A Field Guide

part 2

- Entrez searching
- Genome resources
- Sequence similarity

January 30, 2007
Washington University, St. Louis
### GenBank Record: Feature Table, con't.

<table>
<thead>
<tr>
<th>Number</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>211.1961</td>
<td>gene</td>
<td>1</td>
</tr>
</tbody>
</table>

### GenBank Record: sequence

```
000018831
```

### Indexing for Nucleotide UID 59958365

<table>
<thead>
<tr>
<th>Field</th>
<th>Indexed Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>primary accession</td>
<td>NM_001012399 [acces]</td>
</tr>
<tr>
<td>title</td>
<td>Bos taurus hemochromatosis (hfe), mRNA.</td>
</tr>
<tr>
<td>(organism)</td>
<td>1168 [orgn]</td>
</tr>
<tr>
<td>(sequence length)</td>
<td>2005/02/19 [mdat]</td>
</tr>
<tr>
<td>(modification date)</td>
<td>2006/01/2006/08[mdat]</td>
</tr>
<tr>
<td>(properties)</td>
<td>gbd inv mam srcb refseq</td>
</tr>
</tbody>
</table>

### Global Entrez Search: HFE

- UniGene: gene-centered clusters of transport sequences
- GOA: conserved protein domain databases
- 3D Domains: domain from Entrez Genome
- UniRef50: markers and mapping data
- PopSet: population study data sets
- GEO Profiles: expression and nucleotide abundance profiles
- GEO Datasets: experimental sets of GEO data
- Cancer Chromosomes: cytogenetic databases
- PubChem Compound: small molecule chemical structures
- PubChem Substance: chemical substances screened for bioactivity
- Genomes Project: genome project information
hfe\textit{[title]} AND human\textit{[orgn]} (con't)

- **NM_001012399**
  Homo sapiens hemochromatosis (HFE), mRNA
  gi|69590335|ref|NM_001012399.1|f|69590385|

- **NM_0023032**
  Homo sapiens hemochromatosis (HFE), complete cds
  gi|55759117|ref|NM_0023032.1|||55759117|

- **NM_139011**
  Homo sapiens hemochromatosis (HFE), transcript variant 11, mRNA
  gi|21040356|ref|NM_139011.1||21040356|

- **NM_139010**
  Homo sapiens hemochromatosis (HFE), transcript variant 10, mRNA
  gi|21040354|ref|NM_139010.1||21040354|

---

**Preview/Index**

*Gateway to Advanced Searches*

- **1**: NM_139011
  Homo sapiens hemochromatosis (HFE), transcript variant 11, mRNA
  gi|21040356|ref|NM_139011.1||21040356|

- **2**: NM_139010
  Homo sapiens hemochromatosis (HFE), transcript variant 10, mRNA
  gi|21040354|ref|NM_139010.1||21040354|

- **3**: NM_139009
  Homo sapiens hemochromatosis (HFE), transcript variant 9, mRNA
  gi|21040352|ref|NM_139009.1||21040352|

- **4**: NM_139008
  Homo sapiens hemochromatosis (HFE), transcript variant 8, mRNA
  gi|21040350|ref|NM_139008.1||21040350|

- **5**: NM_139007
  Homo sapiens hemochromatosis (HFE), transcript variant 7, mRNA
  gi|21040349|ref|NM_139007.1||21040349|
'Properties' Search Field

#1 hfe 137
#2 hfe [title] AND human [orgn] 42
#3 #2 AND srcdb refseq [prop] 11
#4 #2 AND srcdb ddbj/embl/genbank [prop] 31
#5 #4 AND gbdiv pri [prop] 29
#4 #4 AND gbdiv est [prop] 2

Primate division gbdiv pri [prop]
EST division gbdiv est [prop]

'Properties' Search Field: biomol

#1 hfe 116
#2 hfe [title] AND human [orgn] 42
#3 #2 AND biol mol mrna [prop] 29
#4 #2 AND biol mol genomic [prop] 13

Genomic DNA biol mol genomic [prop]
cDNA biol mol mrna [prop]

More Queries…

Entrez Nucleotide

Reviewed RefSeqs with transcript variants:
srcdb refseq reviewed [prop] AND transcript [title] AND variant [title]

Entrez Gene

Topoisomerase genes from Archaea:
topoisorase [gene name] AND archaea [organism]

Genes on human chromosome 2 with OMIM links
2 [chromosome] AND human [organism] AND "gene omim" [filter]

Membrane proteins linked to cancer:
"integral to plasma membrane" [gene ontology] AND cancer [dis]
Other Entrez Databases

**UniGene**: rat clusters that have at least one mRNA
rat(organism) **NOT** @mRNA count

**SNP**: uniquely mapped microsatellites on human chr2
microsat(SNP Class) AND 1(Map Weight) AND 2(Chromosome) AND human(organ)

**UniSTS**: markers on the Genethon map of human chromosome 12
Genethon(Map Name) AND human(organism) AND 12(chromosome)

**Structure**: structures of bacterial kinases with resolutions below 2 Å
bacteria(organism) AND kinase AND 0.00:0.002.00(resolution)
Map Viewer – Genome Annotation Updates

<table>
<thead>
<tr>
<th>Species</th>
<th>Build</th>
<th>Map Viewer Release</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bos taurus (cows)</td>
<td>3.1</td>
<td>January 3, 2007</td>
</tr>
<tr>
<td>Gallus gallus (chickens)</td>
<td>2.1</td>
<td>November 30, 2006</td>
</tr>
<tr>
<td>Oryctolagus cuniculus</td>
<td>2.1</td>
<td>November 21, 2006</td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
<td></td>
<td>November 2, 2006</td>
</tr>
<tr>
<td>Dictyostelium discoideum</td>
<td></td>
<td>October 18, 2006</td>
</tr>
<tr>
<td>Strongylocentrotus purpuratus</td>
<td></td>
<td>October 5, 2006</td>
</tr>
<tr>
<td>Pan troglodytes</td>
<td></td>
<td>September 14, 2006</td>
</tr>
<tr>
<td>Homo sapiens (human)</td>
<td></td>
<td>Amel_4.0</td>
</tr>
<tr>
<td>Apis mellifera (bee)</td>
<td></td>
<td>August 11, 2006</td>
</tr>
<tr>
<td>Rattus norvegicus (rat)</td>
<td>RGSC</td>
<td>July 6, 2006</td>
</tr>
<tr>
<td>Macaca mulatta (chinese macaque)</td>
<td>1.1</td>
<td>June 23, 2000</td>
</tr>
</tbody>
</table>

Genome Resources

- Entrez Genome
- Fungal Genomes Central
- Genome Projects Database
  - Eukaryotic Genome Resources
  - Fungi Resources
  - Invertebrates Resources
  - Viral Resources
  - Retroviruses Resources
  - Viral Genomes Resources

Genome Project

13 Eukaryotic Genome Sequencing Projects Selected: Complete - 0, Assembly - 2, In Progress - 11

<table>
<thead>
<tr>
<th>Organism</th>
<th>Group</th>
<th>Subgroup</th>
<th>TaxID</th>
<th>Genome Size (Mb)</th>
<th># Chrs</th>
<th>Status</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aphasia californica</td>
<td>Animals</td>
<td>Other Animals</td>
<td>63900</td>
<td>1800</td>
<td>17</td>
<td>In Progress</td>
<td>WGS</td>
</tr>
<tr>
<td>Bombyx mori</td>
<td>Animals</td>
<td>Other Animals</td>
<td>6526</td>
<td></td>
<td></td>
<td>In Progress</td>
<td>WGS</td>
</tr>
<tr>
<td>Ciona intestinalis</td>
<td>Animals</td>
<td>Other Animals</td>
<td>2719</td>
<td>155</td>
<td>14</td>
<td>Assembly</td>
<td>WGS</td>
</tr>
<tr>
<td>Ciona savignyi</td>
<td>Animals</td>
<td>Other Animals</td>
<td>51311</td>
<td>180</td>
<td></td>
<td>Assembly</td>
<td>WGS</td>
</tr>
<tr>
<td>Daphnia pulex</td>
<td>Animals</td>
<td>Other Animals</td>
<td>6669</td>
<td></td>
<td></td>
<td>In Progress</td>
<td>WGS</td>
</tr>
<tr>
<td>Hydra</td>
<td>Animals</td>
<td>Other Animals</td>
<td>6083</td>
<td></td>
<td></td>
<td>In Progress</td>
<td>WGS</td>
</tr>
<tr>
<td>Nematostella vectensis</td>
<td>Animals</td>
<td>Other Animals</td>
<td>45351</td>
<td></td>
<td></td>
<td>In Progress</td>
<td>WGS</td>
</tr>
</tbody>
</table>
Homologene

- No longer UniGene based
- Protein similarities first
- Guided by taxonomic tree
- Includes orthologs and paralogs

Orthologs

Gene duplication

A-chain gene

Early globin

Rice Homolog

Genome Resources

Genomic Biology

Homologene

Map Viewer

Entrez Gene
MapViewer: Mouse ADAR, 28 Hits

Mus musculus (house mouse) genome view

Search results for query "adar": 28 hits
Mouse MapViewer: Gene Filter

Search results for query: "adar AND (gene<object_type>)": 4 hits

<table>
<thead>
<tr>
<th>Chr</th>
<th>Assembly</th>
<th>Match</th>
<th>Map element</th>
<th>Type</th>
<th>Map</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>reference</td>
<td>all matches</td>
<td>Adar: double-stranded RNA-specific adenine deaminase</td>
<td>Adar</td>
<td>GENE</td>
</tr>
<tr>
<td>3</td>
<td>Cebna</td>
<td>Adar</td>
<td>Adar: double-stranded RNA-specific adenine deaminase</td>
<td>Adar</td>
<td>GENE</td>
</tr>
<tr>
<td>3</td>
<td>MGSCv3</td>
<td>Adar</td>
<td>Adar: double-stranded RNA-specific adenine deaminase</td>
<td>Adar</td>
<td>GENE</td>
</tr>
</tbody>
</table>

Maps & Options

- Sequence maps
- Gene maps
- Cytogenetic maps
- Idogram
- MapViewer
- RefSeq RNA
- SNPs

Gene annotations
Tiling path
Variations

Mouse MapViewer Build 36.1 (Current)
BLAST The Mouse Genome

exon
3' UTR
Basic Local Alignment Search Tool

Web BLAST
- pre-computed results
- how BLAST works
  - words; scoring matrices; statistics
- specialized BLAST algorithms
- what's new, or important
- example oligo search

BLAST Web Searches, 2006

BLAST Web Searches, 2005
Precomputed BLAST Services

- Nucleotide or protein: Related Sequences
- BLAST link: Blink
- Transcript clusters: UniGene
- Protein homologs: Homologene

Link to Related Sequences

BLink (BLAST Link)

Related Sequences

<table>
<thead>
<tr>
<th>Rank</th>
<th>Species</th>
<th>Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552914</td>
<td>Single small insert library clone 323188</td>
</tr>
<tr>
<td>2</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552915</td>
<td>Single small insert library clone 323444</td>
</tr>
<tr>
<td>3</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552916</td>
<td>Single small insert library clone 323445</td>
</tr>
<tr>
<td>4</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552917</td>
<td>Single small insert library clone 323446</td>
</tr>
<tr>
<td>5</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552918</td>
<td>Single small insert library clone 323447</td>
</tr>
<tr>
<td>6</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552919</td>
<td>Single small insert library clone 323448</td>
</tr>
<tr>
<td>7</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552920</td>
<td>Single small insert library clone 323449</td>
</tr>
<tr>
<td>8</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552921</td>
<td>Single small insert library clone 323450</td>
</tr>
<tr>
<td>9</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552922</td>
<td>Single small insert library clone 323451</td>
</tr>
<tr>
<td>10</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552923</td>
<td>Single small insert library clone 323452</td>
</tr>
<tr>
<td>11</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552924</td>
<td>Single small insert library clone 323453</td>
</tr>
<tr>
<td>12</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552925</td>
<td>Single small insert library clone 323454</td>
</tr>
<tr>
<td>13</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552926</td>
<td>Single small insert library clone 323455</td>
</tr>
<tr>
<td>14</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552927</td>
<td>Single small insert library clone 323456</td>
</tr>
<tr>
<td>15</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552928</td>
<td>Single small insert library clone 323457</td>
</tr>
<tr>
<td>16</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552929</td>
<td>Single small insert library clone 323458</td>
</tr>
<tr>
<td>17</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552930</td>
<td>Single small insert library clone 323459</td>
</tr>
<tr>
<td>18</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552931</td>
<td>Single small insert library clone 323460</td>
</tr>
<tr>
<td>19</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552932</td>
<td>Single small insert library clone 323461</td>
</tr>
<tr>
<td>20</td>
<td>Drosophila melanogaster (Dm), eDNA</td>
<td>EMB552933</td>
<td>Single small insert library clone 323462</td>
</tr>
</tbody>
</table>

Most similar

Least similar
To evaluate evolutionary relationships
To identify and annotate sequences
Other:
  - model genomic structure (e.g., Splign)
  - check primer specificity in silico
Global vs Local Alignment

Seq1: WHEREISWALTERNOW (16aa)
Seq2: HEWASHEREBUTNOWISHERE (21aa)

Global

Seq1: 1 W--HEREISWALTERNOW 16
Seq2: 1 HERE IS HERE 21

Local

Seq1: 1 W--HERE 5
Seq2: 3 WASHERE 9

How BLAST Works

1. Make lookup table of “words” for query
2. Scan database for hits
3. Extend alignment both directions
   - Ungapped extensions of hits (initial HSPs)
   - Gapped extensions (no traceback)
   - Gapped extensions (traceback – alignment details)

Nucleotide Words

Make a lookup table based on the word size.

11-mer

ATGCTGCTAGT CGATGACGTAGCTA
ATGCTGCTAGT
TGCTGCTAGTC
GCTGCTAGTC

Protein Words

AIEKCYTGCTLAQEADDTA
AIEKCYT
LEK, IDK, IQK, IER, IDR, etc
EKC
KCY
CYT
...

Lookup table, including neighborhood words, is based on word size, score matrix, and threshold.
Scoring Systems - Nucleotides

Identity matrix

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>G</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>+1</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>G</td>
<td>-3</td>
<td>+1</td>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>C</td>
<td>-3</td>
<td>-3</td>
<td>+1</td>
<td>-3</td>
</tr>
<tr>
<td>T</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
<td>+1</td>
</tr>
</tbody>
</table>

\[ -r \ 1 \ -q \ -3 \]

CAGGTAGCAAGCTTGCATGTCA || ||||||||||||  ||||| raw score = 19 - 9\* = 10\*
CACGTAGCAAGCTTG-GTGTCA

* ignores gap costs

Scoring Systems - Proteins (BLOSUM62)

A  R  N  D  C  Q  E  G  H  I  L  K  M  F  P  S  T  W  Y  V  X

\[ (\text{leucine or isoleucine} \) \] \[ \text{D (pyrolysine)} \]

Local Alignment Statistics

**Expect Value**

\[ E = \text{number of database hits you expect to find by chance, > 5} \]

\[ E = K m n e^{\lambda S} \quad \text{or} \quad E = mn^{2-5} \]

K = scale for search space
\( \lambda S = \text{scale for scoring system} \)
\( S = \text{bitscore} = 0.5 \times \ln K/\ln 2 \)
\( m = \text{query length} \)
\( n = \text{database length} \)

E is dependent on m \( \times \) n (search space)

More info: The Statistics of Sequence Similarity Scores

E is dependent on m \( \times \) n (search space)

More info: The Statistics of Sequence Similarity Scores
Scoring Systems - Proteins

Position Independent Matrices

PAM Matrices (Percent Accepted Mutation)
- Derived from observation; small dataset of alignments
- Implicit model of evolution
- All calculated from PAM1
- PAM250 widely used

BLOSUM Matrices (BLOck SUBstitution Matrices)
- Derived from observation; large dataset of highly conserved blocks
- Each matrix derived separately from blocks with a defined percent identity cutoff
- BLOSUM62 - default matrix for BLAST

Position-Specific Score Matrix

Serine/Threonine protein kinases catalytic loop

<table>
<thead>
<tr>
<th>PSSM scores</th>
<th>1</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>4</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAF-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Position-Specific Score Matrix

An alignment BLAST cannot make:

<table>
<thead>
<tr>
<th>Reason:</th>
</tr>
</thead>
<tbody>
<tr>
<td>no contiguous exact match of 7 bp.</td>
</tr>
</tbody>
</table>
An Alignment BLAST Can Make

Score = 290 hits (741), Expect = 7e-77
Identities = 147/331 (44%), Positives = 206/331 (61%), Gaps = 8/331 (2%)
Frame = +3

Megablast: NCBI’s Genome Annotator

- Long alignments of similar DNA sequences
- Greedy algorithm
- Concatenation of query sequences
- Faster than blastn; less sensitive

Other BLAST Algorithms

- Megablast
- Discontiguous Megablast
- PSI-BLAST
- PHI-BLAST

MegaBLAST & Word Size

<table>
<thead>
<tr>
<th>WORD SIZE</th>
<th>default</th>
<th>minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>blastn</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>megablast</td>
<td>28</td>
<td>8</td>
</tr>
<tr>
<td>blastp</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
### Word Size

**Trade-off: sensitivity vs speed**

![Image](Too fast for you?)

### Discontiguous Megablast

- Uses **discontiguous** word matches
- Better for **cross-species comparisons**

### Templates for Discontiguous Words

<table>
<thead>
<tr>
<th>W</th>
<th>t</th>
<th>coding:</th>
<th>non-coding:</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>16</td>
<td>1110101101101101</td>
<td>1111101101101101</td>
</tr>
<tr>
<td>12</td>
<td>16</td>
<td>1111010110110111</td>
<td>1111010110110111</td>
</tr>
<tr>
<td>11</td>
<td>18</td>
<td>1110101101101101</td>
<td>1110101101101101</td>
</tr>
<tr>
<td>12</td>
<td>18</td>
<td>1011011011011011</td>
<td>1010110110110111</td>
</tr>
<tr>
<td>11</td>
<td>21</td>
<td>1111010110110111</td>
<td>1111010110110111</td>
</tr>
<tr>
<td>12</td>
<td>21</td>
<td>1001101101101101</td>
<td>1001101101101101</td>
</tr>
</tbody>
</table>

W = word size, t = matches in template, coding: = matches in coding, non-coding: = matches in non-coding


<table>
<thead>
<tr>
<th>BLAST</th>
<th>Nucleotide</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Quickly search for nearly similar sequences (megablast)</td>
<td>- Protein-protein BLAST (blastp)</td>
<td>- Position-specific iterated and pattern-matched BLAST (IPS and PHM-BLAST)</td>
</tr>
<tr>
<td>- Search for short, nearly exact matches</td>
<td>- Search for short, nearly exact matches</td>
<td>- Search the conserved domain database (pfam)</td>
</tr>
<tr>
<td>- Search through archives with megablast</td>
<td>- Search through archives with megablast</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Translated</th>
<th>Genomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Translated query vs protein database ( blastx)</td>
<td>- Human, mouse, rat, chimpanzee, cow, pig, dog, sheep, cat</td>
</tr>
<tr>
<td>- Protein query vs translated database ( blastx)</td>
<td>- Chicken, buffalo, zebrafish</td>
</tr>
<tr>
<td>- Translated query vs translated database ( blastx)</td>
<td>- Environmental samples</td>
</tr>
<tr>
<td></td>
<td>- Malaria</td>
</tr>
<tr>
<td></td>
<td>- Insects, nematodes, plants, fungi, microbial genomes, other eukaryotic genomes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Special</th>
<th>Meta</th>
</tr>
</thead>
</table>
Discontiguous (Cross-species) MegaBLAST

**Search**

Load query file from disk:

Set **database**:

Choose database:

Return alignment only:

Discontiguous Word Options

**Options** for advanced mining:

- **Load by**
- **File size**
- **Word size**
- **Percent identity**
- **Percent total matches**
- **Percent total mismatches**

**Discontiguous Word Options**

- **Template length**
- **Template type**
- **Distance**
- **Distance**
- **Distance**

**Ex: Discontiguous MegaBLAST**

- **Score**: 365 bits (190), **Expect**: 3e-97
- **Identities**: 624/836 (74%), **Gaps**: 2/836 (0%)
- **Strand**: Plus/Plus

**Disco. Megablast Example . . .**

- **Query**: NM_078651
  - Drosophila melanogaster CG18582-PA (mbt) mRNA, (3244 bp)
  - **Note**: mushroom bodies tiny; synonyms: Pak2, STE20, dPAK2
- **Database**: nr (nt), Mammalia[orgn]

- > MegaBLAST = “No significant similarity found.”
- > Discontiguous megaBLAST = numerous hits . . .
Ex: BLASTN

Score = 87.7 bits (44), Expect = 9.8e-14
Identities = 74/84 (88%), Gaps = 0/84 (0%)
Strand = Plus/Plus

Nucleotide BLAST Databases

- nr (nt)
  - Traditional GenBank Divisions
    - NM_, XM_, NR_
  - refseq_rna
    - NM_, XM_, NR_
  - refseq_genomic
    - NC_, NT_, NC_
  - est
    - EST Division
  - hgs
    - HTG division
  - dbsts
    - STS Division

- chromosome
  - NC genomic records
- gss
  - GSS division
- pat
  - PAT Division
- wgs
  - wgs entries from traditional divisions
- pdb
  - Nucleotide sequences from structures
- env_nt
  - environmental samples

Protein BLAST Databases

Protein
- nr
  - traditional GenBank records
- refseq = NP_, XP_
- swissprot
- pdb
- pat
- env_nr

What's New?

Save your searches

Select biological sequences, protein families, homologies, and pathways.

Zoom in and out, change colors and fonts.
New Nucleotide Databases

New Formatter

BLAST Output: Alignments & Filter

New Output View
New Output View

- Results can be sorted
- Transcript & genomic hits separated

Sorting Results

- Resorted by Total score

Sorting Hits: by Score

- Sort by Score: longest exon usually first

Sorting Hits: by Query Start

- Sort by Query start: Proper exon order
Example: Mapping Oligos Onto a Genome

Forward primer: CCATGGCGACCCTGGAAAAGC
Reverse primer: CAGCAGCGGCTGTGCCTGCGG

Genome BLAST Results

Database: contig
498 sequences; 3,820,300,372 total letters

If you have any questions or comments about this search, please refer to the BLAST FAQs.
Primer Alignments

**reverse primer**

- Score = 42.1 Bits (11), Expect = 0.001
- Identity = 52/52 (100%)
- Strand = Plus / Plus

```
query: 52 5acaagagctcatctctgag
objects: 463105 5acaagagctcatctctgag 663205
```

**forward primer**

- Score = 42.1 Bits (11), Expect = 0.001
- Identity = 52/52 (100%)
- Strand = Plus / Plus

```
query: 1 5acaagagctcatctctgag
objects: 463105 5acaagagctcatctctgag 463140
```

MapViewer

MapViewer

Sequence View (sv)
## Service Addresses

<table>
<thead>
<tr>
<th>Service</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLAST</td>
<td><a href="mailto:blast-help@ncbi.nlm.nih.gov">blast-help@ncbi.nlm.nih.gov</a></td>
</tr>
<tr>
<td>General Help</td>
<td><a href="mailto:info@ncbi.nlm.nih.gov">info@ncbi.nlm.nih.gov</a></td>
</tr>
<tr>
<td>Wayne Matten</td>
<td><a href="mailto:matten@ncbi.nlm.nih.gov">matten@ncbi.nlm.nih.gov</a></td>
</tr>
</tbody>
</table>