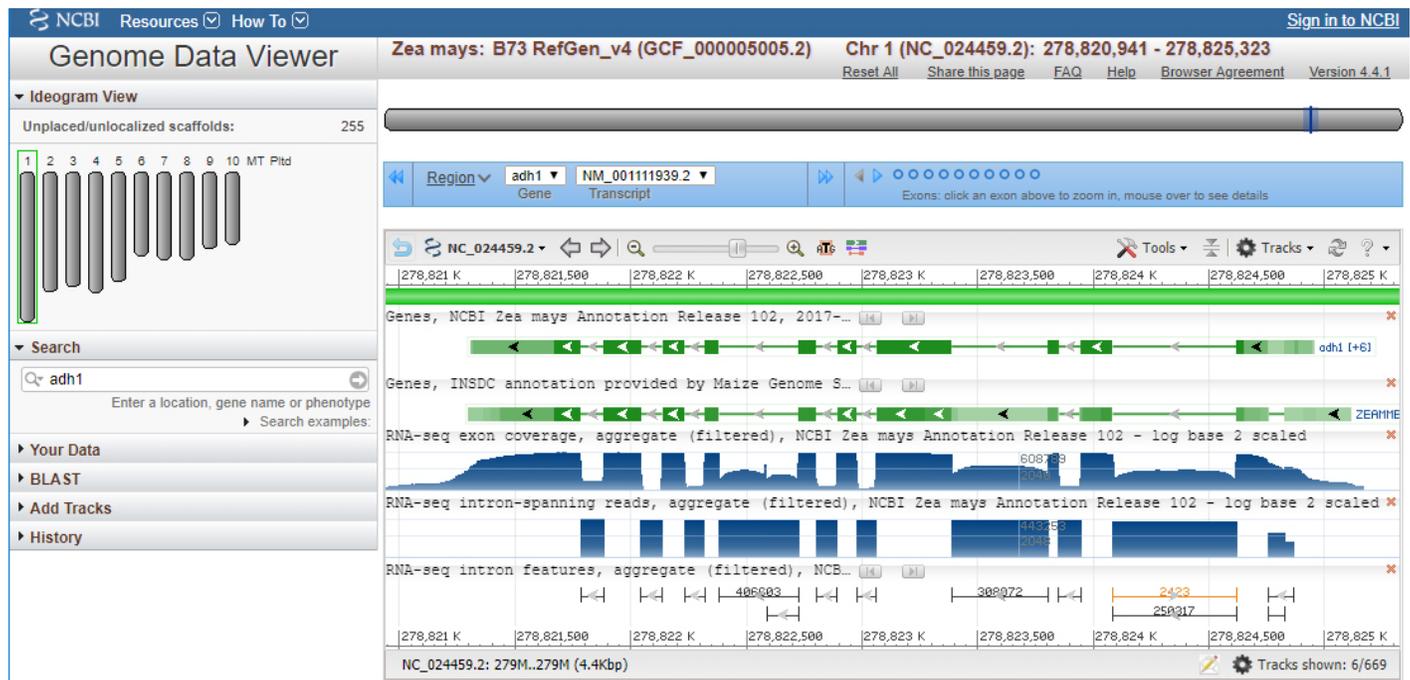
Examples: adh1, chr1:278820000-278826000, DNA repair

Assembly

B73 RefGen\_v4

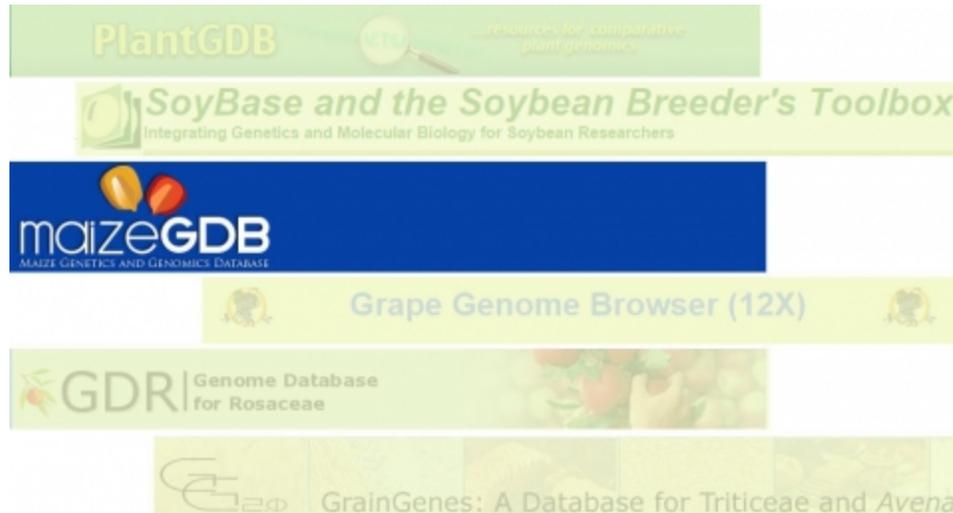
## Try This: Using NCBI BLAST

The NCBI Map View search for *adh1* on the maize genome produces these results. "Ideogram view" highlights chromosome 1 to show that the *adh1* gene is located on chromosome 1.



# Plant Species Sequence Databases

The advent of genomics has resulted in a number of plant species specific sequence databases. For this lesson, Maize Genetics and Genomics Database (MaizeGDB) will be the focus.



# MaizeGDB

MaizeGDB was first released in 1991 (as MaizeDB) and has transitioned from a focus on curation of genetic maps and stocks to the handling of reference maize genome sequence, multiple maize genomes, and sequence-based gene expression data. MaizeGDB relies on the research community for data and on expertise distributed across the USA. We recommend the use of an internet browser other than Internet Explorer (e.g. Google Chrome) to access the [MaizeGDB site](#).

1. Tutorials are available by clicking the video icon.

Chinese Version (中文版) Download Log in/Create account

maizeGDB  
MAIZE GENETICS AND GENOMICS DATABASE

Home About Community Genome Browsers Genomes Tools Data Centers Search Feedback

The AGP v4 genome has officially been published and released. To convert between the old GRMZM annotations and the new v4 Zm annotations you may use our [translation tool](#) or [download the full gene model associations list between v3 and all other assemblies in our database](#).

We would like to remind the maize community that under the [Toronto Agreement](#), whole-genome research on the B104 genome assembly may not be submitted for publication until the official B104 genome paper is published.

**Welcome to MaizeGDB!**

MaizeGDB is a community-oriented, long-term, federally funded informatics service to researchers focused on the crop plant and model organism *Zea mays*.

**Quick Links**

Genome Browser BLAST qTeller MaizeMine

Diversity SNPs Traits Newly Characterized Genes Metabolic Pathways Expression

Hot New Papers Project Portal Maize Meeting Bin Viewer

**Reference Assembly**

B73 ASSEMBLY B73 ANNOTATION ALL GENOMES

[Report assembly error](#) [Report gene model error](#)

B73 RefGen\_v4 is now available

**News**

3, July 2018: qTeller, a comparative RNA-seq expression platform, is now available at MaizeGDB with multiple expression datasets mapped to B73v4! Eventually, datasets mapped to Mo17, W22, PH207 and more will be incorporated!

12, June 2018: A [MaizeGDB pedigree viewer](#) is now available!

5, June 2018: RNA-Seq tracks from [Walley JW et. al \(2016\)](#) are now available on our B73 v4 genome browser!

[more news](#)

**MaizeGDB**  
Maize Genetics and Genomics Database

[Visit the classic MaizeGDB website.](#)

**Database**  
Last update: July 3, 2018  
Next update: August 7, 2018

**Funding Sources**

USDA United States Department of Agriculture NSF

1  
YouTube RSS Facebook Twitter  
Like 436 Share

# MaizeGDB: Tutorials

[Useful tutorials](#) are available to help the user familiarize with MaizeGDB.

The screenshot displays the MaizeGDB website interface. At the top, the MaizeGDB logo and 'Maize Genetics and Genomics Database' are visible, along with a 'Log In' link. A large banner features the text 'MaizeGDB Tutorial' and 'Using the MaizeGDB Genome Browser' over a background of corn cobs. To the right of the banner is a 'MaizeGDB Genome Browser Tutorial' section with a 'Watch Now' button. Below the banner is a search bar. The main content area is divided into two columns. The left column contains a grid of tutorial cards, each with a MaizeGDB logo, a title, a brief description, and a 'read more' link. The right column is titled 'Links to Other Tutorials' and contains a list of links to various resources. At the bottom right, there is a 'Tags' section.

**MaizeGDB Tutorial**  
Maize Genetics and Genomics Database

**Using the MaizeGDB Genome Browser**

**MaizeGDB Genome Browser Tutorial**  
Oct 2010: Learn the difference between the various B73 Genome Assemblies, and learn how to view them with the MaizeGDB Genome Browser. ...  
[Watch Now](#)

**MaizeGDB Tutorial**  
Using the MaizeGDB Genome Browser  
MaizeGDB Genome Browser Tutorial  
Oct 2010: Learn the difference ... read more  
October 3rd, 2010

**MaizeGDB Tutorial**  
How the B73 Genome Sequencing Consortium Sequenced B73  
How was each B73 BAC sequenced?  
The Maize Genome Sequencing Co ... read more  
August 17th, 2010

**MaizeGDB Tutorial**  
What's B73 RefGen\_v1?  
All About B73 RefGen\_v1  
Wonder what a "pseudomolecule" ... read more  
August 17th, 2010

**MaizeGDB Tutorial**  
Find the Genetic Map

**Links to Other Tutorials**

- [Bill Tracy on Corn Breeding](#)
- [Controlled Pollinations of Maize](#)
- [Description of Genetic Maps in MaizeGDB](#)
- [Learn how to add your own custom tracks to the MaizeGDB genome Browser from G-Browse](#)
- [Learn how to sequence DNA from JGI](#)
- [Maize Genetics, Genomics, and Bioinformatics Workshop 2004, Mexico](#)
- [Maize Pollination Videos](#)
- [Wisconsin Plant Breeding Maize Pollination Video on Youtube](#)
- [Wisconsin Plant Breeding Maize Pollination Videos](#)

**Tags**

# Try This: Using MaizeGDB

## Perform a Basic Search

Similar to NCBI, the MaizeGDB is a composite database allowing you to search broadly among databases or to restrict your query to a single database.

1. Open your web browser and go to <https://www.maizegdb.org>.
2. Enter **adh1** into the search box.
3. Press **Enter** or click the Search icon to search within all available data.

Click the image below to see a larger version.



## Explore the Search Results

This search will lead you to a window containing various options.

1. Click on **Locus Lookup (1)** in the left-hand menu.
2. Click on **Gene Models (15)** in the left-hand menu.

Explore the other data available to you by clicking the links in the green box. Click the image below to see a larger version.

The image displays two screenshots of the MaizeGDB Data Search interface, illustrating the steps to explore search results.

**MaizeGDB Data Search**

**1** Locus Lookup (1)

Gene Models (15)  
Variation (100)  
Locus (8)  
Phenotype (4)  
Reference (100)  
Sequence (100)  
Stock (22)  
Marker (3)  
Synonyms (1263)  
Locus Lookup (1)  
CFL (0)  
EST (0)  
BAC (0)

Locus Lookup results based on:  
**Gene models:**  
The Locus **adh1** is between **273,983,286** and **273,986,641** on Chromosome **1** based on gene model **GRMZM2G442658**.  
**Physically mapped:**  
The Locus **adh1** has not been physically mapped.  
**Placed BACs:**  
The Locus **adh1** is not associated with physically mapped probes.  
**Genetically mapped:**  
The Locus **adh1** is between **274,684,822** and **276,776,374** on Chromosome **1** based on the following nearest loci on the IBM2 2008 Neighbors map that are physically placed: (TDP9031 and RF639426).

**2** Gene Models (15)

**Gene Search Results**

There are **15** genes/gene models matching the term **adh1**.  
Results may include multiple versions and transcripts of the same gene model.

1 2 

No gene models were found matching the term **adh1**.

## Access the Genome Browser

Access the genome browser to obtain information about maize *adh1*.

1. Type **adh1** into the Search bar.
2. Select **loci** from the Search options.
3. Click the Search icon.

This time when the results load, you will see only the loci associated with *adh1*.

4. Click the link for **adh1 alcohol dehydrogenase1** to take a closer look at the gene.

Click the image below to see a larger version.

The screenshot displays the MaizeGDB Data Search interface. At the top, there is a search bar containing the text 'adh1' (marked with a '1'). To the right of the search bar is a search icon (marked with a '3'). Below the search bar, a dropdown menu is open, showing various search options, with 'loci' selected (marked with a '2'). The main content area is titled 'MaizeGDB Data Search' and 'Locus Search Results - Complete List'. It states 'Here are the 8 loci matching the term adh1.' and lists several entries. The first entry, 'Gene: adh1 alcohol dehydrogenase1' (marked with a '4'), is highlighted. The page also features a navigation menu with options like Home, About, Community, Genome Browsers, Genomes, Tools, and Data Centers. A notice at the top indicates that searches are now run over HTTPS.

## Explore the Gene Record

The locus record screen provides detailed information on the *adh1* gene. Explore the genetic information for ***adh1* alcohol dehydrogenase1**.

1. Click on **Chromosome Coordinates** when ready to proceed.

Click the image below to see a larger version.

**Gene record**

[Report an assembly error](#) [Report a gene model error](#)

GRMZM2G442658 (***adh1* - alcohol dehydrogenase1**) [Classical Gene List]

**GENE MODEL** **SEQUENCE** **GENETIC INFORMATION**

**Note:** A gene is a specific type of locus; the word "gene" should not be considered to be synonymous with "locus".

**Overview**

**Gene name:** *adh1* (alcohol dehydrogenase1)  
**Synonyms:** Adh (per Schwartz, D), Adh-1 (per Various), Adh2 (per Scandalios, J), AY111936, bnl(*adh1*), CL22280\_1, IDP1964, IDP35, magi75238, np121-*adh1* (per Wright, S), np121(*adh1*) (per Wright, S), PC0141653, rgpc496a(*adh*), umc1726, umc(*adh1*) (per Burr, B)  
**Gene Products:**  
alcohol dehydrogenase

**Gene**

- Overview
- Annotations
- Chromosome Coordinates** **1**
- Map Coordinates
- Nearby Loci
- Allele/variation/polymorphism
- Genetic information
- References
- External Links

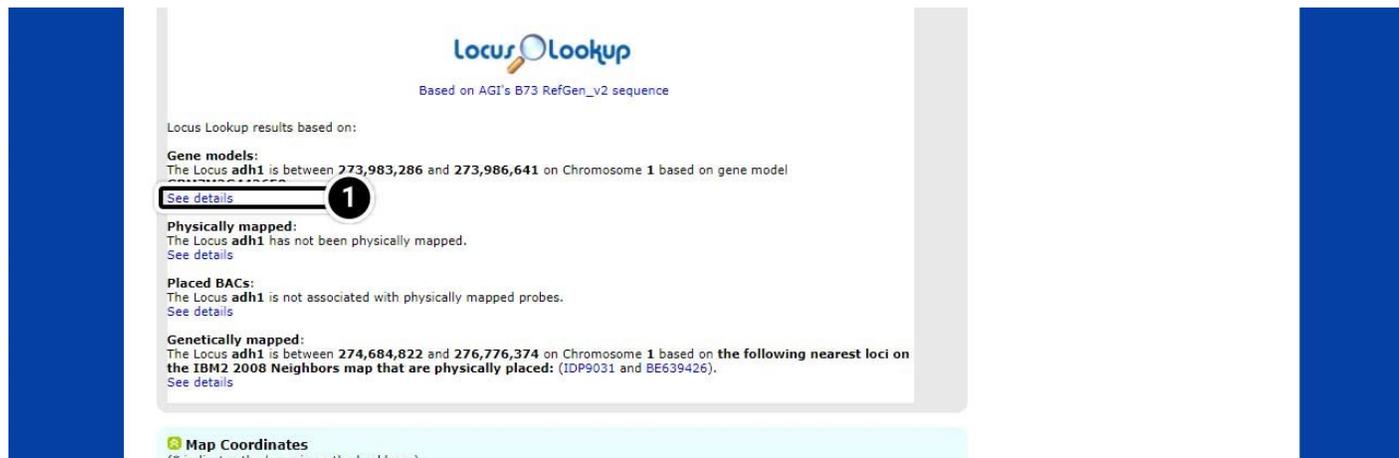
[Open Previous Search Results](#)

## See Details in Locus Lookup

The page will scroll down to the Locus Lookup section.

1. Click on **Show details** to expand this section of the results.

Click the image below to see a larger version.



The screenshot displays the Locus Lookup interface. At the top, the logo "Locus Lookup" is centered, with a magnifying glass icon over the word "Lookup". Below the logo, it states "Based on AGI's B73 RefGen\_v2 sequence". The main content area is titled "Locus Lookup results based on:" and lists several categories:

- Gene models:** The Locus **adh1** is between **273,983,286** and **273,986,641** on Chromosome **1** based on gene model **AT1G15430.1**. A "See details" link is present, with a large number "1" overlaid on it.
- Physically mapped:** The Locus **adh1** has not been physically mapped. A "See details" link is present.
- Placed BACs:** The Locus **adh1** is not associated with physically mapped probes. A "See details" link is present.
- Genetically mapped:** The Locus **adh1** is between **274,684,822** and **276,776,374** on Chromosome **1** based on **the following nearest loci on the IBM2 2008 Neighbors map that are physically placed:** (IDP9031 and BE639426). A "See details" link is present.

At the bottom of the screenshot, a light blue bar contains the text "Map Coordinates" with a small icon, followed by "(# indicates the locus is on the backbone)".

## Expanded Details in Locus Lookup

When the details have loaded, explore the available information.

Note the position of *adh1* based on “AGIs B73 RefGen\_v2 sequence” (*adh1* is located between 273,983,286 and position 273,986,641 on chromosome 1.

1. Click on the map image to launch the MaizeGDB genome browser.

Click the image below to see a larger version.

The screenshot displays the Locus Lookup interface. At the top, the logo "Locus Lookup" is shown with a magnifying glass icon, followed by the text "Based on AGI's B73 RefGen\_v2 sequence". Below this, it states "Locus Lookup results based on:" and "Gene models: The Locus **adh1** is between 273,983,286 and 273,986,641 on Chromosome 1 based on gene model **GRMZM2G442658**." A link "Hide details" is provided. The text "This region is 3,355 base pairs. Click on images to go to the MaizeGDB genome browser." is followed by two views: "Genome View:" and "Genome Browser View:". The "Genome View" shows a chromosome map with 10 chromosomes, where chromosome 1 is highlighted in green and circled with a "1" in a black circle. The "Genome Browser View" shows a detailed view of the locus on chromosome 1, with coordinates 273,983,286 to 273,986,641. It displays the B73 RefGen\_v2 sequence, gene models for GRMZM2G442658, and various transcript models (T01-T07) with exons and introns represented by green and black bars.

## View Datasets in the MaizeGDB Genome Browser

The MaizeGDB Genome Browser is displayed. Click the image below to see a larger version.

1. Here you can use the other datasets available in MaizeGDB including **B73 RefGen\_v1 sequence**, **B73 RefGen\_v3 sequence**, **B73 RefGen\_v4 sequence**, and **BAC-based genome assembly**.

The screenshot displays the MaizeGDB Genome Browser interface. At the top, there is a navigation bar with links for Home, About, Community, Genome Browsers, Genomes, Tools, Data Centers, Search, and Feedback. A search box is located in the top right corner. Below the navigation bar, there is a section for reporting assembly or gene model structure problems, followed by a menu with File and Help options. The main content area features a title for the dataset: "Maize B73 RefGen\_v2: 3.356 kbp from Chr1:273,983,286..273,986,641". A text box explains that this region lies at approximately 323.8 cM on the ISU Integrated IBM 2009 genetic map for chromosome 1. Below this, there are several green buttons: "GO TO THE B73\_REFGEN\_V1 BROWSER", "GO TO THE B73\_REFGEN\_V3 BROWSER", "GO TO THE B73\_REFGEN\_V4 BROWSER", "See this region at maizesequence.org", "See this region at PlantGDB", and "See this region at CoGe". A section titled "SEE THIS REGION AT GENOMAIZE PSU" includes links for GBrowse2 Documentation, Browser, Select Tracks, Snapshots, Community Tracks, Custom Tracks, and Preferences. The "Search" section contains a "Landmark or Region" input field with the value "Chr1:273,983,286..273,986,641" and a "Search" button. Below this is a "Locus Lookup" section with a "Locus Lookup" button and a "Design PCR primers" dropdown menu. A "Data Sources" dropdown menu is set to "Maize B73 RefGen\_v2" and is highlighted with a red circle and the number "1". The "Overview" section at the bottom shows a genomic map of chromosome 1 with a scale from 0M to 300M and a vertical line indicating the current region.

### MaizeGDB Study Questions

Find the position of *adh1* on chromosome 1 in "B73 RefGen\_v1 sequence" and "BAC-based genome assembly."

B73 RefGen\_v2 sequence: Between 273,983,286 and 273,986,641

B73 RefGen\_v1 sequence: Between  and

BAC-based genome assembly: Between  and

Check

What might be the reason for the discrepancy in *adh1* position among the three datasets?

Enter your answer here.

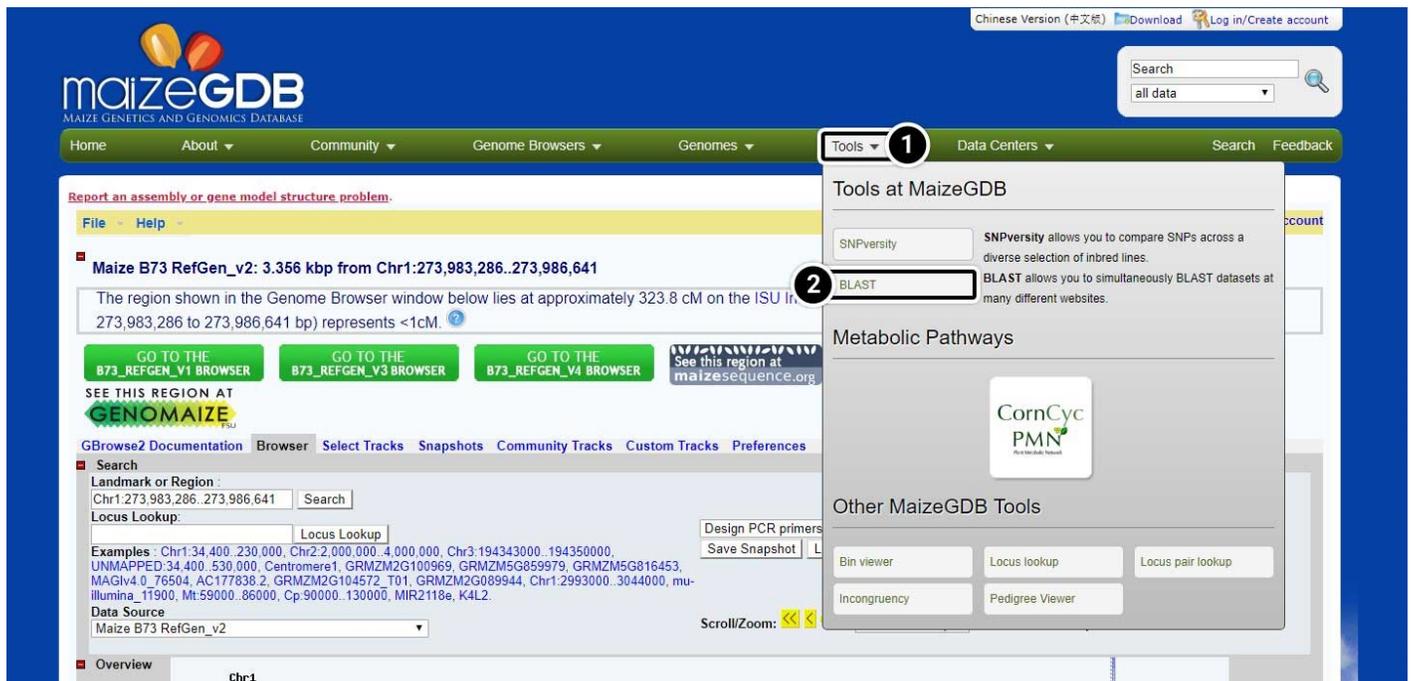
Show Answer

## Navigate to BLAST

Next, we will conduct a BLAST search for *adh1* in maize GDB using *adh1* mRNA from GenBank.

1. In the navigation bar, hover over **Tools**
2. Then click the **BLAST** button.

Click the image below to see a larger version.



The screenshot displays the MaizeGDB website interface. The navigation bar at the top includes links for Home, About, Community, Genome Browsers, Genomes, Tools, Data Centers, Search, and Feedback. The 'Tools' dropdown menu is open, showing options for SNPiversity, BLAST, Metabolic Pathways, and Other MaizeGDB Tools. The 'BLAST' option is highlighted with a red box and a circled '2'. A circled '1' is placed over the 'Tools' dropdown arrow in the navigation bar. The main content area shows a search for 'Maize B73 RefGen\_v2: 3.356 kbp from Chr1:273,983,286..273,986,641' and a search bar for 'Landmark or Region'.

## Input the BLAST Parameters

1. Enter the adh1 mRNA sequence in FASTA format in the box.
2. Use the default parameters to search for adh1 and click the BLAST button.

**MaizeGDB BLAST**

The GenBank BLAST targets have been removed.  
Please use the NCBI BLAST tool to BLAST against GenBank targets.

BLAST up to 5 sequences of up to 35,000 bp total length against maize datasets. [?](#)

Other BLAST pages: [Gramene.org](#), [PlantGDB](#), [NCBI](#), [Panzea](#)

**Step 1: input your sequences** (Raw, FASTA, or Genbank IDs) [?](#)

Sequence type:  Nucleotides  Amino Acids  
Enter your sequence: ([example](#)) **1**

or upload from file:  No file chosen

**Step 2: select datasets** [?](#)

All supported  
Select a dataset

[B73 RefGen\\_v4 \(CSHL\)](#)

**Step 3: select BLAST parameters** [?](#)

Optimize for:  High similarity  Low similarity  Short sequences  
 advanced settings

**Step 4: select output type** [?](#)

Enhanced output  BLAST table output  BLAST text output

Name your sequence (optional)  [?](#)  **2**

E-mail (optional)  [?](#)

**recent queries**  
(none) [\[see all\]](#)

**Overview of BLAST**  
Basic Local Alignment Search Tool, or BLAST, is an algorithm for comparing similarity in biological sequences. It operates on such sequences as the sequence of amino acids in proteins or the sequence of nucleotides of DNA or RNA molecules.  
Learn more about BLAST [here](#).  
The BLAST home page is [here](#).

## View the BLAST Results

The table of BLAST results includes information on chromosomes, probability values, sequence identity, and the number of likely candidates (hits). Also, you can view a representation of the entire genome in the context of where *adh1* may be located.

1. Click the arrow next to “Whole genome view” to see the entire genome in context of where *adh1* may be located.

**BLAST results**

[Edit this query and re-submit](#)

Result set name:  [change name](#)

Saved for 1 week here:  [change](#)

Query sequence:  [change](#)

(Click [?](#) for more information.)

---

**MaizeGDB - B73 RefGen\_v4**  
(10)

### B73 RefGen\_v4 (CSHL)

Input parameters: E-value cutoff: 1e-4, max hits: 500  
Description: BLAST was executed at MaizeGDB, against the sequence database B73 RefGen\_v4 (CSHL), using BLAST program blastn.  
B73 Reference Genome, assembly B73 RefGen\_v4  
Please cite [Jiao, Y et al. \(2017\)](#) if you use these data.

Query sequence 1: [gi|162463221|ref|NM\\_001112073.1| Zea mays liguleless1 \(lg1\), mRNA](#)

Target ID	definition	e-value	percent identity	#
Chr2	Chr2	0	99.19	4
Chr10	Chr10	3.834e-158	83.02	7
Chr6	Chr6	1.634e-37	86.27	3
Chr9	Chr9	1.272e-33	85.23	2
Chr5	Chr5	1.272e-33	84.87	2
Chr4	Chr4	2.754e-30	83.55	7
Chr8	Chr8	1.005e-19	81.48	2
Chr3	Chr3	4.674e-18	80.74	3
Chr7	Chr7	2.813e-15	80.00	3
Chr1	Chr1	7.877e-11	81.25	4

[Download target sequences as fasta](#)

Whole genome view **1**

Visual alignment for Chr2

Alignment details for Chr2

**recent queries**  
(none) [\[see all\]](#)

**Overview of BLAST**  
Basic Local Alignment Search Tool, or BLAST, is an algorithm for comparing similarity in biological sequences. It operates on such sequences as the sequence of amino acids in proteins or the sequence of nucleotides of DNA or RNA molecules.  
[Learn more about BLAST here.](#)  
The BLAST home page is [here.](#)

## Explore the Whole Genome View

The whole genome view allows visualization of the 10 chromosomes of maize including, the predicted position match the *adh1* sequence.

1. Click on "Chr1" corresponding to the red box on Chromosome 1 (E-value = 0).
2. Now, click on "View at MaizeGDB", next to the hit on Chr1.

Query sequence 1: NM\_001111939.2 *Zea mays* alcohol dehydrogenase 1 (*adh1*), mRNA

### Target summary table

Target ID	definition	e-value	percent identity	#
<a href="#">Chr1</a>	Chr1	0	100.00	28
<a href="#">Chr4</a>	Chr4	1.533e-68	81.62	16
<a href="#">Chr10</a>	Chr10	7.387e-42	94.02	6
<a href="#">Chr5</a>	Chr5	5.875e-23	81.82	13
<a href="#">Chr3</a>	Chr3	4.606e-14	94.55	6
<a href="#">Chr2</a>	Chr2	2.143e-12	90.32	7
<a href="#">Chr9</a>	Chr9	7.708e-12	94.12	5
<a href="#">Chr7</a>	Chr7	2.772e-11	92.45	5
<a href="#">Chr8</a>	Chr8	9.971e-11	92.31	4
<a href="#">Chr6</a>	Chr6	4.639e-9	90.38	7
UNMAPPED	UNMAPPED	4.639e-9	91.84	2

[Download target sequences as fasta](#)

### Whole genome view



Visual alignment for Chr1

Alignment details for Chr1

Alignment details for hit #1 for Chr1

Score = 465 bits (859.812), Expect = 0  
Identities = 465/465 (1.0000%), Gaps = 0 (0.0000%)  
Strand = Plus / Minus



```
Query 1315      AGGAGCTGGAGGTGGAGAAGTTCATCACGCACAGCGTCCCGTTCGCCGAGATCAACAAGG 1374
                |||
Sbjct 278821785  AGGAGCTGGAGGTGGAGAAGTTCATCACGCACAGCGTCCCGTTCGCCGAGATCAACAAGG 278821726

Query 1375      CGTTCGACCTGATGGCCAAGGGGGAGGGCATCCGCTGCATCATCCGCATGGAGAACTAGA 1434
                |||
Sbjct 278821725  CGTTCGACCTGATGGCCAAGGGGGAGGGCATCCGCTGCATCATCCGCATGGAGAACTAGA 278821666

Query 1435      TTTCGCTGTCTAGTTTGTGATCTGGCC TGGGCTTGGGGTTAATAAAAAGAGGCAATGCTAG 1494
                |||
Sbjct 278821665  TTTCGCTGTCTAGTTTGTGATCTGGCC TGGGCTTGGGGTTAATAAAAAGAGGCAATGCTAG 278821606

Query 1495      CCTGCCCTTTCGATGAGGAGGTACATACACGCTGGCGATGGACCGCGCTTGTGTGTCGCG 1554
                |||
Sbjct 278821605  CCTGCCCTTTCGATGAGGAGGTACATACACGCTGGCGATGGACCGCGCTTGTGTGTCGCG 278821546

Query 1555      TTCAGTTTGGCTTTTGCCAAGCAGTAGGGTAGCTTCCCGTGTGCGTAATTATATGGTATG 1614
                |||
Sbjct 278821545  TTCAGTTTGGCTTTTGCCAAGCAGTAGGGTAGCTTCCCGTGTGCGTAATTATATGGTATG 278821486

Query 1615      AACCATCACCTTTTGGCTCTACATGGTATGAACGTAAGATACAAATTC AACTACCTCTA 1674
                |||
Sbjct 278821485  AACCATCACCTTTTGGCTCTACATGGTATGAACGTAAGATACAAATTC AACTACCTCTA 278821426

Query 1675      GCTCGCTTGTGTGGTATCTGTATCAGTATTCATGTGTTGTTTGTGTTATGTGTTTGTG 1734
                |||
Sbjct 278821425  GCTCGCTTGTGTGGTATCTGTATCAGTATTCATGTGTTGTTTGTGTTATGTGTTTGTG 278821366

Query 1735      CTTGTATTTGCTGGTGCTTGTATCGCGGGATGCAATGAGTTGTTG 1779
                |||
Sbjct 278821365  CTTGTATTTGCTGGTGCTTGTATCGCGGGATGCAATGAGTTGTTG 278821321
```

## Change the Data Source

The selection you made on the last screen will open a new window containing information on the position of *adh1* and data sources.

1. Click on the pull-down menu of “data source” (arrow) to explore other data sets.

The screenshot displays the Maize GDB (Maize Genetics and Genomics Database) interface. At the top, there is a navigation bar with links for Home, About, Community, Genome Browsers, Genomes, Tools, Data Centers, Search, and Feedback. A search bar is located in the top right corner. Below the navigation bar, there is a section for reporting assembly or gene model structure problems, followed by a 'File - Help' menu and a 'Log in / create account' link.

The main content area features a section titled 'Maize B73 RefGen\_v2: 3.356 kbp from Chr1:273,983,286..273,986,641'. A text box explains that this region is located at approximately 323.8 cM on the ISU Integrated IBM 2009 genetic map for chromosome 1. Below this, there are several green buttons: 'GO TO THE B73\_REFGEN\_V1 BROWSER', 'GO TO THE B73\_REFGEN\_V3 BROWSER', and 'GO TO THE B73\_REFGEN\_V4 BROWSER'. There are also links to 'maizesequence.org', 'PlantGDB', and 'CoGe'.

Underneath, there is a section titled 'SEE THIS REGION AT GENOMAIZE PSU'. Below that, there are links for 'GBrowse2 Documentation', 'Browser', 'Select Tracks', 'Snapshots', 'Community Tracks', 'Custom Tracks', and 'Preferences'.

The 'Search' section includes a 'Landmark or Region' input field with the value 'Chr1:273,983,286..273,986,641' and a 'Search' button. Below this is a 'Locus Lookup' section with a 'Locus Lookup' button and a 'Design PCR primers' dropdown menu. There are also 'Save Snapshot' and 'Load Snapshot' buttons.

Examples of genomic coordinates are listed: Chr1:34,400..230,000, Chr2:2,000,000..4,000,000, Chr3:194343000..194350000, UNMAPPED:34,400..530,000, Centromere1, GRMZM2G100969, GRMZM5G859979, GRMZM5G816453, MAGiv4.0\_76504, AC177838.2, GRMZM2G104572\_T01, GRMZM2G089944, Chr1:2993000..3044000, mu-illumina\_11900, Mt:59000..86000, Cp:90000..130000, MIR2118e, K4L2.

The 'Data Source' dropdown menu is highlighted with a red circle and the number '1', showing 'Maize B73 RefGen\_v2' selected. Below the dropdown, there are 'ScrollZoom' controls with arrows and a 'Show 3.356 kbp' button, along with a 'Flip' checkbox.

At the bottom, there is an 'Overview' section showing a genomic map of chromosome 1 with a scale from 0M to 300M. A vertical line indicates the current region's position on the chromosome.

### *BLAST Study Questions*

Find the position of *adh1* using BLAST results and all versions of datasets. Compare BLAST results with maize genome browse results. If the results are not similar, what might be the reason for the differences?

Position of the *adh1* locus in:

#### **B73 RefGen\_v2 sequence**

Maize genome browse result: Between 273,983,286 and 273,983,641

BLAST: Between  and

#### **B73 RefGen\_v1 sequence**

Maize genome browse result: Between 272,905,082 and 272,905,552

BLAST: Between  and

#### **BAC based genome assembly**

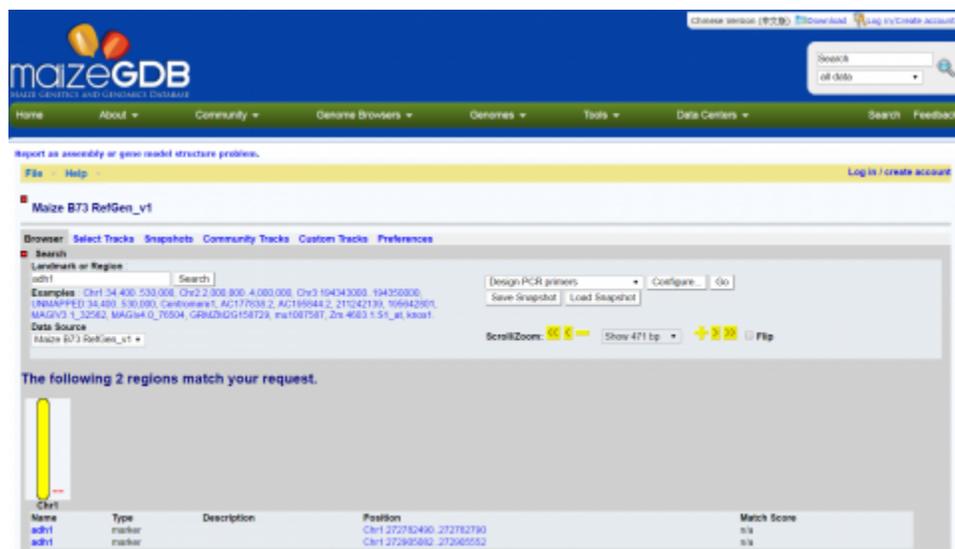
Maize genome browse result: Between 262,399,900 and 262,556,700

BLAST: Since the introduction of the pseudomolecule, the BAC based genome assembly is no longer supported.

Check

## Possible Explanation for BLAST Study Question Results

One reason for discrepancies might be that there are in this genomic region several copies of the gene (eventually ancient duplication no longer actively transcribed due to mutations or whatever). Depending on the origin of your query sequence you use to find the gene, they might show different hit scores from these versions of the gene. As for the version2 pseudo- molecule the location seems to be quite similar...



The screenshot shows the Maize GDB website interface. At the top, there is a navigation bar with links for Home, About, Community, Genome Browsers, Genomes, Tools, Data Centers, Search, and Feedback. A search box is located in the top right corner. Below the navigation bar, there is a section for 'Maize B73 RefGen\_v1'. The main content area displays search results for the query 'adh1'. A table lists two matching regions:

Chr1	Type	Description	Position	Match Score
adh1	marker		Chr1:272752490-272782790	n/a
adh1	marker		Chr1:272905892-272936192	n/a

Fig. 1 Screenshot of the BLAST search output page.

# Multiple Sequence Alignment

Some of the key steps in building a multiple alignment include:

1. Obtaining the sequences to align by database searching
2. Running the multiple alignment program and,
3. Identifying the residues that differ or are conserved among the sequences (finding polymorphisms)

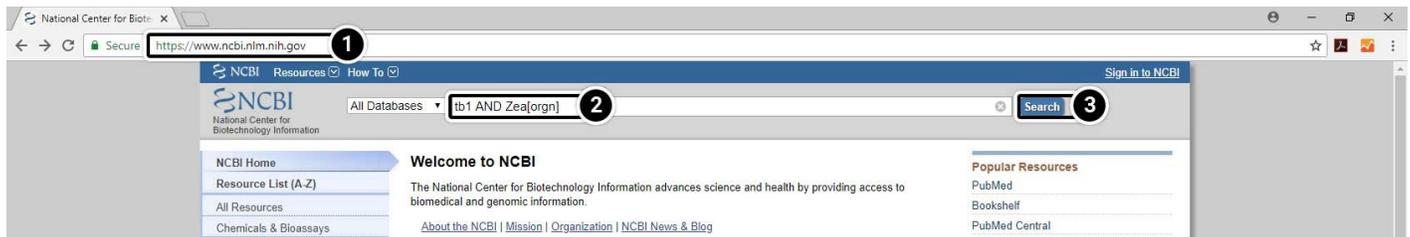
Enter the NCBI site and use the following steps to guide your activity.

# Try This: Multiple Sequence Alignment

## Search the NCBI Website for the Allelic Sequences

Find the allelic sequences for a maize gene. Here we will use teosinte branched1 (tb1) gene from maize as an example.

1. Open your web browser and go to <https://www.ncbi.nlm.nih.gov/>.
2. Enter **tb1 AND Zea[orgn]** into the search box.
3. Press **Enter** or click **Search**.



## Narrow the Search Results

### 1. Select PopSet (population data sets).

 U.S. National Library of Medicine  
National Center for Biotechnology Information [Log in](#)

**Search NCBI databases**

tb1 AND Zea[orgn] ✕ [Search](#)

Results found in 12 databases for **tb1 AND Zea[orgn]**

#### Literature

<a href="#">Bookshelf</a>	1	Books and reports
<a href="#">MeSH</a>	1	Ontology used for PubMed indexing
<a href="#">NLM Catalog</a>	0	Books, journals and more in the NLM Collections
<a href="#">PubMed</a>	57	Scientific and medical abstracts/citations
<a href="#">PubMed Central</a>	266	Full-text journal articles
<a href="#">PubMed Health</a>	0	Clinical effectiveness, disease and drug reports

#### Genes

<a href="#">EST</a>	0	Expressed sequence tag sequences
<a href="#">Gene</a>	8	Collected information about gene loci
<a href="#">GEO DataSets</a>	1	Functional genomics studies
<a href="#">GEO Profiles</a>	0	Gene expression and molecular abundance profiles
<a href="#">HomoloGene</a>	0	Homologous gene sets for selected organisms
<a href="#">PopSet</a>	18	Sequence sets from phylogenetic and population studies
<a href="#">UniGene</a>	1	Clusters of expressed transcripts

1

## Choose the Specific Search Result

2. Select the result containing 17 aligned sequences of tb1 partial cds from a population study. (UID 209362237)

Summary ▾ 200 per page ▾

Send to: ▾ Filters: [Manage Filters](#)

### Search results

Items: 18

[Zea mays subsp. mays teosinte branched 1 protein \(tb1\), gene, partial cds.](#) 2

1. population study, 17 aligned sequences  
UID: 209362237  
[Protein](#) [PubMed](#) [Taxonomy](#)

[Zea mays subsp. mays unknown gene.](#)

2. population study, 61 aligned sequences  
UID: 209362113  
[Protein](#) [PubMed](#) [Taxonomy](#)

[Zea mays subsp. mays putative zinc-finger protein gene, partial cds.](#)

3. population study, 61 aligned sequences  
UID: 209361769  
[Protein](#) [PubMed](#) [Taxonomy](#)

[Zea mays subsp. mays DWARF8 protein \(D8\), gene, partial cds.](#)

4. population study, 19 aligned sequences  
UID: 209361731  
[Protein](#) [PubMed](#) [Taxonomy](#)

[\[No title available\]](#)

5. phylogenetic study, 25 aligned sequences  
UID: 42601454  
[Protein](#) [PubMed](#) [Taxonomy](#)

---

Find related data

Database: Select ▾

---

### Search details

("Thyridaria broussonet OR tb1[All Fields]) AND

---

### Recent activity

🔍 tb1 AND Zea[orgn] (18)

## Explore the Alignment

Scroll down to see the alignment of the 17 tb1 partial cds.

1. Click the + sign until you can see the nucleotides.
2. Click the arrow to pan right in the sequences until you can see the region between 1500 and 1530.

### Alignment

Description	Marker	Seq. Size	First	Alignment	Org. Name	Last	Seq. End	Seq. Len
				1,681 1,690 1,700 1,714				
FJ201118.1	▶ ⊕	1	1,498	A	Zea mays subsp. mays	1,501	2,686	2,686
FJ201119.1	▶ ⊕	1	1,496	A	Zea mays subsp. mays	1,499	2,684	2,684
FJ201120.1	▶ ⊕	1	1,496	A	Zea mays subsp. mays	1,499	2,684	2,684
FJ201121.1	▶ ⊕	1	1,496	A	Zea mays subsp. mays	1,503	2,688	2,688
FJ201122.1	▶ ⊕	1	1,496	A	Zea mays subsp. mays	1,503	2,688	2,688
FJ201123.1	▶ ⊕	1	1,498	A	Zea mays subsp. mays	1,498	2,682	2,682
FJ201124.1	▶ ⊕	1	1,498	A	Zea mays subsp. mays	1,498	2,682	2,682
FJ201125.1	▶ ⊕	1	1,496	A	Zea mays subsp. mays	1,529	2,714	2,714
FJ201126.1	▶ ⊕	1	1,499	A	Zea mays subsp. mays	1,499	2,683	2,683
FJ201127.1	▶ ⊕	1	1,498	A	Zea mays subsp. mays	1,498	2,682	2,682
FJ201128.1	▶ ⊕	1	1,496	A	Zea mays subsp. mays	1,528	2,713	2,713
FJ201129.1	▶ ⊕	1	1,496	A	Zea mays subsp. mays	1,503	2,688	2,688
FJ201130.1	▶ ⊕	1	1,679	A	Zea mays subsp. mays	1,689	2,874	2,874
FJ201131.1	▶ ⊕	1	1,498	A	Zea mays subsp. mays	1,498	2,682	2,682
FJ201132.1	▶ ⊕	1	1,498	A	Zea mays subsp. mays	1,498	2,682	2,682
FJ201133.1	▶ ⊕	1	1,500	A	Zea mays subsp. mays	1,508	2,693	2,693
FJ201134.1	▶ ⊕	1	1,498	A	Zea mays subsp. mays	1,498	2,682	2,682



## Finding Polymorphisms

### Using Clustal Omega

To detect polymorphisms in a set of candidate genes requires a program that aligns multiple sequences. Clustal Omega is one of the commonly used programs. Clustal Omega is a hierarchical multiple alignment program that combines a robust method for multiple sequence alignment with a user-friendly interface. There are different webservers that provide access to Clustal Omega. For this lesson we will use the European Bioinformatics Institute webserver. Clustal Omega can also be downloaded to a personal computer for more routine use. The following is an example of how to use Clustal Omega to detect polymorphisms.

# Try This: Using Clustal Omega

## Search the NCBI Website

1. Go to [the NCBI website](#) and search for **tb1 AND Zea[orgn]**.
2. Click **Search**



The screenshot shows the top navigation bar of the NCBI website. On the left is the NCBI logo with the text "National Center for Biotechnology Information". To its right is a search bar with a dropdown menu set to "All Databases". The search input field contains the text "tb1 AND Zea[orgn]" and is marked with a circled "1". To the right of the search bar is a "Search" button, also marked with a circled "2". Below the search bar is a navigation menu with four items: "NCBI Home", "Resource List (A-Z)", "All Resources", and "Chemicals & Bioassays". The "NCBI Home" item is highlighted with a blue arrow. To the right of the navigation menu is a "Welcome to NCBI" section with a paragraph of text and a link to "About the NCBI | Mission | Organization | NCBI News & Blog". On the far right is a "Popular Resources" section with three items: "PubMed", "Bookshelf", and "PubMed Central".

## Explore the Search Results

### 1. Select PopSet (population data sets).

 U.S. National Library of Medicine  
National Center for Biotechnology Information Log in

---

**Search NCBI databases**

tb1 AND Zea[orgn] x Search

Results found in 12 databases for **tb1 AND Zea[orgn]**

#### Literature

<a href="#">Bookshelf</a>	1	Books and reports
<a href="#">MeSH</a>	1	Ontology used for PubMed indexing
<a href="#">NLM Catalog</a>	0	Books, journals and more in the NLM Collections
<a href="#">PubMed</a>	57	Scientific and medical abstracts/citations
<a href="#">PubMed Central</a>	266	Full-text journal articles
<a href="#">PubMed Health</a>	0	Clinical effectiveness, disease and drug reports

#### Genes

<a href="#">EST</a>	0	Expressed sequence tag sequences
<a href="#">Gene</a>	8	Collected information about gene loci
<a href="#">GEO DataSets</a>	1	Functional genomics studies
<a href="#">GEO Profiles</a>	0	Gene expression and molecular abundance profiles
<a href="#">HomoloGene</a>	0	Homologous gene sets for selected organisms
<a href="#">PopSet</a>	18	Sequence sets from phylogenetic and population studies
<a href="#">UniGene</a>	1	Clusters of expressed transcripts

## Select the Population Set

1. Click on the population set we studied earlier (UID 209362237)

Summary ▾ 200 per page ▾

**Search results**  
Items: 18

Send to: ▾ Filters: [Manage Filters](#)

**Find related data**

Database: Select ▾

Find items

**Search details**

("Thyridaria broussonet  
OR tb1[All Fields]) AND

Search

**Recent activity**

tb1 AND Zea[orgn] (18)

[Zea mays subsp. mays teosinte branched 1 protein \(tb1\), gene, partial cds.](#) **1**

1. population study, 17 aligned sequences  
UID: 209362237  
[Protein](#) [PubMed](#) [Taxonomy](#)

[Zea mays subsp. mays unknown gene.](#)

2. population study, 61 aligned sequences  
UID: 209362113  
[Protein](#) [PubMed](#) [Taxonomy](#)

[Zea mays subsp. mays putative zinc-finger protein gene, partial cds.](#)

3. population study, 61 aligned sequences  
UID: 209361769  
[Protein](#) [PubMed](#) [Taxonomy](#)

[Zea mays subsp. mays DWARF8 protein \(D8\), gene, partial cds.](#)

4. population study, 19 aligned sequences  
UID: 209361731  
[Protein](#) [PubMed](#) [Taxonomy](#)

[\[No title available\]](#)

5. phylogenetic study, 25 aligned sequences  
UID: 42601454  
[Protein](#) [PubMed](#) [Taxonomy](#)

[Add new comment](#)

## Create a FASTA File

Create a FASTA file of the 17 tb1 sequences.

1. Click the pull-down menu **Send to:** at the top right of the screen
2. In the menu that appears, select **File** for the destination
3. Select the **FASTA** format
4. Finally, click **Create File**.

The screenshot shows a web interface with a search bar at the top. Below it, a 'Send to:' dropdown menu is open, displaying options: File (selected), Clipboard, Collections, and Analysis Tool. Below the menu, it says 'Download 1 items.' and 'Format' with a dropdown menu showing 'FASTA'. A 'Create File' button is visible below the format menu. Four numbered callouts (1-4) highlight the steps: 1 points to the 'Send to:' dropdown, 2 points to the 'File' option, 3 points to the 'FASTA' format dropdown, and 4 points to the 'Create File' button. The background shows a search result for 'varf8 region in maize. Tenailon, M.I.' with a 'Data set' link and a section titled 'Other data sets from this' with a link to 'Zea mays subsp. mays unknoc'.

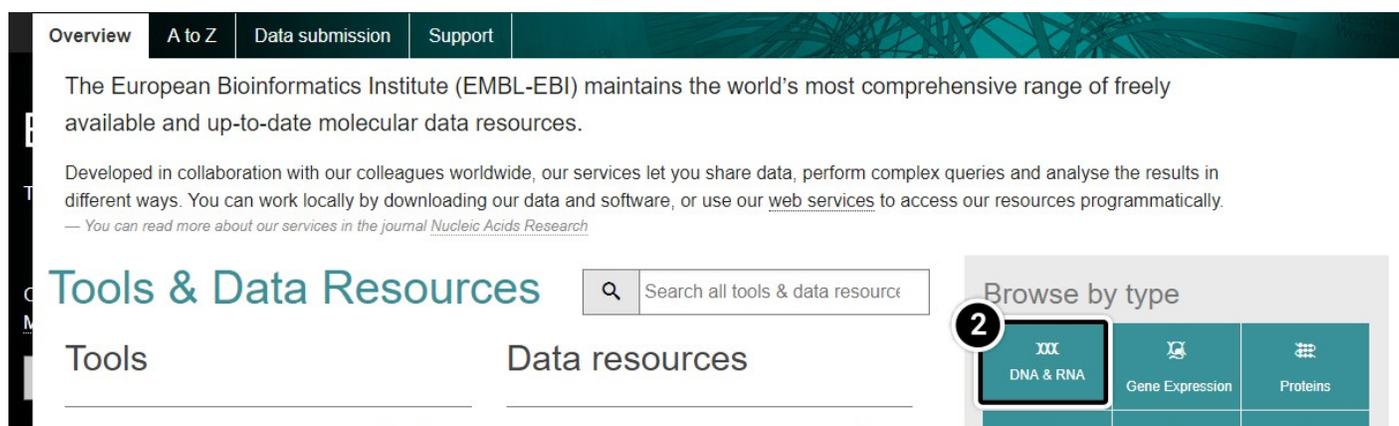
## Access Clustal Omega

Access the Clustal Omega program through [EMBL-EBI](#).

1. Click the Services link
2. Under Browse by type, click **DNA & RNA**



The screenshot shows the EMBL-EBI homepage. At the top, there is a navigation bar with links for Home, **Services** (highlighted with a circled '1'), Research, Training, and About us. The main header features the EMBL-EBI logo and the text 'The home for big data in biology'. A prominent statistic states '27 million Average requests per day to EMBL-EBI websites.' Below this, there are three main sections: 'Our unique Search service helps you explore dozens of biological data resources.' with a 'More about EBI Search' link; 'Find a tool for your data analysis.' with a 'Find a tool' button; and 'Share your scientific data with the world.' with a 'Deposit data' button. A search bar is located at the bottom of the header, containing the text 'Find a gene, protein or chemical' and a search icon. Below the search bar, there are example searches: 'blast keratin bfl1'.



The screenshot shows the 'Tools & Data Resources' page on the EMBL-EBI website. The page has a navigation bar with links for Overview, A to Z, Data submission, and Support. The main content area contains the following text: 'The European Bioinformatics Institute (EMBL-EBI) maintains the world's most comprehensive range of freely available and up-to-date molecular data resources.' and 'Developed in collaboration with our colleagues worldwide, our services let you share data, perform complex queries and analyse the results in different ways. You can work locally by downloading our data and software, or use our [web services](#) to access our resources programmatically.' Below this text, there is a quote: '— You can read more about our services in the journal [Nucleic Acids Research](#)'. The page is divided into two main sections: 'Tools' and 'Data resources'. A search bar is located at the top of the 'Tools & Data Resources' section, containing the text 'Search all tools & data resource'. On the right side, there is a 'Browse by type' section with three buttons: 'DNA & RNA' (highlighted with a circled '2'), 'Gene Expression', and 'Proteins'.

## Perform Alignment

Perform alignment of tb1 partial cds using Clustal Omega. Within the Clustal Omega window you have the option of pasting sequences, or uploading files containing your sequences in FASTA format. We will upload the FASTA file you created in Step 2. As you may notice in this window, the default is set as "PROTEIN." Since you wish to align tb1 DNA sequences, you must change this parameter. Upload your file and click **Submit**.

1. Click **Clustal Omega**
2. Select **DNA** from the dropdown
3. Click **Choose File** to browse for the file you created.
4. Click **Submit**

Overview | A to Z | Data submission | Support

DNA & RNA | [View all tools and data resources >](#)

Filter these tools & data resour

## Tools & Data Resources

### Tools

**BLAST [nucleotide]**   
Fast local similarity search tool for nucleotide sequence databases.  
**Sequence similarity search**

**Clustal Omega**   
Multiple sequence alignment of DNA or protein sequences. Clustal Omega replaces the older ClustalW alignment tools.  
**Multiple sequence alignment**

### Data resources

**DGVa**   
A repository that provides archiving, accessioning and distribution of publicly available genomic structural variants, in all species.  
**DGVa**

**EGA**   
A service for permanent archiving and sharing of all types of potentially identifiable genetic, molecular and phenotypic data resulting from biomedical research projects.  
**EGA**

**ENA** **FNA**

STEP 1 - Enter your input sequences

Enter or paste a set of

PROTEIN  
PROTEIN  
DNA  
RNA

2

Or, upload a file:  chosen

[See example inputs](#)

STEP 2 - Set your parameters

OUTPUT FORMAT

ClustalW with character counts

The default settings will fulfill the needs of most users.

(Click here, if you want to view or change the default settings.)

STEP 3 - Submit your job

Be notified by email (Tick this box if you want to be notified by email when the results are available)

4

## Explore the Output

It will take a moment before you obtain a report of your job request. You can click and save the “Your Job Output” URL to view your results for up to seven days.

1. Click the **Job ID** link
2. You can click the **Download Alignment File** but that is not necessary for this activity
3. Click **Result Summary**

## Your job is currently running... please be patient

The result of your job will appear in this browser window.

Job ID [clustalo-I20180712-160918-0487-55474579-p1m](#) **1**

### Please note the following

- You may press Shift+Refresh or Reload on your browser at any time to check if results are ready.
- You may bookmark this page to view your results later if you wish.
- Results are stored for 7 days.

## Results for job clustalo-I20180712-161111-0992-19931388-p2m

[Alignm \*\*3\*\*](#) [Result Summary](#) [Phylogenetic Tree](#) [Submission Details](#)  
[Download Alignment File \*\*2\*\*](#) [result with Jalview](#) [Send to Simple Phylogeny](#) [Send to MView](#)

CLUSTAL O(1.2.4) multiple sequence alignment

```
FJ201130.1    GGCTTGCCCCCATGTACAAC TTTATGCTGCGATAATATAGGAGTAGTTGGGCCTCTGCT    60
FJ201133.1    GGCTTGCCCCCATGTACAAC TTTATGCTGCGATAATATAGGAGTAGTTGGGCCTCTGCT    60
FJ201121.1    GGCTTGCCCCCATGTACAAC TTTATGCTGCGATAATATAGGAGTAGTTGGGCCTCTGCT    60
FJ201122.1    GGCTTGCCCCCATGTACAAC TTTATGCTGCGATAATATAGGAGTAGTTGGGCCTCTGCT    60
FJ201125.1    GGCTTGCCCCCATGTACAAC TTTATGCTGCGATAATATAGGAGTAGTTGGGCCTCTGCT    60
```

## View Result with Jalview

1. Click **View result with Jalview**
2. Once Jalview opens, click **Colour** then **Nucleotide**
3. Use the scroll bar to navigate to the alignment.
4. Scroll to align the region from 1680 to 1740.

## Results for job clustalo-I20180712-161111-0992-19931388-p2m

Alignments **Result Summary** Phylogenetic Tree Submission Details

### Input Sequences

[clustalo-I20180712-161111-0992-19931388-p2m.input](#)

### Tool Output

[clustalo-I20180712-161111-0992-19931388-p2m.output](#)

### Alignment in CLUSTAL format with base/residue numbering

[clustalo-I20180712-161111-0992-19931388-p2m.clustal\\_num](#)

### Phylogenetic Tree

[clustalo-I20180712-161111-0992-19931388-p2m.ph](#)

### Percent Identity Matrix

[clustalo-I20180712-161111-0992-19931388-p2m.pim](#)

### Jalview

1

[View result with Jalview](#)



### Simple Phylogeny

[Send to Simple Phylogeny](#)



### MView

[Send to MView](#)



http://www.ebi.ac.uk/Tools/services/rest/clustalo/result/clustalo-l20180712-161111-0992-19931388-p2m/al...

File Edit Select View Annotations Format Colour Calculate Web Service

10 50 60

FJ201130.1/1-2874 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201133.1/1-2693 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201121.1/1-2688 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201122.1/1-2688 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201125.1/1-2714 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201128.1/1-2713 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201119.1/1-2684 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201120.1/1-2684 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201118.1/1-2686 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201129.1/1-2688 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201126.1/1-2683 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201132.1/1-2682 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201124.1/1-2682 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201123.1/1-2682 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201127.1/1-2682 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201134.1/1-2682 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA  
FJ201131.1/1-2682 GGCTTGCCCCCATGTAC AGTAGTTGGGCCTCTGCTAAA

Consensus

2

- Apply Colour To All Groups
- Text Colour...
- None
- Clustal
- Blosum62
- % Identity
- Zappo
- Taylor
- Hydrophobic
- Helix Propensity
- Strand Propensity
- Turn Propensity
- Buried Index
- Nucleotide**
- Purine/Pyrimidine

http://www.ebi.ac.uk/Tools/services/rest/clustalo/result/clustalo-l20180712-161111-0992-19931388-p2m/aln-clustal\_num

File Edit S View Annotations Format Colour Calculate Web Service

1680 1690 1700 1710 1720 1730 1740

FJ201130.1/1-2682 G T T T T T C T C T C T C . . . . . T C T C T C T C T C T C T C T C T C T C T T A C A A G C C T A G A

FJ201133.1/1-2693 G T T T T T A A T C . . . . . T C T C T C T C T C T C T C T C T C T C T T A C A A G C C T A G A

FJ201121.1/1-2688 G T T T T T A A C T C T C T C T C . . . . . T C T C T C T C T C T C T C T C T C T C T T A C A A G C C T A G A

FJ201122.1/1-2688 G T T T T T A A C T C T C T C T C . . . . . T C T C T C T C T C T C T C T C T C T C T T A C A A G C C T A G A

FJ201125.1/1-2714 G T T T T T A A C T C T C T C T C T C T C T C T C T C T C T C T C T C T C T C T C T T A C A A G C C T A G A

FJ201128.1/1-2713 G T T T T T A A T C T C T C T C T C T C T C T C T C T C T C T C T C T C T C T C T T A C A A G C C T A G A

FJ201119.1/1-2684 G T T T T T A A C . . . . . T C T C T C T C T C T C T C T C T C T C T C T T A C A A G C C T A G A

FJ201120.1/1-2684 G T T T T T A A C . . . . . T C T C T C T C T C T C T C T C T C T C T C T T A C A A G C C T A G A

FJ201118.1/1-2686 G T T T T T A A C . . . . . T C T C T C T C T C T C . . . . . T C T C T C T C T T A C A A G C C T A G A

FJ201129.1/1-2688 G T T T T T A A C . . . . . T C T C T C T C T C T C T C T C T C T C T C T T A C A A G C C T A G A

FJ201126.1/1-2683 G T T T T T A A C . . . . . T C T C T C T C T C . . . . . T C T C T C T C T T A C A A G C C T A G A

FJ201132.1/1-2682 G T T T T T A A C . . . . . T C T C T C T C T C . . . . . T C T C T C T C T T A C A A G C C T A G A

FJ201124.1/1-2682 G T T T T T A A C . . . . . T C T C T C T C T C . . . . . T C T C T C T C T T A C A A G C C T A G A

FJ201123.1/1-2682 G T T T T T A A C . . . . . T C T C T C T C T C . . . . . T C T C T C T C T T A C A A G C C T A G A

FJ201127.1/1-2682 G T T T T T A A C . . . . . T C T C T C T C T C . . . . . T C T C T C T C T T A C A A G C C T A G A

FJ201134.1/1-2682 G T T T T T A A C . . . . . T C T C T C T C T C . . . . . T C T C T C T C T T A C A A G C C T A G A

FJ201131.1/1-2682 G T T T T T A A C . . . . . T C T C T C T C T C . . . . . T C T C T C T C T T A C A A G C C T A G A

Consensus G T T T T T A A C C T T A C A A G C C T A G A

Occupancy

3

Sequence 8 ID: 1120.1 nucleotide: Guanine (1409)



## Developing Marker Assays

Recall in Module 2 you learned how SSR and SNP can be analyzed by PCR and restriction enzymes. In lesson 8 of this course, you will learn additional strategies to detect DNA polymorphisms for marker development.

## Summary

Biological sequence databases serve an important role of providing access to sequence information to the research community. Searches can be restricted to a single database or expanded to include all other databases. Whole genomes can be explored to predict positions that match a specific sequence. To detect polymorphisms in a set of candidate genes a program that aligns multiple sequences is required. The detected polymorphisms can be used to develop markers to assist in selection.

## Reflection

The **Module Reflection** appears as the last "task" in each module. The purpose of the Reflection is to enhance your learning and information retention. The questions are designed to help you reflect on the module and obtain instructor feedback on your learning. Submit your answers to the following questions to your instructor.

1. In your own words, write a short summary (< 150 words) for this module.
2. What is the most valuable concept that you learned from the module? Why is this concept valuable to you?
3. What concepts in the module are still unclear/the least clear to you?

# Acknowledgements

This module was developed as part of the Bill & Melinda Gates Foundation Contract No. 24576 for Plant Breeding E-Learning in Africa.

**Molecular Plant Breeding Introduction to Bioinformatics Author:** Ursula Frei, Walter Suza, Thomas Lübberstedt, and Madan Bhattacharyya (ISU)

**Multimedia Developers:** Gretchen Anderson, Todd Hartnell, and Andy Rohrback (ISU)

**How to cite this module:** Frei, U., W. Suza, T. Lübberstedt, and M. Bhattacharyya. 2016. Introduction to Bioinformatics. *In* Molecular Plant Breeding, interactive e-learning courseware. Plant Breeding E-Learning in Africa. Retrieved from <https://pbea.agron.iastate.edu>.

---

**Source URL:** <https://pbea.agron.iastate.edu/course-materials/molecular-plant-breeding/introduction-bioinformatics-0?cover=1>