

Module 2: Maps and Genome Sequence (NCBI)

iv - How Do I Find a Zebrafish Gene in the Genome?

Aims

- Introduce search tools
- Suggest methods for text or sequence searches
- Provide example text and sequence searches
- Show alternative ways to identify genomic placement of your gene or cDNA

Introduction

Locating the placement of a gene on the genome can be accomplished through sequence comparisons or by position information of the gene or other related mapped objects, such as BAC clones, SNPs, ESTs or STS markers.

To compare two sequences by BLAST, select a BLAST tool from the BLAST home page (<http://www.ncbi.nlm.nih.gov/BLAST/>) or by choosing a BLAST database from the zebrafish BLAST page (<http://www.ncbi.nlm.nih.gov/genome/seq/DrBlast.html>).

Additional related records can be viewed by following the provided Links to Assembly, WGS Project and UniSTS records from the GenBank record.

Another option to view a gene in its genomic context is to align a representative mRNA to a genomic sequence by NCBI's Splign tool which produces spliced global alignments (<http://www.ncbi.nlm.nih.gov/sutils/splign/splign.cgi>).

Exercises:

- 1. BLAST:** identifying genomic placement via sequence comparison using MegaBLAST
- 2. Zebrafish BLAST:** querying against the WGS contigs or genomic (reference) sequence
- 3. Splign:** aligning an mRNA against a genomic contig to view spliced global alignments

1. BLAST

From the BLAST home page, choose one of the BLAST tools to submit a search.

Search the trace archives with MegaBLAST to identify the most similar WGS, EST or other Danio rerio trace sequence.

NCBI → BLAST Latest news: 13 June 2005 : BLAST 2.2.11 released

About

- Getting started
- News
- FAQs

More info

- NAR 2004
- NCBI Handbook
- The Statistics of Sequence Similarity Scores

Software

- Downloads
- Developer info

Other resources


- References
- NCBI Contributors
- Mailing list
- Contact us

The **Basic Local Alignment Search Tool (BLAST)** finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

Nucleotide <ul style="list-style-type: none"> Quickly search for highly similar sequences (megablast) Quickly search for divergent sequences (discontiguous megablast) Nucleotide-nucleotide BLAST (blastn) Search for short, nearly exact matches Search trace archives with megablast or discontiguous megablast 	Protein <ul style="list-style-type: none"> Protein-protein BLAST (blastp) Position-specific iterated and pattern-hit initiated BLAST (PSI- and PHI-BLAST) Search for short, nearly exact matches Search the conserved domain database (rpsblast) Protein homology by domain architecture (cdart)
Translated <ul style="list-style-type: none"> Translated query vs. protein database (blastx) Protein query vs. translated database (tblastn) Translated query vs. translated database (tblastx) 	Genomes <ul style="list-style-type: none"> Human, mouse, rat, chimp NEW, cow, pig, dog, sheep, cat Chicken, puffer fish, zebrafish Environmental samples Malaria Insects, nematodes, plants, fungi, microbial genomes, other eukaryotic genomes
Special <ul style="list-style-type: none"> Search for gene expression data (GEO BLAST) Align two sequences (bl2seq) Screen for vector contamination (VecScreen) Immunoglobulin BLAST (IgBlast) SNP BLAST 	Meta <ul style="list-style-type: none"> Retrieve results

Organism-specific BLAST

MegaBlast



megablast **BLAST**

[Nucleotide](#) [Protein](#) [Translations](#) [Retrieve results for an RID](#)

Trace Archive database Mega BLAST search

[Search](#)

NM_001024735

Load query file from disk

Browse...

[Set subsequence](#)

From: To:

[Choose database](#)

Danio rerio - WGS

[Return alignment endpoints only](#)

☐

Now:

BLAST!

or

Reset query

Reset all

Options for advanced blasting

[Hits computed](#) 250

[Choose filter](#) ☒ Low complexity ☒ Rodent Repeats ☒ Mask for lookup table only ☐ Mask lower case

Choose a Query sequence

Choose a database from the pull down menu

Submit for BLAST

BLAST Results:



results of BLAST

BLASTN 2.2.10 [Oct-19-2004]

RID: 1119441055-19332-153186667233.BLASTQ2

Database: Danio_rerio_WGS

22,634,736 sequences; 21,084,232,232 total letters

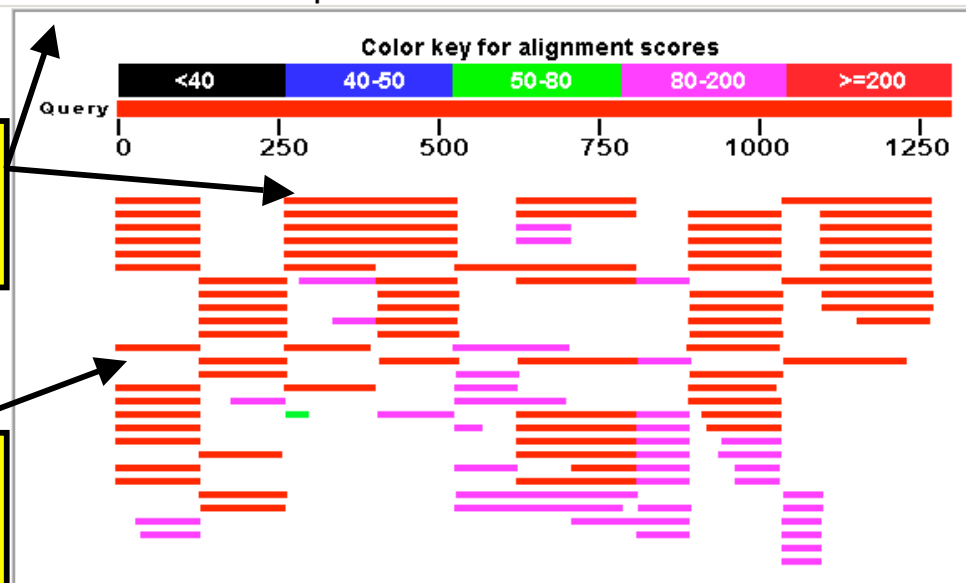
If you have any problems or questions with the results of this search please refer to the [BLAST FAQs](#)

[Taxonomy reports](#)

Query= gi|67078407|ref|NM_001024735.1| Danio rerio zgc:63996 (zgc:63996), mRNA
(1293 letters)

Distribution of 139 Blast Hits on the Query Sequence

611790792 zDH86-629e15.p1k S=234 E=7.1e-58



Mouse-over
display to view
Trace id

Click on any bar
to jump to the
alignment

>[gnl|ti|611790792](#) zDH86-629e15.p1k
 Length=846
 Score = 272 bits (137), Expect = 3e-69
 Identities = 140/141 (99%), Gaps = 0/141 (0%)
 Strand=Plus/Minus

Query 264 AGATGTGGAGAGTGTGATGAACAGCATCGTGTCTCTGCTGCTGATTCTGGAGACGGAGAA 323
 |||
 Sbjct 457 AGATGTGGAGAGTGTGATGAACAGCATCGTGTCTCTGCTGCTGATTCTGGAGACGGAGAA 398



Query 324 GCAGGAGGCTCTTATTGAAAGCTTATGTGAGAAGCTGGTGAAGTTTCGTGAGGGTGAACG 383
 |||
 Sbjct 397 GCAGGAGGCTCTGATTGAAAGCTTATGTGAGAAGCTGGTGAAGTTTCGTGAGGGTGAACG 338

Query 384 GCCCTCGCTTCGGATGCAGCT 404
 |||
 Sbjct 337 GCCCTCGCTTCGGATGCAGCT 317

Click on this link to go to the trace archive record

Trace Archive for ti:611790792

(<http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?cmd=retrieve&dopt=fasta&val=611790792>)



Trace Archive

[Home](#)
[Search](#)
[Site Map](#)
[Main](#)
[Statistics](#)
[Tracking System](#)
[Obtaining Data](#)
[BLAST](#)
[Documents](#)

[Searching Tips](#)
[Searchable Fields](#)
[Registered Species](#)
[Submitting Centers](#)
[FTP](#)

☒ [Style](#): [Blue Sky](#)

☒ [.gz file](#)

☐ [All](#)
☒ [FASTA](#)
☐ [Quality](#)
☐ [SCF](#)
☐ [Info-XML](#)
☐ [Info-Table](#)
☐ [Mate Pair](#)

☐ [in color](#)
☐ [NULL values](#)

Search result:

Your request is: 611790792

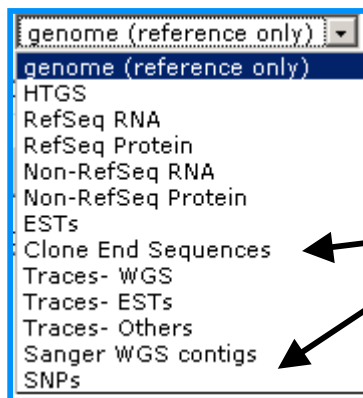
```

>gnl|ti|611790792 name:zDH86-629e15.plk mate:611686865 mate_name:zDH86-629e15.q1k template:zDH86-629e15
CAGCTCGGTACCCTTCATTTTATTTCATGATTAACCAACAAAGACTATATACCTTCATATTTTCAAACAC
AGTTTAAATAAAATGACGGCGGTTTCTATACCTGGTCTAGGTCGTGATGGCATGAAGGTGATGGCGTTGCAG
GTTGCCGCCACTTTAATCAGACTGCAGTACACAGTGTGTCTCACTGGAGTGTCTCGTCCATACCATGGA
ACAGATTACTCAAACCTAAGAGAAGAAATACGAGTTCATACAAAGCAAAACATACAAAATTAATACAAAGGA
AATTAACAAAAAAGCAAAACAGTCTAGACACTCACAGCTGCATCCGAAGCGAGGGGCCGTTACCCCTCAC
GAAACTTCACCGAGCTTCTCACATAAAGCTTTCAATCAGAGCCTCCTGCTTCTCCGCTCCAGAAATCAGCAG
CAGAGACACGATGCTGTTTCATCACACTCTCCACATCTGCAGACATCAGCAGAGGGGGACAAAAAAGGTT
ATTGTGTATCCGATCTATAAAGAGAAGCTCTGATTAGATTTCGCGAGATATCAGTATTGGTCAATAATAA
AAAAATATTTCTTGACTAACAGATGTAGGACAAAAGGGGACACCAAAATAAGTTCAAGTAAAGGGGCCAAAG
AGGGCTTCAGAAATGCAATTAGTAATGATTTTAGGAAAAGAGGTTCTAGATTTTCTGTTCTGCAGTT
AAGGTTAAACACATAACAAATGATCCGCTGATTTGTGTTTGTCTAATCCTTTTACATTTGTTTCATTGAG
CAGAATCGTTTATACTATAAACTAAAAACCAGAGATCACAAGCAGCATTATATTGACCTTGTTCATAAA
ATATGA
    
```

2. Zebrafish BLAST

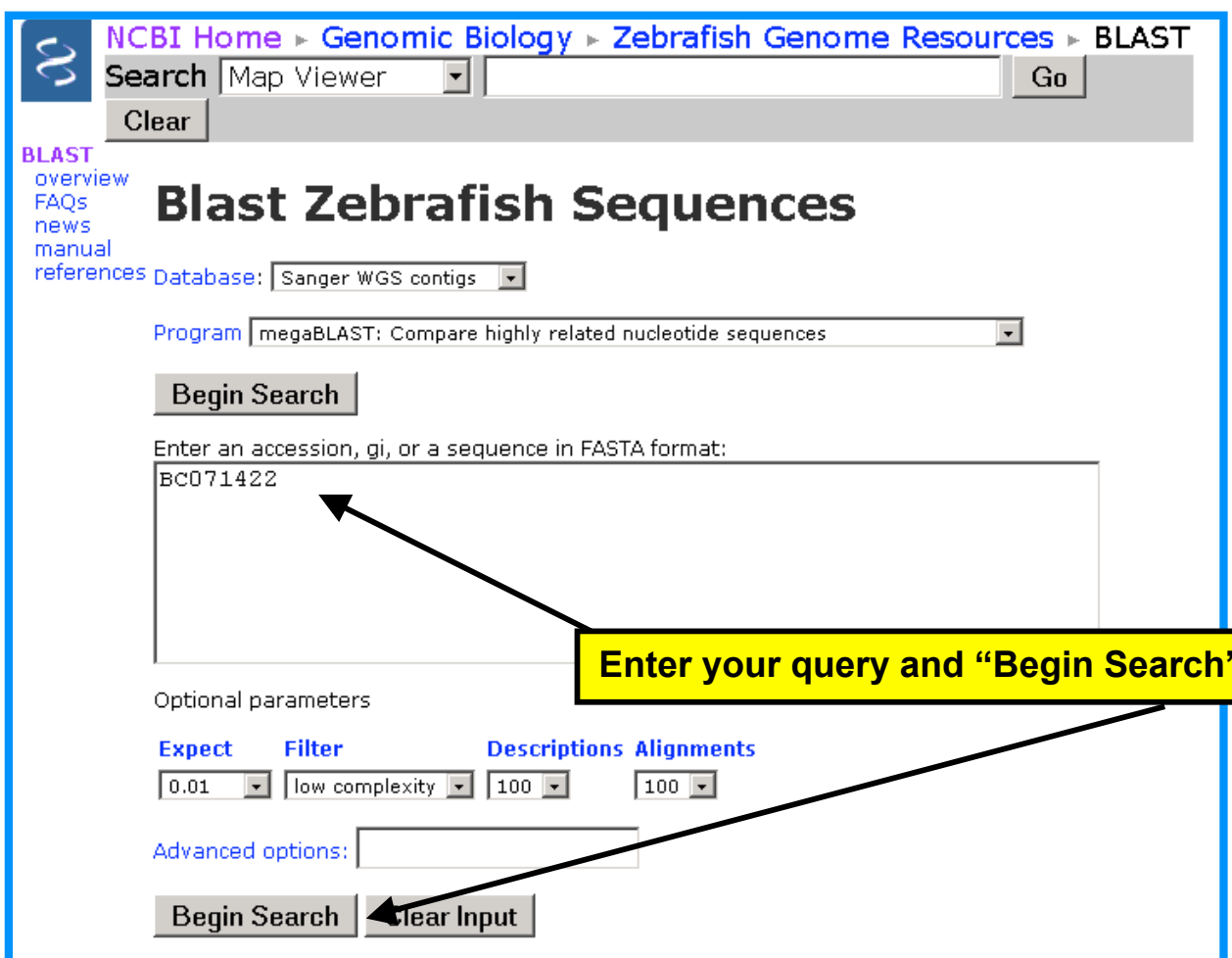
The Zebrafish BLAST page allows you to submit a query sequence against zebrafish-specific databases.

(<http://www.ncbi.nlm.nih.gov/genome/seq/DrBlast.html>)



A screenshot of a dropdown menu for selecting a sequence database. The menu is open, showing a list of options. The first option, 'genome (reference only)', is highlighted in blue. Below it are several other options: HTGS, RefSeq RNA, RefSeq Protein, Non-RefSeq RNA, Non-RefSeq Protein, ESTs, Clone End Sequences, Traces- WGS, Traces- ESTs, Traces- Others, Sanger WGS contigs, and SNPs. Two black arrows point from a yellow text box to the 'Clone End Sequences' and 'Sanger WGS contigs' options.

**Choose a sequence database:
BLAST your ZGC clone against the
Zebrafish Clone End Sequences or
the Sanger WGS contigs**



A screenshot of the NCBI Blast Zebrafish Sequences web interface. The page has a blue header with the NCBI logo and navigation links: 'NCBI Home', 'Genomic Biology', 'Zebrafish Genome Resources', and 'BLAST'. Below the header is a search bar with a 'Map Viewer' dropdown and a 'Go' button. A 'Clear' button is also present. On the left, there are links for 'BLAST overview', 'FAQs', 'news', 'manual', and 'references'. The main heading is 'Blast Zebrafish Sequences'. Below this, there is a 'Database:' dropdown set to 'Sanger WGS contigs' and a 'Program:' dropdown set to 'megaBLAST: Compare highly related nucleotide sequences'. A 'Begin Search' button is located below the program dropdown. A text input field contains the sequence 'BC071422'. Below the input field, there are 'Optional parameters' for 'Expect' (0.01), 'Filter' (low complexity), 'Descriptions' (100), and 'Alignments' (100). There is an 'Advanced options:' text input field. At the bottom, there are 'Begin Search' and 'Clear Input' buttons. A yellow text box with a black border contains the text 'Enter your query and "Begin Search"', with two black arrows pointing to the sequence input field and the 'Begin Search' button.

BLAST Results



BLASTN 2.2.10 [Oct-19-2004]

RID: 1119444995-16813-41293302633.BLASTQ4

Database: sanger_wgs_contigs
21,333 sequences; 1,560,480,686 total letters

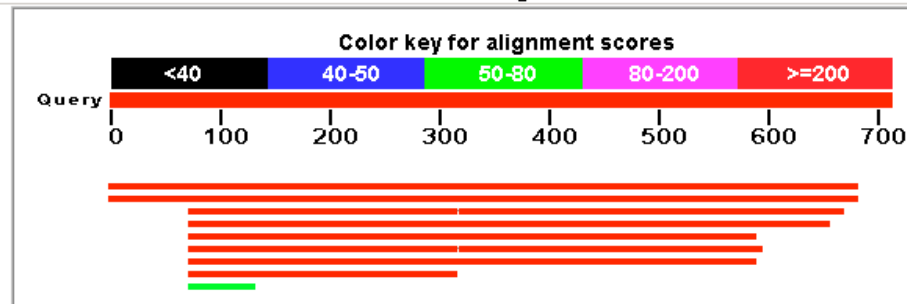
If you have any problems or questions with the results of this search please refer to the [BLAST FAQs](#)

[Taxonomy reports](#)

Query= gi|47937949|gb|BC071422.1| Danio rerio zgc:86750, mRNA (cDNA clone MGC:86750 IMAGE:6899054), complete cds
(714 letters)

Distribution of 34 Blast Hits on the Query Sequence

Mouse over to see the define, click to show alignments




Sequences producing significant alignments:				Score (Bits)	E Value
emb CAAK01000032.1 	Danio rerio whole genome shotgun, scaffol...			<u>700</u>	0.0
emb CAAK01019643.1 	Danio rerio whole genome shotgun, scaffol...			<u>700</u>	0.0
emb CAAK01000072.1 	Danio rerio whole genome shotgun, scaffol...			<u>517</u>	1e-144
emb CAAK01000709.1 	Danio rerio whole genome shotgun, scaffol...			<u>469</u>	1e-130

Click to view GenBank
record

Click on Score to view
the alignment

Alignments:

Alignments

>[emb|CAAK01000032.1](#)  Danio rerio whole genome
genome shotgun sequence
Length=1718556

Score = 700 bits (364), Expect = 0.0
Identities = 366/367 (99%), Gaps = 0/367 (0%)
Strand=Plus/Plus

Query	318	GTACAGGGGATCCTACAGAATGAGGATCTACGAGAGGGACAACCTTCATGGGTCAGATGTA	377
Sbjct	205834	GTACAGAGGATCCTACAGAATGAGGATCTACGAGAGGGACAACCTTCATGGGTCAGATGTA	205893
Query	378	CGAGATGATGGATGACTGTGACAACATCATGAACCGTTACCGCATGTCTCACTGCCAGTC	437
Sbjct	205894	CGAGATGATGGATGACTGTGACAACATCATGAACCGTTACCGCATGTCTCACTGCCAGTC	205953
Query	438	CTGTCATGTGATGGATGGCCACTGGCTCTTTTATGACCAGCCCAACTACAGAGGCAGGAT	497
Sbjct	205954	CTGTCATGTGATGGATGGCCACTGGCTCTTTTATGACCAGCCCAACTACAGAGGCAGGAT	206013
Query	498	GTGGCACTTCGGGCCTGGGCAGTACAGGAACCTTCAGCAATTATGGTGGCATGAGATTTCAT	557
Sbjct	206014	GTGGCACTTCGGGCCTGGGCAGTACAGGAACCTTCAGCAATTATGGTGGCATGAGATTTCAT	206073
Query	558	GAGCATGAGGCGCATCATGGACTCTTGGTACTAGAATTTATTTGAATAAAAAATACTTCTC	617
Sbjct	206074	GAGCATGAGGCGCATCATGGACTCTTGGTACTAGAATTTATTTGAATAAAAAATACTTCTC	206133
Query	618	TAAGATATTAAACATTGTCTTGAATATAATTAATGCCACTAACAATAAAAAACAATATCCA	677
Sbjct	206134	TAAGATATTAAACATTGTCTTGAATATAATTAATGCCACTAACAATAAAAAACAATATCCA	206193
Query	678	CAAATAC 684	
Sbjct	206194	CAAATAC 206200	

**Click to view the GenBank
record and view related
Assembly, WGS Project and**

**Click on the Links pull-down menu to
follow the links to the related Assembly,
WGS Project and UniSTS records**

NCBI Nucleotide search results for **CAAK01000032**.

Search: Nucleotide for [] Go Clear

Display: GenBank Send all to file

Range: from begin to end ☐ Reverse complemented strand **Features:** ☐ SNP ☐

1: CAAK01000032 Reports **Danio rerio whole...[gi:58293633]**

LOCUS CAAK01000032 1718556 bp DNA linear VRT 19-JAN-2005

DEFINITION Danio rerio whole genome shotgun, scaffold Zv4_scaffold32, whole genome shotgun sequence.

ACCESSION CAAK01000032 [CAAK01000000](#)

VERSION CAAK01000032.1 GI:58293633

KEYWORDS WGS.

SOURCE Danio rerio (zebrafish)

ORGANISM [Danio rerio](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.

REFERENCE 1

AUTHORS Jekosch, K., Caccamo, M., Ning, Z., Humphray, S., Scott, C., Barlow, K., Bradley, A., Burton, J., Clark, R., Elliot, D., Grafham, D., Hunt, A., Jones, M., Lloyd, D., Lloyd, C., Matthews, L., McLaren, S., McLay, K., Oliver, K., Palmer, S., Plumb, R., Quail, M., Riddle, C., Shownkeen, R., Sims, S., Threadgold, G., Willey, D., Windsor, C., Hubbard, T., Beck, S. and Rogers, J.

Links

- Assembly
- WGS Project
- Taxonomy
- UniSTS

Assembly: WGS Assembly Chromosome 1 Contig

NCBI Nucleotide

Search Nucleotide for [] Go Clear

Limits Preview/Index History Clipboard Details

Display GenBank Send all to file

Range: from begin to end ☐ Reverse complemented strand Features:

☐ SNP ☐ CDD ☒ MGC ☐ HPRD ☐ STS

☐ 1: [NW_633982](#). Reports ...[gi:67044306]

[Click here to see all features and the sequence of this contig record.](#)

LOCUS NW_633982 1718556 bp DNA linear CON 08-JUN-2005

DEFINITION Danio rerio chromosome 1 genomic contig.

ACCESSION NW_633982

VERSION NW_633982.1 GI:67044306

KEYWORDS .

SOURCE Danio rerio (zebrafish)

ORGANISM [Danio rerio](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.

COMMENT GENOME ANNOTATION [REFSEQ](#): NCBI contigs are derived from assembled genomic sequence data.
Also see:
[Documentation](#) of NCBI's Annotation Process

The DNA sequence is from the genome assembly released by the Wellcome Trust Sanger Institute as Zv4, 12 July 2004 (see http://www.sanger.ac.uk/Projects/D_rerio/Zv4/assembly_information.shtml).

FEATURES

source Location/Qualifiers
1..1718556
/organism="Danio rerio"
/mol_type="genomic DNA"
/strain="Tuebingen"
/db_xref="taxon:7955"
/chromosome="1"

CONTIG join([BX119979.9](#):1..194023, [CAAK01000032.1](#):194024..649943, complement([BX511095.5](#):39775..154840), [CAAK01000032.1](#):765010..828010, [BX537295.6](#):1..183025, [CAAK01000032.1](#):1011036..1152097, [BX247884.6](#):1..108530, [CAAK01000032.1](#):1260628..1495069, complement([BX323083.6](#):56674..106933), [CAAK01000032.1](#):1545330..1546329, [BX119983.4](#):1..172227)

//

View NCBI annotation documentation

View the assembly instructions for the Chromosome 1 contig and follow links to clone and WGS sequences

UniSTS: view markers placed on the genomic contig by ePCR

NCBI **UniSTS** Integrating Markers and Maps

PubMed All Databases BLAST OMIM Taxonomy Structure

Search UniSTS for [] Go Clear

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Send to

All: 88 in_Gene: 43 Mapped: 5 OMIM: 0 with_SNP: 2

Items 1 - 20 of 88 Page 1 of 5 Next

1: [UniSTS:190496](#) [Links](#)
fb36h10.x1
Danio rerio locus
Found by e-PCR in sequences from *Canis familiaris*, *Danio rerio*, *Mus musculus* and *Rattus norvegicus*.

2: [UniSTS:202319](#) [Links](#)
fa11f10.s1
Danio rerio locus wdhd1
Found by e-PCR in sequences from *Danio rerio*.

3: [UniSTS:193148](#) [Links](#)
fc23d08.x1
Danio rerio multiple loci
Found by e-PCR in sequences from *Danio rerio*, *Homo sapiens*, *Mus musculus* and *Rattus norvegicus*.

About Entrez

Entrez UniSTS

Help

Query tips

Submit

Submit map

FTP site

Statistics

Related sites

e-PCR

Map Viewer

Gene

UniGene

dbSNP

GeneMap'99

GDB

MGI

RGD

RHdb

Click on the UniSTS id to view the e-PCR results and links to Gene and MapViewer

Go to UniSTS for fa11f10.s1

UniSTS: (<http://www.ncbi.nlm.nih.gov/genome/sts/sts.cgi?uid=202319>)

Entrez UniSTS

[Help](#)

[Query tips](#)

[Submit](#)

[Submit map](#)

[FTP site](#)

[Statistics](#)

Related sites

[e-PCR](#)

[Map Viewer](#)

[Gene](#)

[UniGene](#)

[dbSNP](#)

[GeneMap'99](#)

[RHdb](#)

[GDB](#)

[MGD](#)

[ZFIN](#)

Genomic biology

[Bos taurus](#)

[Canis familiaris](#)

[Danio rerio](#)

[Homo sapiens](#)

[Mus musculus](#)

[Rattus norvegicus](#)

[Sus scrofa](#)

UniSTS:202319
[Links](#)

fa11f10.s1

Primer Information ?

Forward primer:	GACTACTGGGTT CAGAAATGGG
Reverse primer:	TTGAAAGCAGTT CCTGTCTCGC
PCR product size:	121 (bp), <i>Danio rerio</i>

Danio rerio

Name:	fa11f10.s1
Also known as:	FA11F10.S1

Cross References ?

Gene	GeneID:	322945	
	Symbol:	wdhd1	
	Description:	WD repeat and HMG-box DNA binding protein 1	
	Position:		
UniGene	Dr.33169	WD repeat and HMG-box DNA binding protein 1	
	Dr.425	Zgc:92585	

Electronic PCR results ?

ESTs (4)

AA494843.1	114 .. 234
BI980695.1	119 .. 239
BM184900.1	247 .. 367
BQ092342.1	110 .. 230

Whole Genome Shotgun sequences (1)

CAAK01000032.1	508682 .. 508802
--------------------------------	------------------

3. Splign

View the spliced alignment of your ZGC clone against the Sanger WGS contig.

The screenshot shows the NCBI Splign web interface. At the top left is the NCBI logo. To its right is the 'Splign' logo, which features a stylized DNA sequence with exons in yellow and introns in green. Below the logos is a navigation bar with links: HOME, SEARCH, SITE MAP, Overview, Online, Download, Documentation, and Contacts. The main content area has a heading 'Please specify input sequences by GI/Accession or in FASTA format. Examples (click to select):' followed by a list of four examples: AB010263 / NT_033778 (1 model, generic), AF034611 / NT_077569 (1 model, 67 exons), NM_020978 / NG_004750 (6 models, unaligned segment), and AF238306 / NT_033777 (1 model, many frameshifts). Below this list are two input fields: 'cDNA:' and 'Genomic:'. The 'cDNA:' field contains the text 'NM_001024735'. The 'Genomic:' field contains the text 'NM_639445'. To the right of the 'Genomic:' field are two input boxes labeled 'From:' and 'To:', with '1' and 'max' respectively. Below the 'cDNA:' field is a checkbox labeled 'Reverse and complement the query (cDNA)'. Below the 'Genomic:' field is a checkbox labeled 'Cross-species mode'. At the bottom left is a button labeled 'Align'. At the bottom right is a button labeled 'Browse...'. A yellow box with the text 'Select query and target sequences' has two arrows pointing to the 'cDNA:' and 'Genomic:' input fields. Another yellow box with the text 'You can also reverse and complement the query sequence or run the comparison in cross-species mode' has an arrow pointing to the 'Cross-species mode' checkbox.

NCBI

Splign

HOME SEARCH SITE MAP Overview Online Download Documentation Contacts

Please specify input sequences by GI/Accession or in FASTA format.
Examples (click to select):

- AB010263 / NT_033778 (1 model, generic)
- AF034611 / NT_077569 (1 model, 67 exons)
- NM_020978 / NG_004750 (6 models, unaligned segment)
- AF238306 / NT_033777 (1 model, many frameshifts)

cDNA: NM_001024735

Genomic: NM_639445 From: 1 To: max

☐ Reverse and complement the query (cDNA)

☐ Cross-species mode

Align

Browse...

You can also reverse and complement the query sequence or run the comparison in cross-species mode

Alignment Results:
Click on each exon to view alignment statistics, cds alignment (pink) and splice sites (green)

NCBI Spleign

HOME SEARCH SITE MAP Overview Online Download Documentation Contacts

#	Query	Subject	Span	Coverage	Overall	Exon	CDS	In-frame
1	NM_001024735(+)	NW_639445(-)	83063-94791	100.00	99.37	99.37	99.47	99.47

Graphics | Text

Model 1

	Coverage	100.00	CDS	99.47	Mismatches and indels	8
Overall	99.37	In-frame	99.47	Exons (min/max/ave)	61 / 175 / 115	
Exon	99.37	Primary transcript	1269	Introns (min/max/ave)	80 / 3936 / 1047	

NM_001024735 (+) Danio rerio zgc:63996 (zgc:63996), mRNA

NW_639445 (-) Danio rerio chromosome Un genomic contig

Segment 8: 82bp (808-889)

83063 94791 [Flip]

Segment Alignment

1 2 3 4 5 6 1:94791

7 8 9 10 11

ATCACCGGCAAGTACACGGATCCTCTTCCGGCCTGAAAAGCTAGGCTGCGAGCGTGGATTTTATCTGTG

ATCACCGGCAAGTACACGGATCCTCTTCCGGCCTGAAAAGCTAGGCTGCGAGCGTGGATTTTATCTGT

71:94721

CTCTTTAAACACTTTAAAGATGAGCGTACCGGCGTTTATTGACATAACGGAGGAGGATCAG.....

CTCTTTAAACACTTTAAATATGAGCGTACCGGCGTTTATTGACATAACGGAGGAGGATCAGGTATG

cds alignment

splice site