



Search Field Descriptions for Sequence Database

Monica Romiti, M.L.S.¹ and Peter Cooper, Ph.D.²

Created: December 3, 2010; Updated: February 9, 2011.

Table 1. Fields available for all Sequence Databases (Nucleotide, Protein, EST, GSS). Fields only available for the EST and GSS databases are given in Table 2.

Search Field	Short Field Specifier	Definition
[Accession]	[ACCN]	The accession number assigned by NCBI. <i>Examples:</i> AF123456[ACCN] Nucleotide NP_000240[ACCN] Protein
[All Fields]	[ALL]	All terms from all search fields in the database. <i>Example:</i> human[All Fields] Nucleotide Protein EST GSS (Compare with human[Organism], see [Organism] entry in this table.)
[Author]	[AU] [AUTH]	All authors from all references in the records. The format is last name [space] first initial(s), without punctuation. <i>Example:</i> venter jc[AUTH] Nucleotide Protein
[EC/RN Number]	[ECNO]	Enzyme Commission (EC) number for an enzyme activity. <i>Example:</i> 5.3.1.9[ECNO]) Protein Nucleotide (glucose-6-phosphate isomerase)

Table 1. continued from previous page.

Search Field	Short Field Specifier	Definition
[Feature Key] (Nucleotide, Protein, GSS)	[FKEY]	Biological features listed in the Feature Table of the sequence records. <i>Examples:</i> polya signal[FKEY] Nucleotide nonstdres[FKEY] Protein gene[FKEY] GSS The GenBank feature table definition has more information on available features.
[Filter]	[FILT] [SB]	Filtered subsets of the database. An important kind of filter is based on the presence of links to other records. Other filters create useful subsets of data such as those set as Filters in the Discovery column of search results <i>Examples:</i> <i>Links</i> nucleotide_protein[Filter] Nucleotide protein_structure[Filter] Protein nucest_unigene[Filter] EST nucgss_unists[Filter] GSS <i>Organism or properties subsets</i> all[filter] Nucleotide Protein EST GSS mrna[filter] Nucleotide refseq[filter] Nucleotide Protein mammals[filter] Nucleotide Protein EST GSS
[Gene Name]	[GENE]	Gene names annotated on database records. For NCBI Reference Sequences, these names correspond to official nomenclature guidelines when possible. Submitters provide the gene names on GenBank/GenPept records. Gene names on submitted records may be historical names or vary from official guidelines for other reasons. <i>Example:</i> BRCA1[GENE] Nucleotide Protein
[Genome Project]	-	The numeric unique identifier for the genome project that produced the sequence records. <i>Examples:</i> 13139[Genome Project] Nucleotide Protein (<i>Oryza sativa Japonica</i>) 21117[Genome Project] Nucleotide EST GSS (Pelagic Microbial Assemblages in the Oligotrophic Ocean)
[Issue]	[ISS]	The issue number of the journals cited on sequence records, not generally useful in sequence databases.

Table 1. continued from previous page.

Search Field	Short Field Specifier	Definition
[Journal]	[JOUR]	<p>The name of the journals cited on sequence records. Journal names are indexed in the database in abbreviated form although many full titles are mapped to their abbreviations. Journals are also indexed by their by International Standard Serial Number (ISSN).</p> <p><i>Examples:</i></p> <p>proceedings of the national academy of sciences of the united states of america[Journal] Nucleotide Protein EST GSS Proc Natl Acad Sci U S A[Journal] Nucleotide Protein EST GSS 0027-8424[Journal] Nucleotide Protein EST GSS</p>
[Keyword]	[KYWD]	<p>Keywords applied by submitter or from controlled vocabularies applied by NCBI or other databases. Except for specific kinds of records, such as the examples given below, the terms in this index are not well controlled. This field is unpopulated for many GenBank/GenPept records.</p> <p><i>Examples:</i></p> <p>BARCODE[KYWD] Nucleotide Protein HTG[KYWD] Nucleotide RefSeqGene[KYWD] Nucleotide WGS_MASTER[KYWD] Nucleotide</p>
[Modification Date]	[MDAT]	<p>The date of most recent modification of a sequence record. The date format is YYYY/MM/DD. Only the year is required. The Modification Date is often used as a range of dates. The colon (:) separates the beginning and end of a date range.</p> <p><i>Examples:</i></p> <p>2009/01/08[MDAT] Nucleotide Protein EST GSS 1995/09[MDAT] Nucleotide Protein EST GSS 2010/01:2010/12/31[MDAT] Nucleotide Protein EST GSS</p>
[Molecular Weight] (Protein only)	[MOLWT]	<p>The molecular weight in Daltons of the protein chain calculated from the amino acids only. This may not correspond to the molecular weight of the protein obtained from biological samples because of incomplete data or post-translational modifications of the protein in living systems. The colon (:) separates the beginning and end of a molecular weight range.</p> <p><i>Examples:</i></p> <p>3039[MOLWT] Protein 25000:75000[MOLWT] Protein</p>
[Organism]	[ORGN]	<p>The scientific and common names for the complete taxonomy of organisms that are the source of the sequence records. This vocabulary includes all available nodes in the NCBI taxonomy database.</p> <p><i>Examples:</i></p> <p>cellular organisms[ORGN] Nucleotide Protein EST GSS firmicutes[ORGN] Nucleotide Protein human[ORGN] Nucleotide Protein EST GSS Escherichia coli O157:H7[ORGN] Nucleotide Protein</p>

Table 1. continued from previous page.

Search Field	Short Field Specifier	Definition
	Sequences_help_apppe.TF.2	
[Page Number]	[PAGE]	The page numbers of the articles that are cited on the sequence record, not generally useful in sequence databases.
[Primary Accession]	[PACC]	<p>The primary accession number of the sequence record. This is the first one appearing on the ACCESSION line in the GenBank/GenPept format. Many records have additional secondary accessions representing records that have been merged. The Accession field indexes both primary and secondary accessions.</p> <p><i>Examples:</i></p> <p>U01317[PACC] Nucleotide M18047[PACC] Nucleotide (Compare: M18047[ACCN] Nucleotide, see [Accession] entry in this table.)</p>
[Primary Organism]	[PORGN]	<p>The primary organism when there is more than one source organism.</p> <p><i>Examples:</i></p> <p>human[PORGN] Nucleotide (Compare with human[ORGN], see [Organism] entry in this table.)</p>
[Properties]	[PROP]	<p>Molecular type, source database, and other properties of the sequence record. Terms indexed for this field are a useful classification system for sequence records.</p> <p><i>Examples:</i></p> <p><i>Molecule type</i></p> <p>biomol_crna[PROP] Nucleotide biomol_genomic[PROP] Nucleotide biomol_mrna[PROP] Nucleotide</p> <p><i>Cellular location</i></p> <p>gene_in_genomic[PROP] Nucleotide Protein gene_in_mitochondrion[PROP] Nucleotide Protein</p> <p><i>GenBank division</i></p> <p>gbdiv_htg[PROP] Nucleotide gbdiv_vrt[PROP] Nucleotide Protein</p> <p>(These GenBank division queries must be combined with srcdb_genbank[PROP] to retrieve only GenBank records.)</p> <p><i>Database source</i></p> <p>srcdb_genbank[PROP] Nucleotide Protein EST GSS srcdb_ddbj/embl/genbank[PROP] Nucleotide Protein EST GSS srcdb_refseq_known[PROP] Nucleotide Protein srcdb_refseq_predicted[PROP] Nucleotide Protein srcdb_swiss-prot[PROP] Protein srcdb_pdb[PROP] Nucleotide Protein</p>

Table 1. continued from previous page.

Search Field	Short Field Specifier	Definition
[Protein Name]	[PROT]	<p>The names of protein products as annotated on sequence records. The content of this field is not well controlled for GenBank/GenPept records and may contain inaccurate or incomplete information.</p> <p><i>Examples:</i></p> <p>aldolase[Protein Name] Nucleotide Protein</p>
[Publication Date]	[PDAT]	<p>The date that records were made public in Entrez. The date format is YYYY/MM/DD. The colon (:) separates the beginning and end of a date range.</p> <p><i>Examples:</i></p> <p>2009/01/08[PDAT] Nucleotide EST GSS 2009/01/10[PDAT] Protein 1995/09[PDAT] Nucleotide Protein EST GSS 2010/01:2010/12/31[PDAT] Nucleotide Protein EST GSS</p>
[SeqID String]	[SQID]	<p>The NCBI identifier string for the sequence record. This is a brief structured format used by NCBI software.</p> <p><i>Example:</i></p> <p>gnl asm gca 000000215 2 chr3 45328308[SeqID String] Nucleotide</p>
[Sequence Length]	[SLEN]	<p>The total length of the sequence – the number of nucleotides or amino acids in the sequence. The colon (:) separates the beginning and end of a length range.</p> <p><i>Examples:</i></p> <p>755[SLEN] Nucleotide Protein EST GSS 100:1000[SLEN] Nucleotide Protein EST GSS</p>
[Substance Name]	[SUBS]	<p>The names of chemical substances associated with a record. This field is only populated for sequences extracted from structure records – PDB derived sequences. The associated residue position is often included.</p> <p><i>Examples:</i></p> <p>mg, 1010[Substance Name] Nucleotide atp[Substance Name] Protein</p>
[Text Word]	[WORD]	<p>Text on a sequence record that is not indexed in other fields. Terms indexed here are included in an All Fields search, not generally useful.</p>
[Title]	[TI] OR [TITL]	<p>Words and phrases found in the title of the sequence record. The title is the DEFINITION line of the GenBank/GenPept format of the record. This line summarizes the biology of the sequence and includes the organism, product name, gene symbol, molecule type, and sequence completeness.</p> <p>complete cds[TI] Nucleotide kinesin[TI] Nucleotide Protein liver[TI] Nucleotide Protein EST uncultured[TI] Nucleotide Protein EST GSS</p>

Table 1. continued from previous page.

Search Field	Short Field Specifier	Definition
	Sequences_help_apppe.TF.2	
[Volume]	[VOL]	Contains the volume number of the journals in references on the sequence record, not generally useful in the sequence databases.

† Queries using any term followed by the full name of the indexed field in square brackets will only retrieve records with the term indexed in that field. For example a search with apolipoprotein[Title] finds only records with “apolipoprotein” indexed for their Title field. Some fields have shorter names that can also be used instead of the full name. These are listed in the **Abbreviated Field Specifier** column of Table 1 when available.

Table 2. Fields available only for EST and GSS databases.

Index Search Field	Description
[Clone ID]	The clone identifier provided by the submitter of the EST or GSS records. <i>Example:</i> image 1000232[Clone ID] EST ZMMBBb0001G04f[Clone ID] GSS
[EST Name] [GSS Name]	The name given to the EST or GSS record by the submitter. <i>Examples:</i> R-OVA-119[EST Name] EST DKFZP761J17121[GSS Name] GSS
[EST ID] [GSS ID]	Legacy dbEST or dbGSS unique identifier provided by NCBI. <i>Examples:</i> 2081316[EST ID] EST 14283478[GSS ID] GSS
[Library Class] (GSS Only)	Information about the kind of genomic DNA library that was the source of the clone. <u>Examples:</u> bac ends[Library Class] GSS methylation filtered [Library Class] GSS cosmid ends[Library Class] GSS shotgun[Library Class] GSS
[Library Name] (EST Only)	The name given to the cDNA library that is the source of the clone, provided by the submitter and taken verbatim from the record. May contain useful information about the cell, tissue, or organ source. <i>Examples:</i> soares fetal liver spleen 1nfls[Library Name] EST full length enriched swine cdna library, adult adrenal gland[Library Name] EST
[Submitter Name]	Submitter name of EST and GSS records. Unlike [Author Name], the Submitter Name content is not controlled and is verbatim from the EST or GSS record <i>Examples:</i> smith tp1[Submitter Name] EST GSS david severson[Submitter Name] EST da lightfoot and chris town[Submitter Name] GSS