



## Searching dbSNP using Sequence Data Homology (Search using BLAST)

Created: July 7, 2005; Updated: February 25, 2014.

### Basic BLAST Search Strategies

#### Can I search dbSNP for sequences using BLAST?

The dbSNP BLAST search is located on the left side bar of the dbSNP homepage, or you can [get there directly](#).

### Search dbSNP using a Sequence

#### Can I determine if a SNP has a refSNP ID number by BLASTing the sequence containing the SNP?

Yes. Use the [BLAST SNP](#) page, which is accessible from left side bar (Header: "Search") on the [dbSNP home page](#). (05/23/08)

#### Do you have a resource that will allow a user to BLAST against dbSNP with sequences as short as 50 bp?

You can access our resource, "BLAST SNP" by clicking on the word "Search" located in the left blue sidebar on the dbSNP home page to release a drop-down menu, and then clicking on the words "BLAST SNP" to go to the [BLAST SNP](#) page. Using BLAST SNP, you can currently find if your sequence aligns with a known SNP flanking sequence. In the future, we will update SNP flanking sequences with known neighboring SNPs using the [IUPAC](#) code. (9/24/07)

#### How do I determine if SNPs I've discovered have already been reported to dbSNP by another lab?

Align your sequences against those in the human SNP database using a program called BLAST.

Click on BLAST SNP on the SNP homepage side bar, which brings you to [SNP BLAST by chromosome](#).

SNP variations are encoded using IUPAC notation in the BLAST database. If the position of your variation does not match any of the SNP positions as indicated by the IUPAC notation, then you have a new SNP.

Example of BLAST output:

```
Query: 241  gtgcttcctgggctccctggctgtgctgctgtgtgtgtgtgcaacggagcgtgtgcagtacta
          |||
Sbjct: 313  gtgcttcctgggctcsctggctgtgctgctgtgtgtgtgtgcaacggagcgtgtgcagtacta
```

You can see in the example BLAST output above that there is an existing variation in dbSNP (note: the IUPAC notation "s", shown above, represents g or c).

#### Is it possible to search dbSNP using a sequence rather than an accession number or rs number?

Yes. You can align your sequences against those in the human SNP database by using a program called BLAST.

Click on BLAST SNP on the SNP homepage side bar, which brings you to [SNP BLAST by chromosome](#). SNP variations are encoded using IUPAC notation in the BLAST database.

Here is an example of the BLAST output:

```
Query: 241  gtgcttcctgggctccctggctgtgctgctgtgtgtgtgcacggagcgtgtgcagtacta
          |||
Sbjct: 313  gtgcttcctgggctcsctggctgtgctgctgtgtgtgtgcacggagcgtgtgcagtacta
```

You can see in the BLAST output above that there is an existing variation in dbSNP (note: the IUPAC notation “s”, shown above, represents g or c).

## BLASTing a Batch of Sequences

### Is there a way to BLAST 500 sequences in a single batch?

dbSNP doesn't have a service for BLASTing a batch of sequences. You'll have to do it programmatically on your side using the [Blast API](#).

The parameters you need to past to the blast API are:

```
ALIGNMENTS=100
ALIGNMENT_VIEW=PairwiseWithIdentities
AUTO_FORMAT=Semin_auto
CMD=Put
DATABASE=human_9606/human_9606
DB_DIR_PREFIX=snp
DESCRIPTIONS=100
EXPECT=0.01
FILTER=L
FILTER=R
FILTER=m
INPUT_TYPE=FASTA+format
MEGABLAST=yes
NOENTREZLINKS=OTHER_ADVANCED
PAGE=Nucleotides
PROGRAM=blastn
SHOW_OVERVIEW=yes
WWW_BLAST_TYPE=SNP
submit2=Submit+Query
QUERY=CATTTTCTAGGTATGAACACTGTGCATGATGAAGTCTTTCCAAGCCACACCAGTGGTTCATGTGTG
TGCACCTCCGGTTTGAGTGCTAGTGAGATACTTCTGTGGTCTGAATTGCCTGACTATTTGGGGTTGTGATAT
TTTCATAAAGATTGATCAACATGTTTTCGAATTTCCCTCCCCAACAGTCTTCCATTACCAAGTAAAGATTCAATTT
TCTGGGACTGAGAGTGAAACCCATAACCAATCAGGCCTTTGAGATTTCTCTGTATGGCACCGTGGCCGAGAGTG
AGAACATCCCATTC R CTCTGTGAGTAGCACAGGGGGGGCGGTCATCATGGCACCCAGTCCCTCTCCTGCCATA
ACCTTGGTCTGAGCAGCAGAAGCAGAGAGCGATGCC TAGAAAAACAAGTCTTTAGTTAAAAAAAATCAGAATTT
CAAAATTGAGGTCTTTCCCTCTATTTGATATTGAGAAAAAAAATGCTTCAAATTGGCCATTTTATTTTCACTTAC
TAGTTATATTTTATTTATTTATCATCTTATATCTGTTTATTTCTTTTATAAAGCTGCTGTTAAACAATATAATT
AAACTATCTCAAAGGTTTGACATT
```

Leave all the parameters the same for each of your BLAST requests and substitute the above query sequence with your query sequence.

(10/15/08)

## Searching dbSNP using Primer Sequence

I have forward and reverse primer sequence, the projected amplicon size and the site of the SNP. What is the best way to retrieve the rsID for this SNP?

We don't have primers sequences stored in dbSNP. Since dbSNP is annotated on uniSTS, you might try searching in dbSTS or uniSTS to find the SNP. (01/08/08)

## Searching for SNPs not yet Submitted to dbSNP

**I have found a novel SNP for and would like to confirm that nobody else has reported it. How do I do this?**

To confirm that no one else has submitted the SNP previous to your submission, try [BLASTing](#) using your flanking sequences.

dbSNP does accept submissions for an existing SNP, however, since these additional submissions serve to validate the existing SNP if the additional submissions were assayed using a different method or were derived from a different sample. We assign a unique submitted SNP (ss) number for each SNP submitted. Please see the "[Submission Quick Start](#)" section of this archive for further details(01/10/08)

**How do I search dbSNP to determine if some SNPs and deletion/insertion variations I have identified are novel?**

Try BLASTing dbSNP with your sequence. To do this, go to the [SNP home page](#), locate the "Search" section in the left-hand side bar and click on "[Blast SNP](#)". (3/8/05)

**How do I determine if a sequence I have contains any of the variations housed in dbSNP?**

You can try to BLAST dbSNP by clicking on the BLAST SNP button located on the dbSNP homepage sidebar, which will take you to [SNP BLAST by chromosome](#).

**How do I determine if SNPs I've discovered have already been reported to dbSNP by another lab?**

Align your sequences against those in the human SNP database using a program called BLAST.

Click on BLAST SNP on the SNP homepage side bar, which brings you to [SNP BLAST by chromosome](#).

SNP variations are encoded using IUPAC notation in the BLAST database. If the position of your variation does not match any of the SNP positions as indicated by the IUPAC notation, then you have a new SNP.

Example of BLAST output:

```
Query: 241  gtgcttctctgggctccctggctgtgctgctgtgtgtgtgcacggagcgtgtgcagtacta
          |||
Sbjct: 313  gtgcttctctgggctcsctggctgtgctgctgtgtgtgtgcacggagcgtgtgcagtacta
```

You can see in the example BLAST output above that there is an existing variation in dbSNP (note: the IUPAC notation "s", shown above, represents g or c).

**How are the contig and chromosome positions of SNPs determined? I found some new polymorphisms and would like to determine these positions.**

Here at dbSNP, we map refSNP sets ourselves using MEGABLAST. Most commonly, the contig position of a polymorphism is found when the SNP is mined computationally. For instance, the [Shttp://www.ncbi.nlm.nih.gov/SNP/snp\\_ss.cgi?subsnp\\_id=22190651](http://www.ncbi.nlm.nih.gov/SNP/snp_ss.cgi?subsnp_id=22190651) SAHASNP program mines SNPs from overlapping traces of the clones used in the genome assembly.

Please look at [this example](#).

Three of the submitter records for refSNP cluster rs3736544 were computationally mined (see those with [http://www.ncbi.nlm.nih.gov/SNP/snp\\_ss.cgi?subsnp\\_id=22190651](http://www.ncbi.nlm.nih.gov/SNP/snp_ss.cgi?subsnp_id=22190651) SSAHASNP in the Handle/Submitter ID field), and the contig coordinates were used to construct the submitter ID.

If you have already sequenced the flanks surrounding your polymorphisms, you can [use the form](#) on the contig BLAST page to determine the contig coordinates.

If you have a large number of contigs, it is more practical to [download the entire genome sequence](#) and BLAST them yourself.

## Unable to get Sequence Comparison Results

I've tried BLASTing dbSNP in *Homo sapiens* chromosome 6 using a sequence that contains a known SNP but get no sequence comparison results—just the RID number (1094764833-2783-170044276144.BLASTQ4).

I noticed that you included your BLAST Request ID (RID) in your question. GOOD!

It is always a good idea to save your Request ID (RID) when it first appears at the top of the format page, which is displayed after you select Submit. The RID will allow you or anyone else to retrieve the formatted BLAST result of your query without redoing the time-consuming BLAST.

I used your RID number to retrieve your BLAST output and found that the alignment was indeed returned. The fact that you did not see an alignment is puzzling. It could be that you possibly did not have the correct formatting options checked on the format page, but I think they are set to sensible defaults automatically.

Please try this to view the results of your 196-base example using the RID you provided and the following instructions:

1. Go to the [main BLAST page](#) and click on Retrieve results by RID under the word Meta, which takes you to the [BLAST Request ID](#) page.
2. Now, enter your request ID, 1094764833-2783-170044276144.BLASTQ4, in the format options box. Please make sure that the number of descriptions and the number of alignments are set greater than zero. Now, click on format.
3. You should see the page displayed. If this doesn't work, please let us know. If it does, try repeating your query to dbSNP, copy your RID somewhere just in case, and click on “format”. If it doesn't format, use your browser to go back to the previous format page and repeat these steps as needed until output appears.
4. Make sure that the formatting options are sensible, and that the number of descriptions and the number of alignments are set greater than zero. Also, indicate that the alignment should be shown in HTML (“show !alignment! in !HTML!”). The alignment view should probably be set to !pairwise! unless you know better.

## Looking for SNP BLAST URL that Includes pre-set Parameters

We are in the process of developing SNP web pages, and was wondering if you have a URL for SNP BLAST that includes the parameters (sequence, organism) already filled in.

The best method is to use URL API, and the best documentation for using this is on the [NCBI Learning Center](#) website. Once you are on the Learning Center website, scroll down the list and select “BLAST URLAPI”.

Toward the middle of the page there is also a link called “Remotely accessible BLAST databases”. Please note, however, that some of the contents of this document are out of date.

Example:

To call the zebrafish snp database, use the following text in your URL:

```
&DATABASE=snp/zebrafish_7955/zebrafish_7955&
```

Other BLAST parameters are specified on the “BLAST URLAPI” page.

(4/19/06)

**I want to download the current set of sequences in dbSNP to format for a local BLAST search. Could I use a "wget --mirror" to pull down the contents of /snp/organisms/\*/ss\_fasta/\*/\*", and then concatenate the files?**

We already have the files you need on the dbSNP FTP site as FASTA files. For example, to get [human ss FASTA data](#), go to the human organism directory, and select ss\_fasta, and then select the year in which the data you need was submitted.

You can also get [rs FASTA organized by chromosome](#) in the dbSNP FTP site, or you could also blast dbSNP rs sequences directly [online](#).

For more information, please check the [dbSNP handbook](#).

**(9/13/06)**