

Welcome to the Gramene BLAST Tutorial

- This tutorial will show you how to conduct a BLAST search.
- With BLAST you may:
 - Search for sequence similarity matches in the Gramene database - ideal for cross-species comparisons.
 - Select the best target database for your search.
 - Choose the best algorithm for your search.
 - Fine-tune search parameters.

Tutorial Tips



If you are viewing this tutorial with Adobe Acrobat Reader, click the "bookmarks" on the left hand side of the Reader for easier navigation.

Note! Although we continually work to make Gramene compatible with all browsers, there are problems with some browser versions. If you're having difficulty viewing Gramene, try using a different browser. Please report any problems with browsers through Gramene Feedback.

Gramene Home Page

www.gramene.org

The screenshot shows the Gramene Home Page with a green header and navigation menu. A search bar is located in the top right. The main content area is divided into several sections: 'Quick Start', 'Visit with us at', 'Gramene Tip', and a list of resources. A blue callout box with a pointer highlights the 'Sequences-BLAST' link in the left-hand navigation menu.

GRAMENE Home Find anything [input] Search

Search Genomes Species Download Resources About Help Feedback

Quick Start

Visit with us at

- July 7-11. ASPB demo
- Aug 29-31 Environmental ontology workshop
- Sept 21-22 Agricultural Ontology Service (AOS) Workshop
- Nov 4-8. Crop Science poster

Gramene Tip:

Search using synonyms. For example, *Pennisetum glaucum*, *Pennisetum americanum*, and *Pearl millet* are synonyms, but produce different search results. See [Ontologies](#) to find synonyms.

- [Browse All Tips](#)

Sequences-BLAST

Click here to open BLAST Home Page

GENETIC DIVERSITY Search for SNP and SSR allelic variation on loci of rice, maize, and wheat germplasms.

BIOCHEMICAL PATHWAYS Search for ALL the rice pathways on starch biosynthesis or get an overview of the metabolic network. Compare rice and wheat biosynthesis pathway datasets.

LITERATURE: Search for literature related to Gramene data for *Oryza*, *Zea*, *Triticum*, *Hordeum*, *Avena*, *Setaria*, *Pennisetum*, and *Brachypodium distachyon*.

SPECIES PAGES Search for species pages for *Oryza*, *Zea*, *Triticum*, *Hordeum*, *Avena*, *Setaria*, *Pennisetum*, and *Brachypodium distachyon*.

SUBMISSION: Submit a question.

Note! Although we continually work to make Gramene compatible with all browsers, if you're having difficulty viewing Gramene, try using a different browser. Please report any problems with browsers through Gramene Feedback.

SPECIES

BLAST Home Page

Sequence Information:
Enter a sequence that you have and are trying to locate, or are trying to find a similar sequence for

Search Information:
where would you like to search for similar sequences, using which tool?

TBLASTX compares a nucleotide query sequence translated in all reading frames against a protein sequence database

The screenshot shows the BLAST Home Page interface. At the top, there are navigation buttons: 'new', 'SETUP' (highlighted in red), 'CONFIG', 'RESULTS', and 'DISPLAY'. On the right side, there are buttons for 'refresh', 'FAQ', 'Online Help', and 'Tutorial'. Below these is a 'Summary' section with links for 'setup', 'configure', and 'results', each with a status indicator 'Not yet initialised'.

The main content area is divided into three sections:

- Enter the Query Sequence:** This section has three options: 'Either Paste sequences (max 10) in FASTA or plain text:' with a large text input field; 'Or Upload a file containing one or more FASTA sequences' with a 'Browse...' button; and 'Or Enter an existing ticket ID:' with a 'Retrieve' button. Below these are radio buttons for 'dna queries' (selected) and 'peptide queries'. A callout bubble points to these radio buttons, containing the text: 'DNA codes contain: ACTG' and 'Peptide codes contain: GALMFWKQESPVICYHRNDT'.
- Select the databases to search against:** This section includes a 'Select species:' dropdown menu with 'Millet', 'Poaceae', and 'Rice' (selected). Below this are radio buttons for 'dna database' (selected) and 'peptide database'. To the right are two dropdown menus: 'Genomic sequence' and 'Peptides (Fgenesh gene models)'.
- Select the Search Tool:** This section has a dropdown menu with 'BLASTN' (selected) and 'TBLASTX'. A 'configure' button and a 'RUN' button are also present. A callout bubble points to the 'BLASTN' option, containing the text: 'BLASTN compares a nucleotide (dna) query sequence against a nucleotide sequence database'.

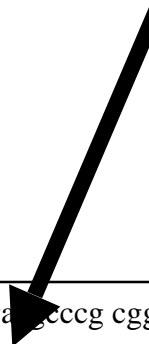
At the bottom, there is a 'Search sensitivity:' dropdown menu with 'Normal' selected and a note: 'Optimise search parameters to find the following alignments'.

Example

BLAST - The Basic Local Alignment Search Tool – is used when searching for related (similar) sequences.

Problem: Finding a Nucleotide Match in the Rice Genome

You have an RFLP (restriction fragment length polymorphism) genetic marker with a **known hybridization sequence**, and want to know where it is located on the rice genome



```
1  ggcatccatg ggcCCAagg cggagaagaa gccggcggcg aaga gcccg cggaggagga
61  gcccgcggcg gagaaggccg agaaggcctg gcggggaaga agccaaggc ggagaagcgt
121 ctccccgccg gcaaggccga gaagagcagc ggcgagggga agaaggcggg gcggaagaag
181 gcgaagaaga gcgtcgagac ctacaagatc tacatctca aggtgctcaa gcaggtccac
241 cccgacatcg gcatctctc caaggccatg tcatcatga actcctcat caacgacatc
301 ttcgagaagc tcgccggga gtccccaag ctgcgcgct acaacaagaa gccaccatc
361 aactnacggg agatccagac ctncgtccgc cttgtc
```

Step 1: Enter the Sequence

The screenshot shows a web interface with a navigation bar at the top containing buttons for 'new', 'SETUP', 'CONFIG', 'RESULTS', and 'DISPLAY'. On the right side, there are buttons for 'refresh', 'FAQ', 'Online Help', and 'Tutorial'. The main content area is titled 'Enter the Query Sequence' and contains three options: 'Either Paste sequences (max 10) in FASTA or plain text:', 'Or Upload a file containing one or more FASTA sequences', and 'Or Enter an existing ticket ID:'. The first option is selected, and a text area contains a FASTA-formatted DNA sequence. Below the text area are radio buttons for 'dna queries' (selected) and 'peptide queries'. Callouts provide instructions: '1. Paste in the sequence' points to the text area; '1a. Select "dna queries" because this is a nucleotide sequence' points to the selected radio button; 'Alternatively, save a sequence to a file and use this box to upload it.' points to the 'Browse...' button; 'Browse FAQ' points to the 'FAQ' button; 'Click for Help' points to the 'Online Help' button; and a larger callout states 'You can search with up to 10 sequences at a time. Simply format them using FASTA format.'

1. Paste in the **sequence**

1a. Select "dna queries" because this is a nucleotide sequence

Alternatively, save a sequence to a file and use this box to upload it.

Browse FAQ

Click for Help

You can search with up to 10 sequences at a time. Simply format them using FASTA format.

Step 2: Choose Target Database

2b. Select your database. For this example, choose “dna database”

2a. Select your target organism(s), in this case “Rice”

The screenshot shows the BLAST search interface. At the top, it says "the databases to search against". Below this, there is a "Select species:" dropdown menu with "Rice" selected. To the left of this menu, there are radio buttons for "dna database" (selected) and "peptide database". To the right of the "dna database" radio button, there is a dropdown menu for "Genomic sequence" and "Peptides (Fgenesh gene models)". Below this, there is a "Select the Search Tool" section with "BLASTN" selected. To the right of this section, there is a "configure" button and a "RUN" button. Below the "RUN" button, there is a "Search sensitivity:" dropdown menu with "Near-exact" selected. At the bottom, there is a "Optimize parameters to find the following alignments" section.

2c. Select a specific database. In this case, “Genomic Sequence”

2e. Select your search sensitivity

2f. Click RUN

Search tool options will change according to species and database.

2d. Select your search tool, “BLASTN” for this example

Used To examine or alter the optimized default parameters (see slide 16)

Use BLASTN for simple nucleotide against nucleotide searches

Use BLASTP for simple protein versus protein searches

Use TBLASTN for a protein query versus a DNA library search

Use BLASTX for DNA query versus protein library

Use TBLASTX for a translated DNA query against a translated DNA database

BLAST Queries

Ticket ID

Use Feedback to make enquiries to Gramene Staff

Gramene BlastSearch (BlastView) - SmartFox Internet Browser

File Edit View Go Bookmarks Tools Help

http://www.gramene.org/Multi/blastView/BLA_laO45mV6D

SmartFox Help SmartFox Support Plug-in FAQ

GRAMENE Multi Find anything Search

Search Genomes Download Resources About Help Feedback

new SETUP CONFIG RESULTS DISPLAY refresh Online Help FAQ Tutorial

Displaying unnamed sequence alignments vs Rice LATESTGP database

Showing top 100 alignments of 167, sorted by Raw Score refresh

Alignment Locations vs. Karyotype (click arrow to hide)

Summary

- setup
- Rice
- Genomic sequence

Click Arrows to hide or reveal results sections.

To sort results select an option and refresh

Results are stored on Gramene's server for one week, so that they can be accessed later with the **Ticket ID** or a bookmark to the results page.

Step 3: Results - Genomic Context

Click "CONFIG" if you want to change parameters & try again (slides 16 - 17).

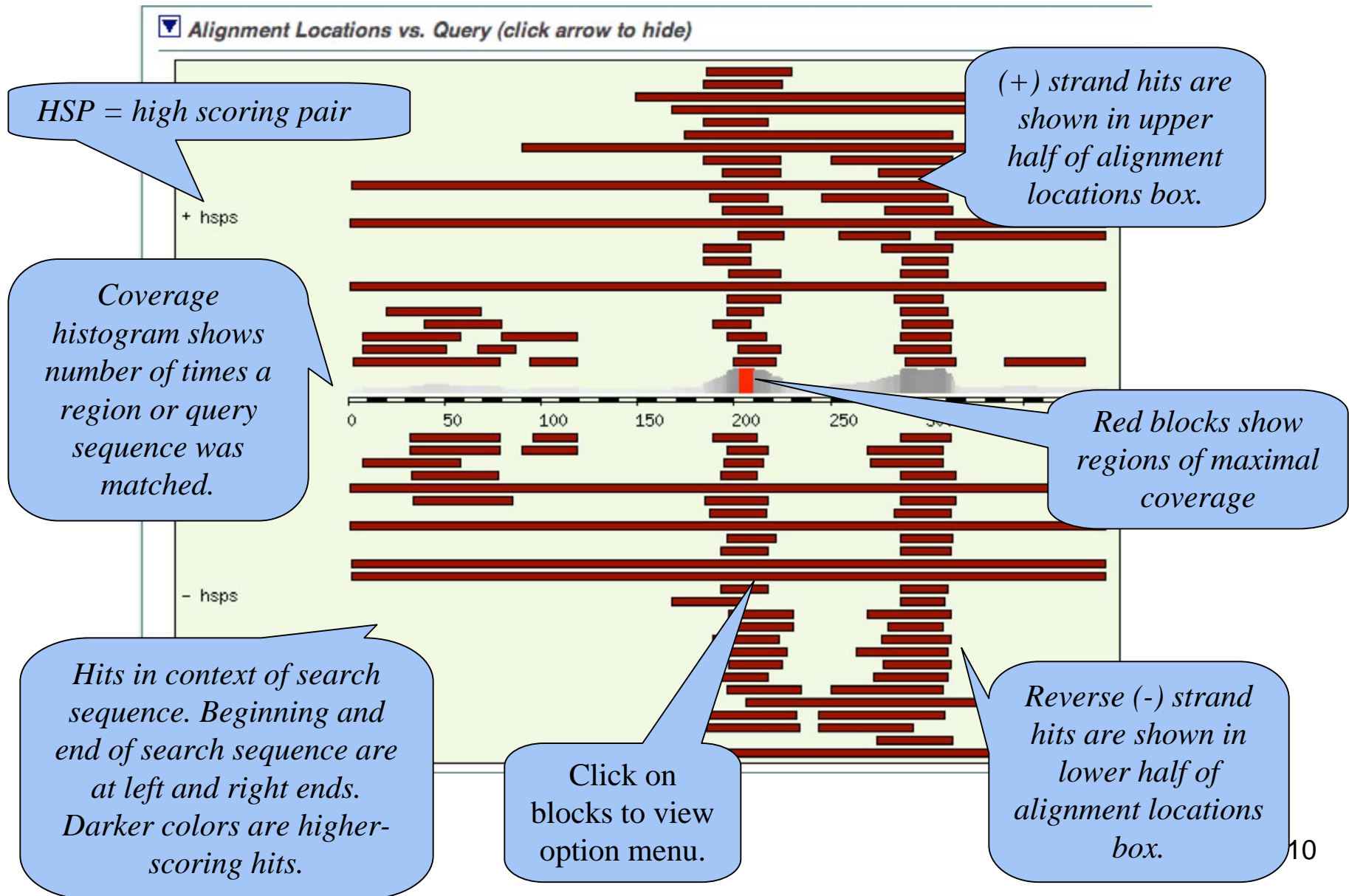
The screenshot displays a web-based interface for sequence alignment. At the top, there are navigation buttons: "new", "SETUP", "CONFIG", and "RESULTS". Below this, a header indicates "Displaying unnamed sequence alignments vs Rice LATESTGP database". A status bar shows "Showing top 100 alignments of 167, sorted by Raw Score" with a "refresh" button. A section titled "Alignment Locations vs. Karyotype (click arrow to...)" shows a karyotype with chromosomes 1 through 7. Red triangles of varying sizes indicate alignment locations. A callout box on the left shows a detailed view of an alignment for "Alignment vs AP002522", listing metrics like Raw Score, PercentID, Length, P-value, and E-value. A "Summary" panel on the right lists parameters under "setup", "configure", "results", and "display".

Identifies number of results

Click on pointers to view option menu.

Hits are in genomic context. Darker triangles indicate higher scoring hits. The best hit is in a box.

Step 3 Results - Search Sequence



Step 3: Results – Alignment Summary

▼ **Alignment Summary (click arrow to hide)**

Select rows to include in table, and type of sort (Use the 'ctrl' key to select multiples)

Query: Name, Start, End, Ori

Subject: _off_ Name, Start, End

Sort By: >Score, <E-val, >E-val, <P-val

5. Click the [C] for “contig” view of the genome (see Genome Browser Tutorial)

4. Click the [A] link of the best (first) hit to view the alignment. (Slide 12)

Select from menus to configure results

Links	Query Start	Query End	Query Ori	Chromosome Name	Start	End	Clone	Score	E-val	%ID	Length
[A] [S] [G] [C]	1	392	-	Chr:1	2817113	2817505	+	379	1.1e-217	98.98	393
[A] [S] [G] [C]	1	392	-	Chr:1	2834924	2835216	+	343	3.2e-196	96.69	393
[A] [S] [G] [C]	1	392	+	Chr:1	2671274	2671666	+	339	4.7e-193	96.44	393
[A] [S] [G] [C]	2	392	-	Chr:1	2854476	2854867	+	334	9.0e-190	96.17	392
[A] [S] [G] [C]	2	392	-	Chr:1	2841532	2841928	+	331	9.0e-190	96.17	392
[A] [S] [G] [C]	392	392	+	Chr:1	2683597	2683989	+	331	6.8e-188	95.93	393
[A] [S] [G] [C]	392	392	+	Chr:1	2868984	2869381	+	300	6.8e-169	95.70	393
[A] [S] [G] [C]	90	90	+	Chr:5	28350554	28350554	+	216	3.8e-155	100.00	77

Shortcut: You can also click on the best hit in the genome graphic (from Slide 9) to open a menu for that alignment.

▼ **Alignment Locations vs. Karyotype (click arrow to hide)**

Alignment vs AP003045

Alignment...

Query Sequence...

Genomic Sequence...

ContigView...

Raw Score: 379

PercentID: 98.98

Length: 393

P-value: 1.1e-217

E-value: 1.1e-217

Alignment...

Query Sequence...

Genomic Sequence...

ContigView...

Raw Score: 343

PercentID: 96.69

Length: 393

P-value: 3.2e-196

E-value: 3.2e-196

Alignment...

Query Sequence...

Click these links to view:

- [A] alignment
- [S] query sequence
- [G] target (genome) sequence
- [C] on genome (“ContigView”)

▼ **Alignment Summary (click arrow to hide)**

Select rows to include in table, and type of sort (Use the 'ctrl' key to select multiples)

Query: Name, Start, End, Ori

Subject: _off_ Name, Start, End

Chromosome: Name, Start

Clone: Name, Start

Stats: Score, E-val

Sort By: >Clone, <Score, >Score

Links	Query Start	Chromosome Name	Start	Score	E-val
[A] [S] [G] [C]	1	Chr:1	2817113	379	1.1e-217
[A] [S] [G] [C]	1	Chr:1	2834924	343	3.2e-196
[A] [S] [G] [C]	1	Chr:1	2671274	339	4.6e-193
[A] [S] [G] [C]	2	Chr:1	2841532	334	8.9e-190
[A] [S] [G] [C]	2	Chr:1	2854476	334	8.9e-190
[A] [S] [G] [C]	1	Chr:1	2683597	331	6.8e-188
[A] [S] [G] [C]	2	Chr:1	2868984	300	6.8e-169
[A] [S] [G] [C]	90	Chr:5	28350554	216	3.8e-155

6/15/07

Step 4: Review Alignment

Query location : unnamed 1 to 392 (-)
Database location : AP002522 124988 to 125380 (+)
Genomic location : 1 2817113 to 2817505 (+)

Alignment score : 379
E-value : 1.1e-217
Alignment length : 393
Percentage identity: 98.98

```
Query: 392 AGGCGGACGNAGGTCTGGATCTCCCGTNAAGTTGATGGTGGGCTTCTTGTGTAG
      |||
Sbjct: 124988 AGGCGGACGGAGGTCTGGATCTCCCGTGAAGTTGATGGTGGGCTTCTTGTGTAGCCGCGC 125047

Query: 332 AGCTTGGCGGACTCCCGCGGAGCTTCTCGAAGATGTCGTTGATGAAGGAGTTCATGATG 273
      |||
Sbjct: 125048 AGCTTGGCGGACTCCCGCGGAGCTTCTCGAAGATGTCGTTGATGAAGGAGTTCATGATG 125107

Query: 212 GACATGGCCCTGGAGGAGATGCCGATGTCGGGGTGGACCTGCTTGAGCACCTTGAAGATG 213
      |||
Sbjct: 125108 GACATGGCCCTGGAGGAGATGCCGATGTCGGGGTGGACCTGCTTGAGCACCTTGAAGATG 125167

Query: 152 TAGATCTTGTAGGTCTCGACGCTCTTCTTCCCGG
      |||
Sbjct: 125228 TAGATCTTGTAGGTCTCGACGCTCTTCTTCCCGGAGACGCTTCTCCGCCTTGGGCTTCTTCCC 125287

Query: 92 GCCAGG-CCTTCTCGGCCTTCTCCGCCGGGGCTCCTCCTCCGGGGCTTCTTCGCCGCC 34
      |||
Sbjct: 125288 GCCAGGCGCTTCTCGGCCTTCTCCGCCGGGGCTCCTCCTCCGGGGCTTCTTCGCCGCC 125347

Query: 33 GGCTTCTTCTCCGCCTTGGGCGCCATGGATGCC 1
      |||
Sbjct: 125348 GGCTTCTTCTCCGCCTTGGGCGCCATGGATGCC 125380
```

The sequence you searched.

This subject looks like a pretty good match for the query.

The result you are viewing.

Note that the long sequences continue to the next line, like reading a sentence

Note the lines identifying matches, and the absence of a line if it doesn't match.

Show or Hide tables by clicking on the [+] or [-] signs.

Step 5: Review Genome with Genome Browser

Position of hit on chromosome

Chromosome 1
Chr. 1

Overview

maize synteny
8>

red rice synteny
3>

Chr. 1 band
AP002540 AP002522 AP003045

DNA(clone)
AP002540 AP002522 AP003045

Markers
RM1869 S14073 RM10137 S20229S RM3740 RM10143
R1841 S14073 RM10138 S20229S RM3740
R1841 S14073 S20229S C1127
R106 S14073 S20229S C1127
S14073 RM10139
S14073

GeneModel TIGR
LOC_Os01g05710 LOC_Os01g05780 NP_909283.1 Q5ZD75 Q5ZD71 NP_909303.1 Q5ZCE
LOC_Os01g05680 Q5ZEK1 NP_909282.1 NP_909294.1 Q5ZD69 LOC_Os01g061
Q5ZEJ0 Q5ZEI3 Q5ZEJ6 NP_909284.1 NP_909295.1 Q5ZD66 LOC_Os01g06
LOC_Os01g05720 LOC_Os01g05820 NP_909291.1 Q5ZD78 Q5ZCE4
LOC_Os01g05760 NP_909285.1 NP_909296.1 NP_909302.1 LOC
LOC_Os01g05770 NP_909287.1 NP_909297.1 LOC_Os01g06080
NP_909288.1 NP_909304.1
LOC_Os01g05910 NP_909289.1 Q5ZCE3
Q5ZD68

AQCF002-SPKNB
COW1-SPKNB
AQCW10.H2O2CN

For more information see Genome Browser Tutorial and Contig View help.

Step 5: Review Best Hit on Genome

The screenshot shows a genome browser interface with several callouts:

- Chromosome 1**: A callout points to the chromosome selection menu, stating "Note that the overview has been hidden (from previous slide)."
- Overview**: A callout points to the "Overview" button, stating "Use menu to select features displayed".
- Detailed view**: A callout points to the "Detailed view" button.
- Features menu**: A callout points to the "Features" dropdown menu, stating "Use menu to select features displayed".
- Zoom**: A callout points to the zoom controls, stating "Position of hit shown at high magnification."
- GeneModel TIGR**: A callout points to the gene model tracks, stating "Collapse or Expand tables by clicking on the [+] or [-] signs."
- RFLP**: A callout points to a red bar representing an RFLP, stating "This is the most probable position of RFLP. You should correlate with genetic position to be sure."

For more information see [Genome Browser Tutorial](#) and [Contig View help](#).

Step 6: Adjusting Settings

Select the databases to search against

Select species:
Use 'ctrl' key to select multiple species

dna database
 peptide database

Select the Search Tool

BLASTN
 TBLASTX

Optimise search parameters to find the following alignments

During the search (Slide 7)
Instead of hitting RUN,
press **configure**.

Step 6: Adjust Settings - Configure

The screenshot shows the BLAST configuration page with the following elements:

- Navigation tabs: new, SETUP, CONFIG (highlighted), RESULTS, DISPLAY
- Buttons: refresh, Online Help, FAQ, Tutorial
- Configuration for BLASTN section with various parameters and their values:

Parameter	Value
Maximum E-value for reported alignments	10
number of database hits to report	10
used to filter query sequence	dust
Sort option for database hits	-sort_by
Statistics option for calculation of alignment score	-statistics
Word size for seeding alignments	15
Step-size for sliding-window used to seed alignments	-wink
One-hit seeding. (One-hit)	-one-hit
Alignment extension cutoff	-X
Additional Other options (not validated)	

See *help* for “*config*” assistance

6b. Click “**RUN**”.

Identifies version of BLAST selected

Each option has listed its explanation

Tune search by adjusting desired BLASTN parameters. Parameter set will depend on the version of BLAST selected in previous step.

6a. Change E-value cutoff to 0.001 to increase stringency of search.

Step 7: Refined Results

Displaying unnamed sequence alignments vs Rice LATESTGP database
Showing top 100 alignments of 167, sorted by Raw Score

Alignment Locations vs. Karyotype (click arrow to hide)

Alignment Locations vs. Karyotype

Notice there are fewer low-scoring hits after doing a config.

The current search parameters are shown in the right panel.

- Rice
- Genomic sequence
- BLASTN
- Custom sensitivity

configure

- -E: 0.0001
- -B: 100
- -filter: dust
- -W: 15
- -M: 1
- -N: -1
- -Q: 3
- -R: 3

results

display

CHLOROPLAST MITOC

new

SETUP

CONFIG

RESULTS

DISPLAY

refresh

Online Help

FAQ

Tutorial

HELP

Help with help!

- General
- Full text search

Entry Points

- Ensembl HelpView
- Gramene BLASTView
- Navigation bar Search / "Search the Gramene Databa...

Data Displays

- Ensembl AlignView
- Ensembl AnchorView
- Ensembl ChromoView
- Ensembl ContigView
- Ensembl CytoView
- Ensembl DomainView
- Ensembl ExonView

Gramene BLASTView

Gramene [BlastView](#) provides access to BLAST sequence similarity search algorithms via a web interface. It allows for simultaneous searches with multiple query sequences against multiple target species.

See the [FAQ](#).

Use the interface by navigating between these four pages:

- '[SETUP](#)'
- '[CONFIG](#)'
- '[RESULTS](#)'
- '[DISPLAY](#)'

You can navigate both forward and backwards between these pages. A 'Summary' column on the right shows the states of the current query.

BLAST (Basic Local Alignment Search Tool)

We currently support [WU-BLAST 2.0](#).

SETUP

The 'SETUP' page allows specification of all basic and required parameters such as the query sequence, the similarity search library, the search algorithm, as well as predefined optimised parameter sets.

1. **Enter the Query Sequence.** Paste in a sequence in **plain text** or [FASTA sequence database](#) format, or upload a sequence file. To submit several sequences at once, upload a file containing the sequences in FASTA format. There is a limit of **10 sequences per query**.

FAQs

Gramene FAQ Database

FAQ Quick Search: Search

Search Genomes Species Download Resources About Help Feedback

Gramene FAQ Database

FAQ Quick Search: Search

- ▶ BLAST
- ▶ Data Curation
- ▶ Diversity
- ▶ Genes
- ▶ Genomes
- ▶ Gramene Program
- ▶ GrameneMart
- ▶ Literature
- ▶ Maps and CMap
- ▶ Markers
- ▶ Ontologies

▼ BLAST

- ▶ 1. Is there a tutorial or help section on how to use BLAST?
- ▼ 2. What are some examples of scientific queries that can be answered using BLAST?

 You can search for the pairwise sequence match to your query sequence. The pairwise match can lead to a region of the rice genome, or help you find a match to an EST or a contig of ESTs, CDSs, or a protein or rice gene homolog.
- ▶ 3. Which BLAST programs are available?
- ▶ 4. Against which sequence databases can I do a BLAST query?
- ▶ 5. Is my BLAST query ...
- ▶ 6. I have submitted a BLAST job, why does ...
- ▶ 7. Can I search with multiple query sequences?

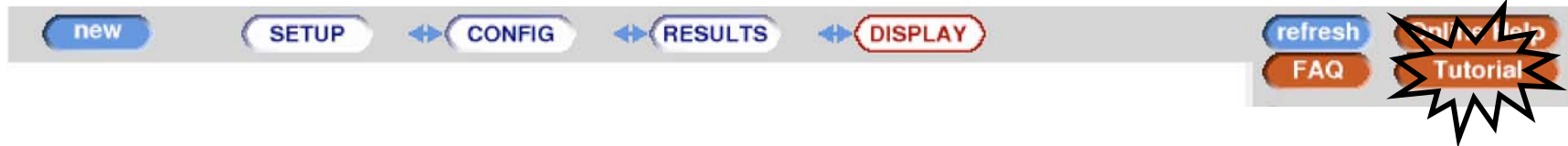
Feedback: Submit a question to Gramene

Search FAQ

Select a module to view it's FAQs

Click arrows to expand or collapse a question or answer

Tutorial



[BLAST Home](#) | [Help Files](#) | [FAQs](#)

BLAST Tutorial

Select your preferred format:

- [PowerPoint](#)
- [Adobe PDF](#)

Tutorials are available in PDF or Powerpoint

Release Notes

Navigate to other Gramene Tutorials.

Gramene Tutorials: [Navigation](#) | [Overview](#) | [Genomes](#) | [GrameneMart](#) | **BLAST** | [Maps](#) | [Markers](#) | [Proteins](#) | [Ontologies](#) | [Genes](#) | [QTL](#) | [Diversity](#) | [Pathways](#) | [Literature](#)

Free associated software:



Download free software for viewing tutorials.

More BLAST Information

Web sites that provide more BLAST information and advice on setting parameters include:

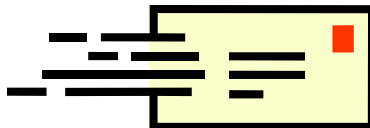
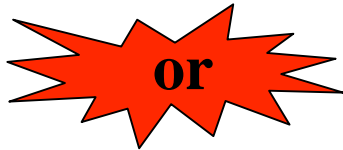
- the WU-BLAST web site <http://blast.wustl.edu/>
- EBI 2can introductions for protein or nucleotide BLAST, <http://www.ebi.ac.uk/2can/>
- NCBI-BLAST (*a good tutorial although focused on a slightly different implementation.*)
<http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/information3.html>

A book on the BLAST family of sequence similarity search algorithms (Joseph Bedell, Ian Korf and Mark Yandell, **BLAST**, O'Reilly & Associates, 2003, ISBN: 0-596-00299-8) provides profound theoretical background, as well as a protocol section covering common practical search problems. The pre-defined optimized parameter sets are based on recommendations in this book.
(<http://www.oreilly.com/catalog/blast/>)

Contact Gramene



Use the feedback button, located at the top of every page, to provide **feedback** or to **ask questions** about Gramene or your search needs.



Email the Gramene list at gramene@gramene.org