GenomeQuest Quick Start Guide

Variant Version

How to enter a search into GenomeQuest

1. Select **IP Search** as search type from the main GenomeQuest page.

2. Select the query sequence type (either nucleotide or protein) sequence in the **Query** box by clicking on the appropriate tab and paste the desired sequence in the box. Remember that it must be in FASTA format, and all query sequences must be identified in the header line. It's best to not have spaces in the header.

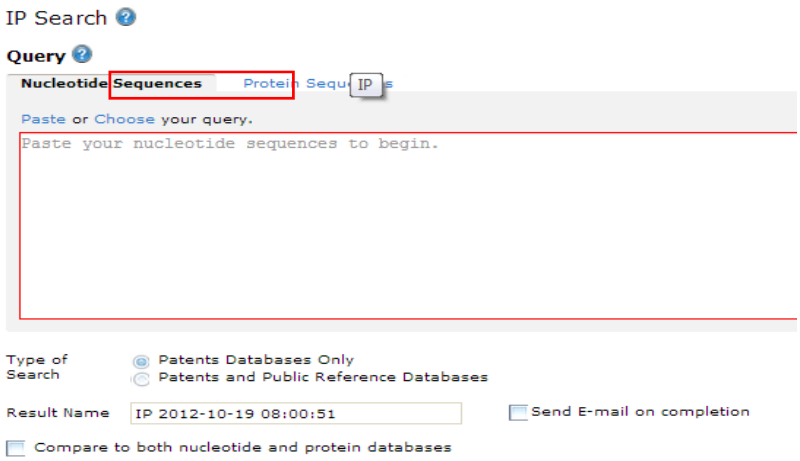
   Example:

   >US20100223293-16257
   CTGCAGGCGAACTGTTCTGTTCCGGTAGGGAATTCACCGCAGGCGAGAACGCGGCCGGCT
   GCAAAGCTTGGCGAGTACGGAGTAAAACTTATTTTTTGAGGGCTGCCGCCTTTGGACAAAT
   CCAGTAAACCCACCGAGCTTCGGAAATGTGGGACTGAGAAGGGACGGCGATCCCAGATCA
   CACAGAGGACAGGGGAAAACGAAGCCACCGAGCCCCCACACGTCGCCATCCATCGCCGTA
   ATCGATCACCGCCGTCTCCTCCCCCACACACCCACCGGAACCGTCGTCCTGACCTCTCGCCA
   GCGATAAGCAAATCCCCTCCCCACTTTATCGTCCACAAAGCCTTCTTCCCGCCCTCCCGAATC
   GCTCCCTCTCTGTCCCTCCGCTCCAGCCGCCGCCGCCGCCTCCGCCCTCCGAATCCCAATGG
   CGTCCGCGGCCGCTCCTCCAACCTCCCTCTCCCTCGCGGCCCGCGCGGCCACCAGGGCAGC
   CGCCGCCGCCGCCGCCCCGC

3. For all search types other than Patentability, select the option of **patents databases only** under the **Type of search**. For patentability searches please select **Patents and public reference databases**.

4. Provide a suitable and easily recognizable name under **Result name,** such that one can easily identify the desired search results.

The screenshot below illustrates how the window for entry of nucleotide query sequences will look. Note that the **NUCLEOTIDE SEQUENCES** tab is selected for this example of a nucleotide query sequence.



5. Optionally, you may select the box "**compare to both nucleotide and protein databases**" for protein or nucleotide query sequences. Some examples when one might NOT to compare against both nucleotide and proteins are if:
    a. nucleotide sequence is not expressed: e.g. promoter, terminator, intron, or other non-expressed sequence;
    b. nucleotide sequence is genomic, or a whole genome (very unlikely to have such a query sequence, but it can happen);
    c. sequences are very short, such as primers, probes;
    d. fusion proteins or other synthetic proteins not from nature;
    e. vectors, plasmids, vector containing a gene (although you may want to search the gene against protein as well separately)'
    f. any other query sequence where it doesn't make scientific sense to search it against both.

6. Select the appropriate algorithm under the **"search strategy"** option:

Query	Algorithm
CDR or other short peptide	GenePast
Sequence longer than 25-50 residues	Blast or GenePast
Sequence with variable positions	MOTIF

For all algorithms, you have the choice to limit subject length.

If you are looking for sequences where you have a reasonable idea of what length to expect, then set the "limit subject length" to reflect this. For example, if you are searching a promoter in isolation, depending on the promoter length you might set this filter to 2000 (or even 10,000) in order to exclude genomic DNA. If you are looking for primers or probes, you might set the filter to between 10 and 50 residues.

MOTIF options (use for variable sequences)

The MOTIF algorithm is used for searches where at least one position can have more than one choice. This can include indels or different residues. Be aware that this does require 100% identity to all residues, so if there is even a single mismatch, you will not find that hit. There are methods available to address this in more detail.

If you have a sequence with a variant position, here is how you would write a single substitution. You can build from here.

Sequence YAGXL where X can be P or T is written. YAG[PT]L, for example.

The link below goes to specific documentation:
https://docs.genomequestlive.com/ - motif

Here is a GenomeQuest blog entry with one strategy for variant searching:
https://www.gqlifesciences.com/hunting-single-point-mutations-biological-sequences-crisprcas9-example/

A slide deck with another MOTIF strategy will be coming in the near future.

General help information: https://docs.genomequestlive.com/

For further assistance: support@gqlifesciences.com