

Q01. Which browser is suitable for your web service?

We have tested our software using the following browsers:

- (1) Mozilla *Firefox* 2.0.x (Linux and Windows)
- (2) Microsoft *Internet Explorer* 7.0.x (Windows)
- (3) Google *Chrome* 1.0.x (Windows)

Q02. What happens after I upload my sequence file in FASTA format?

The screenshot shows the WebGMAP web interface. At the top, the logo 'webGMAP' is displayed in blue and green, with the tagline 'A public web service for cDNA-genome mapping' below it. A navigation menu on the left lists various options: Introduction, Genome Browser, Upload Fasta File (highlighted with red arrows), Dump Sequences To DB, Sequence Viewer, GMAP Configuration and Execution, GMAP Result To DB, and GMAP Result Exploration. The main content area is titled 'Upload a Sequence File in FASTA Format'. It shows a 'Project Name' field containing 'guest@muohio.edu_081230_044635'. Below this, there are two sections: 'A. Requirement for your fasta file' and 'B. A sample Fasta file for download and then upload'. Section A lists requirements: max file size is 1 MB, only the first 100 reads are processed, and file names cannot contain certain special characters. Section B lists sample files: Rice.fasta, Arabidopsis.fasta, and Chlamy.fasta. At the bottom, there is a 'C. Upload File:' section with a text input field, a 'Browse...' button, a 'Reset' button, and a 'Submit' button. The footer contains the copyright information: 'Copyright © 2009. BioInfoLab, Botany Department, Miami University, Oxford, Ohio.'

When a user upload his/her sequence file in FASTA format, a WebGMAP project will be created and a specific project name (e.g., guest@muohio.edu_081230_044635) will be associated uniquely with the uploaded file. Within one project, you can only conduct cDNA-genome mapping once, by selecting your desired genome and specifying GMAP parameter settings. Obviously, the project name is the backbone of all database operations and user interface accessing in the future. *Users can create unlimited projects dependent on their needs, but they have to upload a FASTA sequence file for each project.*

Q03. What if I want to reuse the same sequence file and conduct cDNA-genome mapping using different genomes or parameter settings?

You have to upload your sequence file multiple times and create different projects. Consequently, each of these projects is associated with the same sequence file you uploaded. See Q02 for explanation.

Q04. In my sequence file, can I put amino acid sequences instead of nucleotide sequences?

At this moment, our web service can only handle DNA sequences, not amino acid sequences.

Q05. Why do you have the limitation of 100 sequences with the length of 50-2500 nt?

cDNA-genome mapping using GMAP is resource demanding. By such limitation, we try to maintain a good performance of our web server (www.conifergdb.org), which is housing several public bioinformatics databases.

Q06. Within the same sequence file, can I put cDNA or EST sequences from different species?

You can do whatever you want. However, it is recommended that you should put cDNA or EST sequences from the same species together as one single sequence file. Then, you can take advantage of cDNA-genome mapping options we provide: intra-species versus inter-species (see **Q07** for details).

Q07. What is intra-species or inter-species cDNA-genome mapping?

Currently, we provide the genomes of *Arabidopsis thaliana*, *Oryza sativa* and *Chlamydomonas reinhardtii* for mapping. If you have cDNA or EST sequences that are from the same species, you should select “intra-species” mapping. Otherwise, you need to choose “inter-species” mapping (or cross-species mapping).

What happens behind the scene is that we adopt different filtration conditions for inter-species and intra-species mapping results:

(1) For intra-species mapping, only the hits that satisfy the following criteria are the valid hits.

- (1.1) minimum matched query length=50 (nt)
- (1.2) minimum matched query identity=60 (%)
- (1.3) minimum matched coverage of query sequence=30 (%)

(2) For inter-species mapping, only the hits that satisfy the following criteria are the valid hits.

- (2.1) minimum matched query length=30 (nt)
- (2.2) minimum matched query identity=30 (%)
- (2.3) minimum matched coverage of query sequence=10 (%)

Please keep in mind that in most cases, GMAP will not produce bad mapping results, such as identity of 30%. In the near future, we will allow users to set up the filtration for their valid cDNA-genome mapping hits.

Q08. I got the raw GMAP result file. However, when I clicked “GMAP Result To DB” to proceed, I was told that no hit has been saved successfully into the database. Why?

That is the intra-species or inter-species filtration we established (See Q05 for details). We only save the valid GMAP hits into our database for further exploration.

Q09. What is the main functionality of your sequence viewer?

As you can see in the figures below, you can easily search, filter and order your sequences by their names or lengths, as well as their SeqID, the internal identifier designated by us.

Valid Sequence List

Search:

sid =
 seqname =
 len =

Search type:

Results: 1 - 8 of 8 Pages: <<< << 1 >> >>> Page size: 10

Seq Identifier	Seq Name	Seq Len
228	CACW7241.b1	971
229	CACW6542.b1	988
230	963032D06.y2	775
231	963013B07.y9	813
232	894013E10.y1	1020
233	1031050A07.y1	765
234	1031040E09.y2	748
235	CACW8187.g1	997

Results: 1 - 8 of 8 Pages: <<< << 1 >> >>> Page size: 10

Sequence Control Panel

Protein Translation Frame: Alignment Space Separator Reverse Complement

Motif Search: Motif Max Error:

>894013E10.y1

```

1      11      21      31      41      51      61      71      81      91
0  TTCGAATTAA  CCCTCACTAA  GGAACA AAAA  GCTGGAGCTC  CACCGCGGTG  GCGGCCGCTC  TAGAACTAGT  GGATCCCCCG  GCGTCGAGGA  ATTCCGCACG
1  AGGGTGACGT  TAGATACCTA  CTTGACAAAT  ACAAGCCTTC  TAGAATGAAA  TCGCGGAGAG  CCGCAGCGGT  CTTGCTGCTG  GCGATTATTG  CCTGCAGCGC
2  GCTTTTAGCT  GCTGGCGCCC  CAGCGGATAG  CAAGGATGCC  AAGCCGAAGA  GGAAGAACCT  GCCACCAAG  GCATTCATT  ACGAGGGCAA  GCACTGGACG
3  GTGCACAAGA  CTTCCGAGGG  GCGCGCTTC  TTCTACAACG  TGGGAGCGGG  CGCCACCCAG  TGGACCGACC  CCCGAGTGG  GCAAAATGACG  CCGCAGCAGC
4  GCATCGTGGT  CATCGCCATG  TTTGCCACGG  CCTTTGTGCT  CATGATCGCC  GCGCGCGCGG  CCTACATCCT  CTGGGTCCGG  AGCAAGCACC  CCGAGCTGCT
5  GAAGGGCCCG  AAGAAGGTGA  AGGGCCTCAA  GAACTGGGAG  CCGCGCTGCG  AGCCGCCCAA  GGTGGCCAAG  CCCCGCGGG  GCAGCTCCAG  CCCGCCGCCC
6  GAGGACCGCG  TGAAGCGCGG  CGCCTTGCCG  GCACTGGCCG  CCGGAGGAGA  AAAGGACCC  TAAGGGCCAC  TTGATGAGGC  GTGCTGCTGA  TACCCAGGG
7  CAGGACGAGG  AGGGTAAGCA  TGGAGGGATA  GGAGGGCGCG  TAACCCGGAT  AGACACGGCG  GGGGGGTGT  GTGCCGCGTG  CGCACCACAA  CTGAGGACGT
8  GGCAAGCGCG  CGGCCCTGG  GACACATGAT  TGCAAAAACG  GACCATAAAA  TGGGGGGGGA  GGCAAGCATG  CCCTGGCCTC  CGGGTTTCGG  GTGGACCCGG
9  GGGTGGGCGC  ACAAGCAACA  TTCTTCTTCG  TATTTGAGG  CCCCCGAGCG  GACCTGTGGA  AATTCGGTGA  ACTCCCAAGG  GAGGCCCA  CCCCAGGCTC
10 CCAGGAGTA  ATTGGACCC
    
```

>Motif Alignment: Match1

```

1      TTTTTGAGGC      10
      |||||
932  ATTTTGAGGC      941
    
```

For each individual sequence, you can do 6-frame translations, reverse complement the sequence, or use space separator to track nucleotide positional information. Moreover, we provide fuzzy motif detection that allows users to search a given motif in the sequence, and if existing, we provide sequence alignments for the detected motif(s) in the sequence.

Q10. How can we use your interface to navigate the mapping results?

Search

hit_id =

ref_name =

identity =

Search type:

Results: 1 - 8 of 8 Pages: << << **1** >> >> | Page size:

N #	CStart	Chromosome	QName	Coverage	Identity	QStart	QEnd	QDir	CStart	CEnd	HSP_No
1	191	Chr01	BR060005A10G06.ab1	97	98.4	19	691	Antisense	9817	8499	4
2	192	Chr01	C1220977	96.1	99.6	1	544	Sense	9107	9809	3
3	187	Chr01	H063C04	92.2	93.3	24	528	Sense	4466	6532	2
4	188	Chr01	OSJNEa13L20.f	88.4	98.6	82	711	Sense	2068	4524	4
5	190	Chr01	OSJNEe05H05.f	84.1	98	83	1114	Sense	885	6732	5
6	193	Chr01	OSJNEe05H05.r	88.5	96.5	77	1171	Antisense	9817	7465	5
7	189	Chr01	OSJNEf07N20.f	86.8	92.3	94	1081	Sense	1972	6371	5
8	194	Chr01	OSJNEf07N20.r	92.4	92.8	80	1096	Antisense	9749	7517	5

Results: 1 - 8 of 8 Pages: << << **1** >> >> | Page size:

We provide users a tabulated view for their GMAP results. Again, a user can search, filter and order their results easily. By clicking the CStart number, a graphic interface will be popped up to display the detailed gene structures. By zoom in to the lowest resolution, you can see the sequence alignments at the nucleotide level.

Chromosome Start End Order By Desc Show Filter

Color Legend

Chromosome: + █ - █ Annotated Gene: █ Representative mRNA: 5'UTR █ 3'UTR █ UTR █ CDS █

EST: 5TSS █ 3TSS █ 5TNS █ 3TNS █ polyA █ polyT █ Mismatch Insertion █ Deletion █ Unmapped Seq █

Scroll/Zoom

Unit: 100 bp

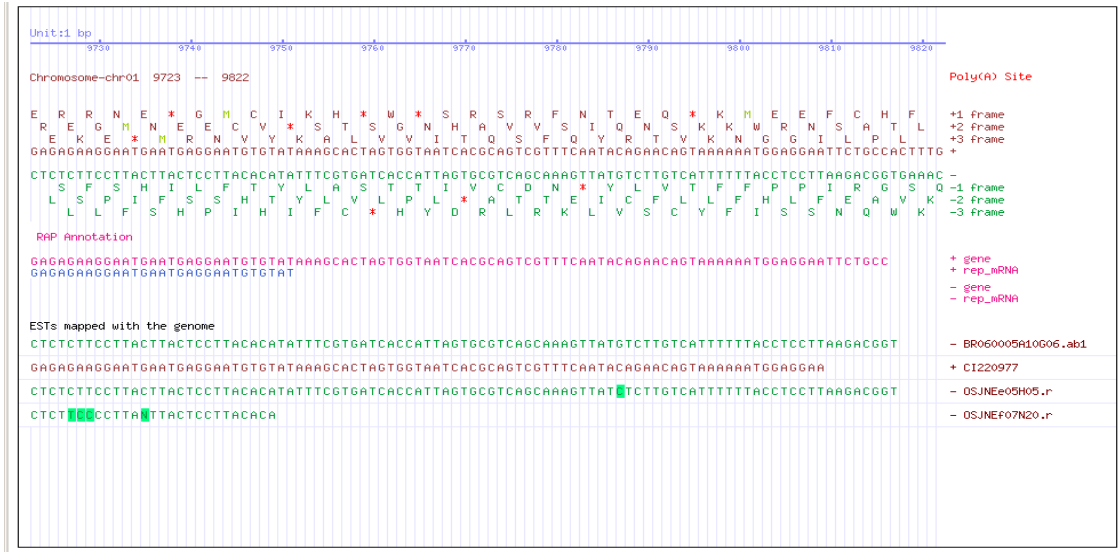
Chromosome—chr01 1 --- 10000

Poly(A) Site +

RNP Annotation

ESTs mapped with the genome

- + gene
- + rep_mRNA
- gene
- rep_mRNA
- BR060005A10G06.ab1
- + C1220977
- + H063C04
- + OSJNEa13L20.f
- + OSJNEe05H05.f
- OSJNEe05H05.r
- + OSJNEf07N20.f
- OSJNEf07N20.r



Q11. Can you process data from new sequencing technologies, for example, 454 and Solexa data?

Currently, we have a sequence length limitation of 50-2500 nt. This should be fine with 454 data which usually have a length of more than 100 nt. In the future release, we will allow users to process their short (*i.e.*, 30 nt) Solexa data.