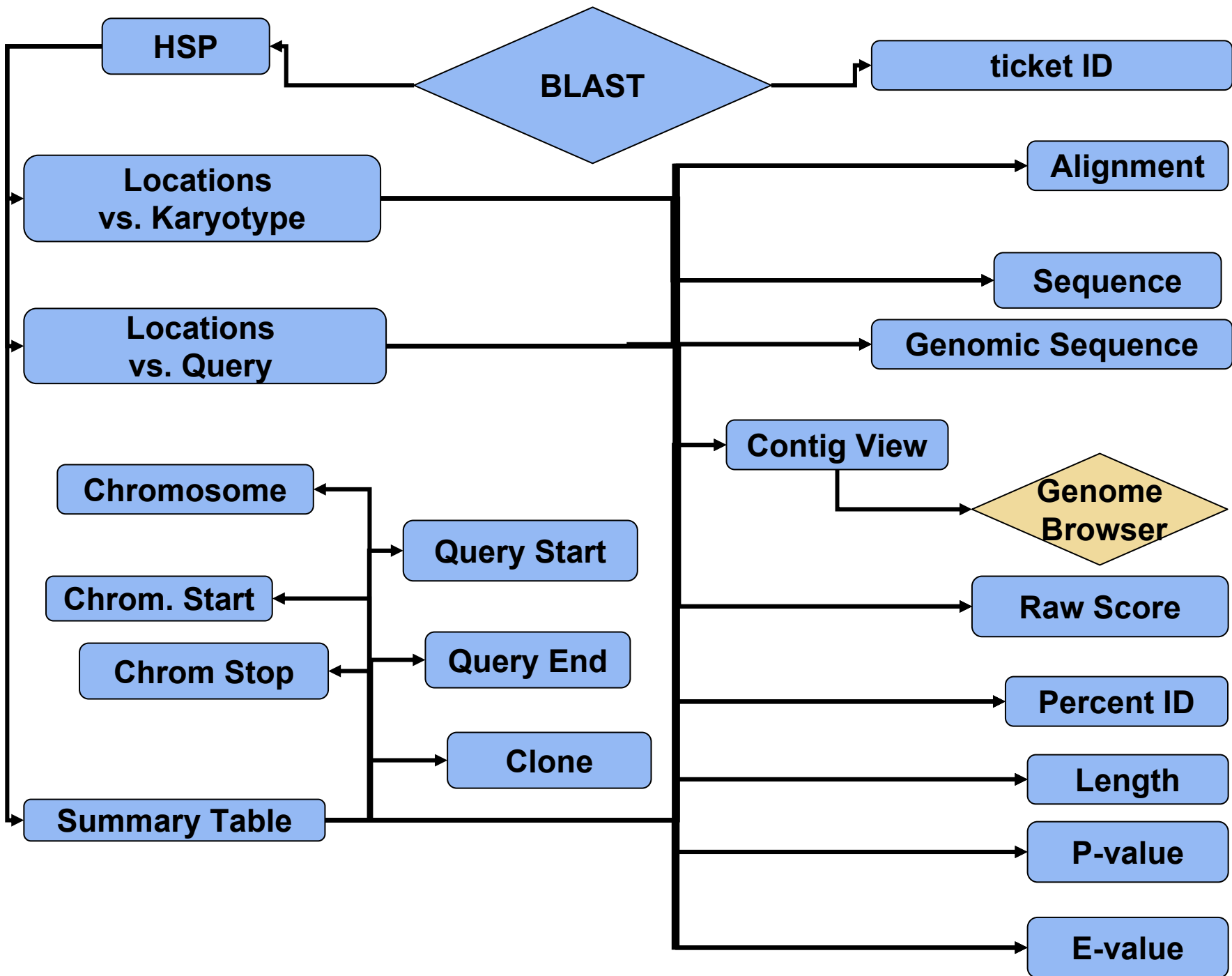


Welcome to the Gramene BLAST Tutorial

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



Tutorial Help



The hand icon indicates a link that allows you to go to the same page in your web browser.



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

Action Options are noted in this font.

Notes or comments use this style font.



Gramene Home Page

GRAMENE

A Resource for Comparative Grass Genomics

v20 (Dec 2005)

[Search](#) | [Genomes](#) | [Download](#) | [Resources](#) | [About](#) | [Help](#) | [Feedback](#)

Quick Search

Find anything

Did you know...?

- Gramene now has [tutorials](#) for every module.

Gramene Tips:

You can view Gramene's archive of all its [News](#) items.

[All Tips](#)

- Genomes-Ensembl
- Maps-CMap
- Markers
- QTL
- Genes
- Proteins
- Ontologies
- Literature
- Sequences-BLAST**
- All-GrameneMart

Quick Start

Genomes for [Rice](#), [Maize](#) & [Arabidopsis](#); Look for search with [GrameneMart](#); Search for sequence [Gene Ontology](#).

[ProSite](#) or Browse by Gene Ontology using [GO](#)

al maps for [Rice](#), [Maize](#), [wheat](#), [Barley](#), [Oats](#), [Sorghum](#), and other grasses, or use [Comparative Map Viewer \(CMap\)](#) to compare maps of different types and

MOLECULAR MARKERS (SSRIT); or search by [Others](#).

TRAITS: Search the [Genes](#) or [QTL](#) database for important phenotype-related loci such as [Rice Genes](#), [Rice QTL](#), [Maize QTL](#). Don't forget to explore traits in [Ontologies](#).

LITERATURE: Search the literature for your friends and topics of interest.

SUBMISSION: Submit a [Rice Gene](#) or [Ontology Term](#) to Gramene.

Featured News

- Breaking news on genome research
- [Rice News Worldwide](#) from IRRI
- [Gramene News Archive](#)

Visit us at

- [Plant and Animal Genom XIV Conference](#), January 14-18, 2006, San Diego, CA, USA
- [Rice Technical Working Group Meeting](#), Feb. 26-Mar. 1, 2006, The Woodlands, Texas, USA



Click here to open BLAST Home Page





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?

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

DNA codes contain: ACTG
Peptide codes contain: GALMFWKQESPVICYHRNDT

compares a nucleotide (dna) query sequence against a nucleotide 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 offers 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 options are radio buttons for 'dna queries' (selected) and 'peptide queries'. A blue callout bubble points to these options, containing the text: 'DNA codes contain: ACTG' and 'Peptide codes contain: GALMFWKQESPVICYHRNDT'.
- Select the databases to search against:** This section has a 'Select species:' dropdown menu with 'Millet', 'Poaceae', and 'Rice' (selected). Below it 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'. To the right is a 'configure' button and a 'RUN' button.

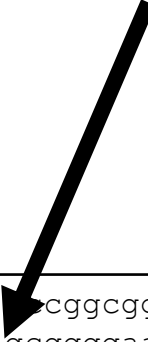
At the bottom, there is a 'Search sensitivity:' section with a dropdown menu set to 'None' and a note: 'miser search parameters to find the following alignments'.

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   ggcattccatg gcgcccaagg cggagaagaa cggcggcg aagaagcccg cggaggagga
61  gcccgcgggc gagaaggccg agaaggcctg gcggggaaga agccaaggc ggagaagcgt
121 ctccccgccg gcaaggccga gaagagcagc ggcgagggga agaaggcggg gcggaagaag
181 gcgaagaaga gcgtcgagac ctacaagatc tacatcttca aggtgctcaa gcaggteccac
241 cccgacatcg gcatctcctc caaggccatg tccatcatga actccttcat caacgacatc
301 ttcgagaagc tcgccgggga gtccgccaag ctcgcgcgct acaacaagaa gccaccatc
361 aactnacggg agatccagac ctncgtccgc cttgtc
```



Step 1: Enter the Sequence

1. Paste in the **sequence**. You must **remove the sequence numbers** first!

new SETUP CONFIG DISPLAY refresh Online Help
FAQ Tutorial

Enter the Query Sequence

Either Paste sequences (max 10) in FASTA or plain text:

```
ggcatccatg gcgcccagg cggagaagaa gccggcggcg aaga  
gcccggcg gagaaggccg agaaggctg gccgggaaga agccc  
ctcccgcg gcaaggccga gaagagcagc gccgagggga agaaggcgg  
gcgaagaaga gcgtcgagac ctacaagatc tacatctca aggtgctca
```

Or Upload a file containing one or more FASTA sequences

Or Enter an existing ticket ID:

dna queries
 peptide queries

results
Not yet initialised

Not yet initialised

Browse FAQ

Click for Help

Browse...

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

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

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 databases to search against

Select species: Poaceae
Rice
Rice_alta

Use 'ctrl' key to select multiple species

dna database
 peptide database

Genomic sequence
Peptides (Fgenesh gene models)

Select the Search Tool

BLASTN
TBLASTX

configure ▶ RUN ▶

Search sensitivity: Near-exact matches

Opt... with parameters to find the following alignments

2c. Select a specific database. In this case, **"Genomic Sequence"**

2e. Select your search sensitivity

2f. Click **RUN**

The available set of databases and search tools will change according to species and type of database, and change/grow with each release

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 about a search

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

Summary

▶ setup

- Rice
- Genomic sequence

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)

E-value
Alignment Length
% Identity
P-value
Raw Score

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

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)

Alignment Locations vs. Karyotype (click arrow to hide)

Alignment vs AP003045

Raw Score: 379
PercentID: 98.98
Length: 393
P-value: 1.1e-217
E-value: 1.1e-217

Alignment vs AP003045

Raw Score: 343
PercentID: 96.69
Length: 393
P-value: 3.2e-196
E-value: 3.2e-196

(%ID): 20 - 40 40 - 60 60 - 80 80 - 100

Summary

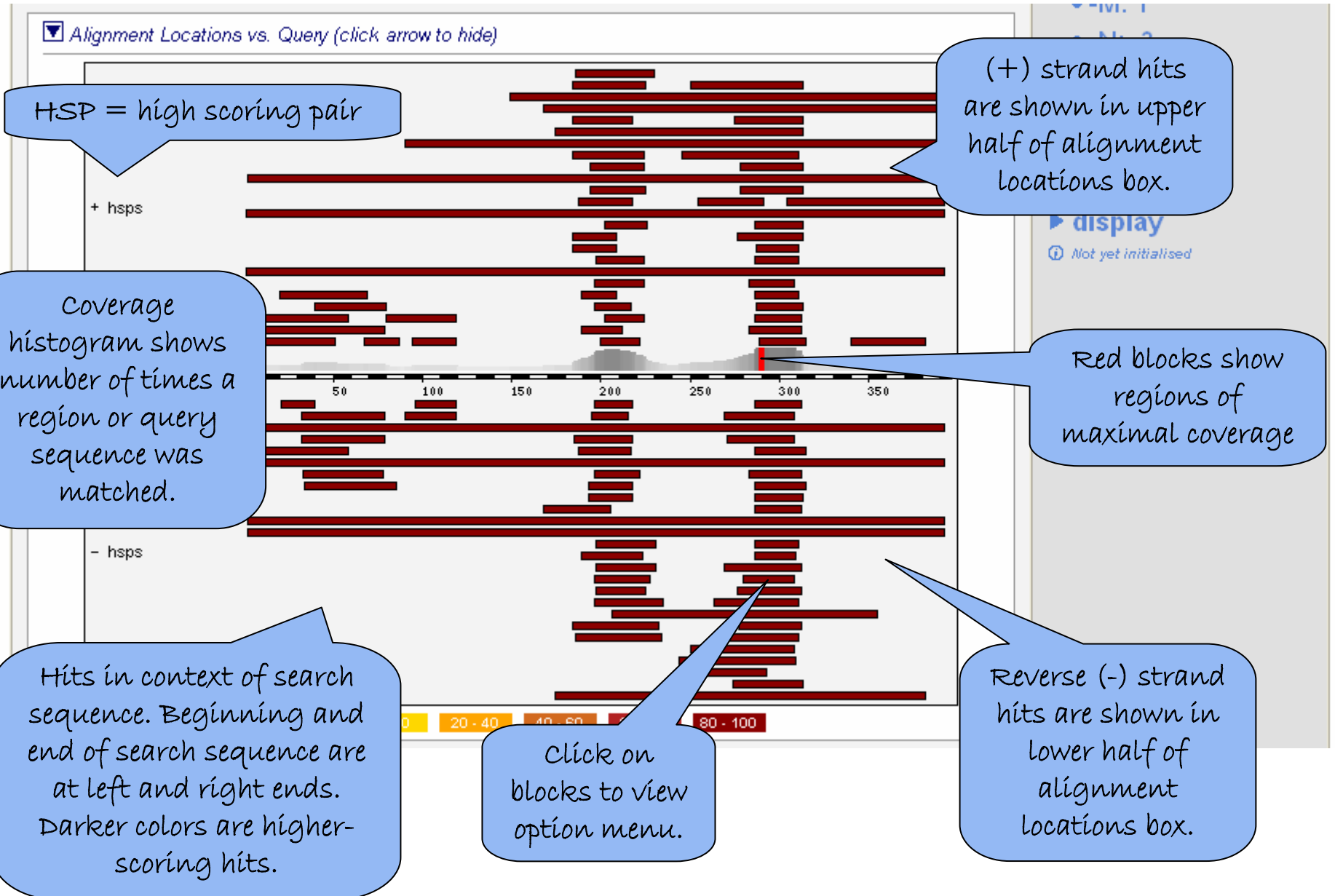
- setup
 - Rice
 - Genomic sequence
 - BLASTN
 - Low sensitivity
- configure
 - E: 10
 - B: 100
 - filter: dust
 - W: 15
 - M: 1
 - N: -3

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

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

Shortcut: You can also click on the best hit in the genome graphic (from Slide 10) to open a menu for that alignment.

Alignment Summary (click arrow to expand)

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

refresh

Query:

Subject:

Chromosome:

Clone:

Stats:

Sort By:

Links

Select from menus to configure results

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

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)

refresh

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	2871274	339	4.6e-193
[A][S][G][C]	2	Chr:1	2841537	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

Alignment Locations vs. Karyotype (click arrow to expand)

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...

Step 4: Review Alignment

The sequence you searched.

```
Query location      : unnamed          1 to      392 (-)
Database location  : AP003045       26309 to 26701 (+)
Genomic location   : 1             2817113 to 2817505 (+)

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

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

```
Query: 392 AGGCGGACGNAGGCTCTGGATCTCCCGTNAGTTGATGGTGGGCTTCTTGTTGTAGCGCGCG 333
      |||
Sbjct: 26309 AGGCGGACGGAGGTCTGGATCTCCCGTGAGGTGATGGTGGGCTTCTTGTTGTAGCGCGCG 26368
```

```
Query: 332 AGCTTGGCGGACTCCCCGGCGAGCTTCTCGAAGATGTCGTTGATGAAGGATGATGATG 273
      |||
Sbjct: 26369 AGCTTGGCGGACTCCCCGGCGAGCTTCTCGAAGATGTCGTTGATGAAGGAGTTCATG 26429
```

```
Query: 272 GACATGGCCTTGGAGGAGATGCCGATGTCGGGGTGGACCTGCTTGAGCACCTT
      |||
Sbjct: 26429 GACATGGCCTTGGAGGAGATGCCGATGTCGGGGTGGACCTGCTTGAGCACCTT
```

```
Query: 212 TAGATCTTGTAGGTCTCGACGCTCTTCTTCGCTTCTTCCGCCCCGCTTCTT
      |||
Sbjct: 26489 TAGATCTTGTAGGTCTCGACGCTCTTCTTCGCTTCTTCCGCCCCGCTTCTTCCCCTCG 26548
```

```
Query: 152 CCGCTGCTCTTCTCGGCTTGCCGGCGGGGAGACGCTTCTCCGCTTGGGCTTCTTCCC 93
      |||
Sbjct: 26549 CCGCTGCTCTTCTCGGCTTGCCGGCGGGGAGACGCTTCTCCGCTTGGGCTTCTTCCC 26608
```

```
Query: 32 GCCAGG-CCTTCTCGGCTTCTCGGCTTGCCGGCGGGGCTCCTCCTCCGCGGGCTTCTTCGCCGCC 34
      |||
Sbjct: 26609 GCCAGGGCCTTCTCGGCTTCTCGGCTTGCCGGCGGGGCTCCTCCTCCGCGGGCTTCTTCGCCGCC 26668
```

```
Query: 33 GGCTTCTTCTCCGCTTGGGCGCCATGGATGCC 1
      |||
Sbjct: 26669 GGCTTCTTCTCCGCTTGGGCGCCATGGATGCC 26701
```

The result you are viewing.

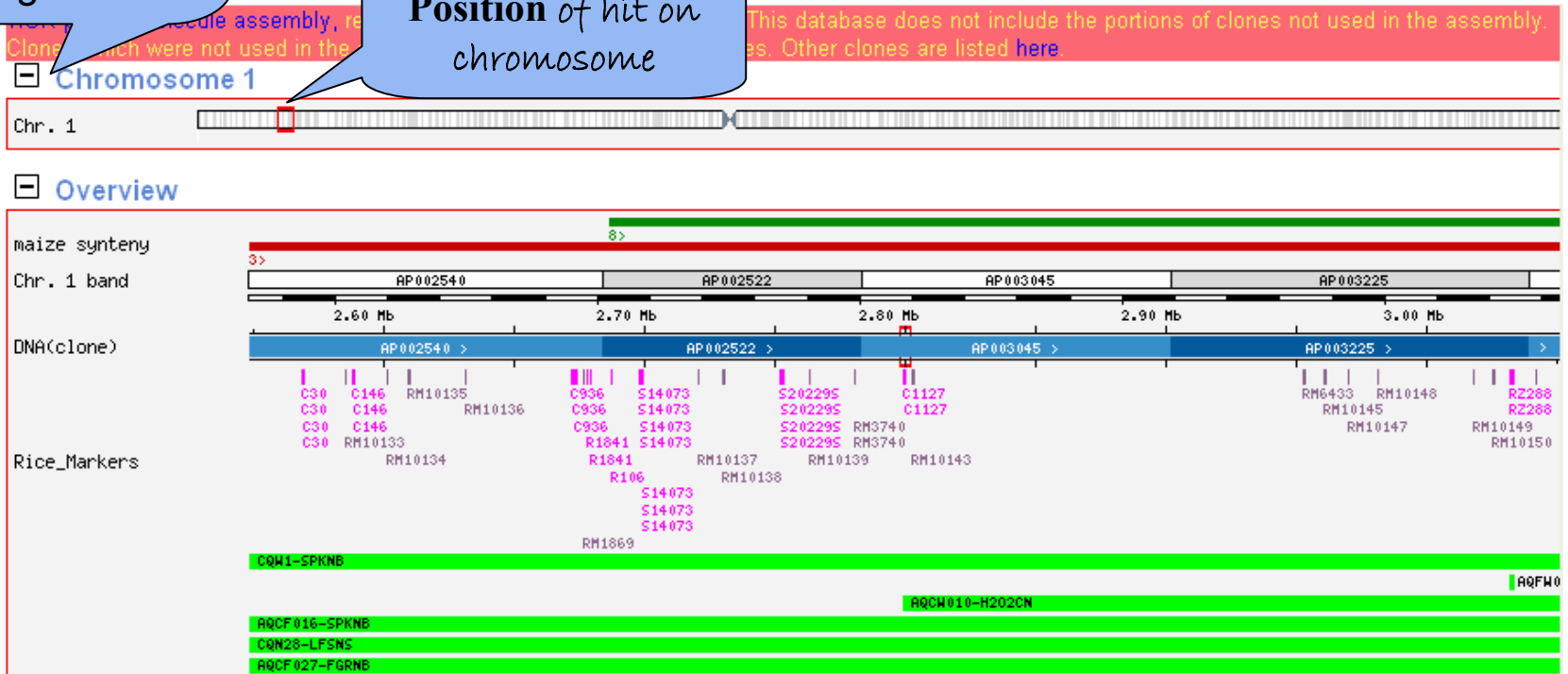
Note the lines identifying matches, and the absence of a line if it doesn't match.

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

Step 5: Review Genome with Genome Browser

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

Position of hit on chromosome



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

Step 5: Review Best Hit on Genome

Note that the overview has been hidden (from previous slide).

Position of hit shown at high magnification.

Use menu to select features displayed

Collaps or Expand tables by clicking on the [+] or [-] signs.

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

Search sensitivity:
Optimise search parameters to find the following alignments

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



Step 6: Adjust Settings - Configure



new SETUP CONFIG RESULTS DISPLAY

refresh Online Help
FAQ

Summary

Run Search

RUN ▶

Configuration for BLASTN

-E	10	Maximum E-value for reported alignments
-B	100000	Maximum number of database hits to report
-filter	1	Program used to filter query sequence
-sort_by	100	Sort option for database hits
-statistics	1	Statistics option for calculation of alignment score
-W	15	Word size for seeding alignments
-wink	1	Step-size for sliding-window used to seed alignments
-hitdist	1	Hit distance for seeding by default
-M	1	Match score
-N	-3	Mismatch score
-Q	3	Cost of first gap character
-R	3	Cost of second and remaining gap characters
-nogap	<input type="checkbox"/>	Turns off gapped alignments
-X	1	Alignment extension cutoff
Additional		Other options (not validated)

Genomic sequence
BLASTN
Low sensitivity

▶ config
① Not yet initialised

▶ results
① Not yet initialised

▶ display
① Not yet initialised

See help for "config" assistance

6b. Click "RUN".

Identifies version of BLAST selected

Each option has listed it's 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.

Click "CONFIG" if you want to change parameters & try again (repeat slide 17).

Step 7: Refined Results

new SETUP CONFIG RESULTS DISPLAY

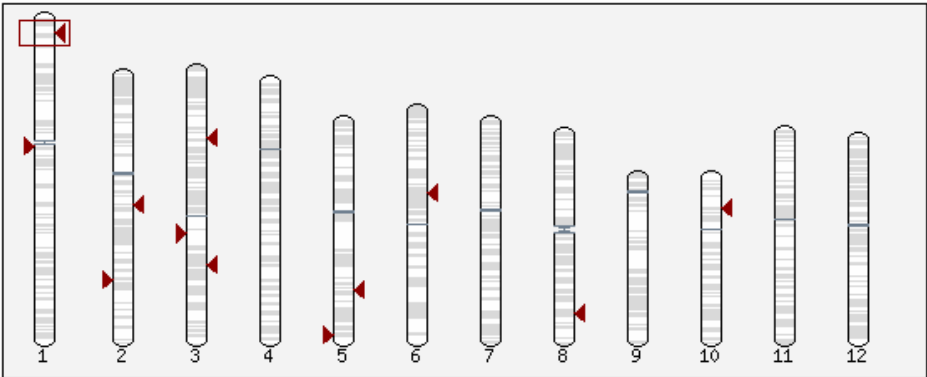
refresh Online Help
FAQ Tutorial

Displaying unnamed sequence alignments vs Rice LATESTGP database

Showing top 100 alignments of 21, sorted by Raw Score

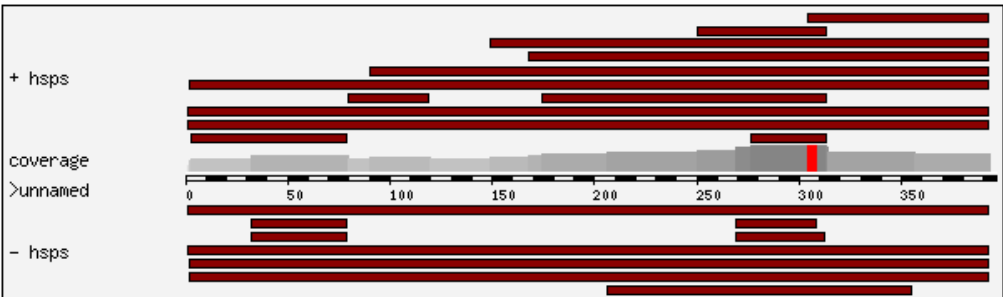
refresh

Alignment Locations vs. Karyotype (click arrow to hide)



Key (%ID): 0 - 20 20 - 40 40 - 60 60 - 80 80 - 100

Alignment Locations vs. Query (click arrow to hide)



Key (%ID): 0 - 20 20 - 40 40 - 60 60 - 80 80 - 100

Alignment Summary (click arrow to hide)

▶ setup

- Rice
- Genomic sequence
- BLASTN
- Custom sensitivity

▶ configure

- -E: 0.001
- -B: 100
- -filter: dust
- -W: 15
- -hitdist: 40
- -M: 1
- -N: -3
- -Q: 3
- -R: 3

▶ results

▶ display

ⓘ Not yet initialised

current parameters in the panel.

More BLAST Information

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



the WU-BLAST web site



EBI 2can introductions for protein or nucleotide BLAST, and



NCBI-BLAST (*a good tutorial although focused on a slightly different implementation.*)

A book on the BLAST family of sequence similarity search algorithms

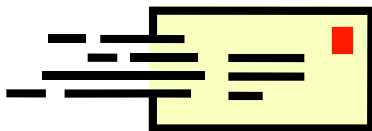
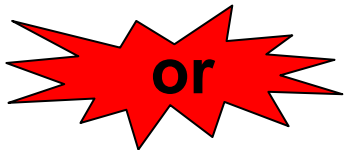


(Joseph Bedell, Ian Korf and Mark Yandell, **BLAST**, O'Reilly & Associates, 2003, ISBN: 0-596-00299-8) provides profound theoretical background, as well as a protocol section covering common practical search problems. The pre-defined optimized parameter sets are based on recommendations in this book

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 Gramene at gramene@gramene.org